The CRing Project

Founded by Akhil Mathew
Continued by Markus J. Pflaum

January 26, 2023

Authors

The following people have contributed to this work, in alphabetical order:

Shishir Agrawal
Eva Belmont
Zev Chonoles
Rankeya Datta
Anton Geraschenko
Sherry Gong
François Greer
Darij Grinberg
Aise Johan de Jong
Adeel Ahmad Khan
Frédéric Latrémolière
Holden Lee
Geoffrey Lee
Daniel Martin
Michael Martinez
Akhil Mathew
Markus J. Pflaum
Ryan Reich
William Wright
Moor Xu

Copyright

Contents

Attribution

The following table lists the authors and contributors for each chapter or section of this monograph. If no source is mentioned, the corresponding author(s) submitted their work to the CRing Project. All contributions have been released under the GNU Free Documentation License, Version 1.3, see Copyright.

Chapter/Section	Author(s)	Contributor(s)	Source
0.1 Category Theory	A. Mathew		
0.2 Number Systems	M.J. Pflaum		
II.1 General Topology	F. Latrémolière	M.J. Pflaum	Liber Mathematicae, Point Set Topology

Part 0.

Foundations

0.1. Category theory

1.1. Introduction

The language of categories is not strictly necessary to understand the basics of commutative algebra or topology. Nonetheless, it is extremely convenient, powerful and actually will become indispensible for advanced topics such as "homological algebra" or "homotopy theory". Moreover, category theory will clarify many of the constructions made in the future when we can freely use terms like "universal property" or "adjoint functor". As a result, we begin this book with an introduction to category theory. The interested reader can pursue further study in Mac Lane (1998) or Kashiwara & Schapira (2006).

For the beginning, the reader is advised not to take the present chapter too seriously; skipping it for the moment to the following chapters and returning here as a reference could be quite reasonable.

1.2. Objects, morphisms, and categories

1.2.1. Definitions and first examples

1.2.2 Categories are supposed to be places where mathematical objects live. Intuitively, to any given type of structure (e.g. groups, rings, etc.), there should be a category of objects with that structure. These are not, of course, the only type of categories, but they will be the primary ones of concern to us in this book.

The basic idea of a category is that there should be objects, and that one should be able to map between objects. These mappings could be functions, and they often are, but they don't have to be. Next, one has to be able to compose mappings, and associativity and unit conditions are required. Nothing else is required.

1.2.3 Definition A (*locally small*) *category* C consists of:

- a collection of sets called *objects*,
- for each pair of objects X, Y a set $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ of *morphisms* from X to Y such that for every quadruple of objects X, X', Y, Y' the morphism sets $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ and $\mathrm{Mor}_{\mathsf{C}}(X', Y')$ are disjoint whenever $(X, Y) \neq (X', Y')$,
- for every object X an *identity morphism* $\mathrm{id}_X \in \mathrm{Mor}_{\mathsf{C}}(X, X)$, and

- for every triple X, Y, Z of objects a *composition law*

$$\circ_{(X,Y,Z)} : \mathrm{Mor}_{\mathsf{C}}(X,Y) \times \mathrm{Mor}_{\mathsf{C}}(Y,Z) \to \mathrm{Mor}_{\mathsf{C}}(X,Z), (f,g) \to g \circ f.$$

It is further required that these data fulfill the following two axioms:

($\mathsf{Cat}1$) The composition law is *associative* which means that for every quadrupel of objects X, Y, Z, W and all $f \in \mathrm{Mor}_{\mathsf{C}}(X,Y)$, $g \in \mathrm{Mor}_{\mathsf{C}}(Y,Z)$ and $h \in \mathrm{Mor}_{\mathsf{C}}(Z,W)$ the relation

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

holds true.

($\mathsf{Cat}2$) The composition law is *unital* with units given by the identity morphism. This means that for each pair of objects X, Y and every morphism $f \in \mathrm{Mor}_{\mathsf{C}}(X,Y)$ the relation

$$\mathrm{id}_Y \circ f = f \circ \mathrm{id}_X = f$$

holds true.

1.2.4 Remarks (a) In practice, a category C will often be the storehouse for mathematical objects such as groups, Lie algebras, rings, manifolds, etc., in which case the corresponding morphisms will be (induced by) ordinary functions preserving the underlying structure of the objects of the category. More precisely, the objects of such categories are *structured sets* that means ordered pairs $(X, \mathcal{S})$, where X is a set, called the (*underlying*) *space*, and $\mathcal{S}$ is the so-called *structure* on X; see (Bourbaki, 2004, Chap. IV) for the theory of structures, and (Moschovakis, 2006, 4.30) for structured sets. A topology on a set X, a group operation plus an identity element, a sheaf of rings on a topological space X, a manifold structure, a σ-algebra with a measure, or (compatible) combinations of these all form examples of a structure on the space X. Morphisms between two structured sets $(X, \mathcal{S})$ and $(Y, \mathcal{T})$ of the same type - meaning the structures are both topologies, or both group operations with identity elements, and so on - are then functions $f : X \to Y$ preserving the structures $\mathcal{S}$ and $\mathcal{T}$. For example, the structure preserving maps are the continuous functions if the structures are topologies and they are homomorphisms when the structures are group operations with identities. There is one - luckily curable - caveat with that concept. Consider for example the category of topological spaces, and consider the set $\mathbb{R}$ of real numbers. There are many topologies on $\mathbb{R}$, so let us pick for example the euclidean topology $\mathcal{T}_{\mathbb{R}}$ and the discrete topology $\mathcal{P}(\mathbb{R})$ (recall that $\mathcal{P}(Y)$ denotes the powerset of a set Y). In the category of topological spaces one then has $(\mathbb{R}, \mathcal{T}_{\mathbb{R}}) \neq (\mathbb{R}, \mathcal{P}(\mathbb{R}))$. The identity map $\mathrm{id}_{\mathbb{R}}$ now is continuous from $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ and continuous from $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ (but not vice versa). Hence, $\mathrm{id}_{\mathbb{R}}$ would be regarded as a morphism both from $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ and from $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ in violation of the requirement that $\mathrm{Mor}_{\mathsf{C}}(X,Y) \cap \mathrm{Mor}_{\mathsf{C}}(X',Y') = \varnothing$ for $(X,Y) \neq (X',Y')$. This deficiency can be healed by a slight modification of the notion of a morphism between structured sets. Let $f : X \to Y$ be a structure preserving map between the underlying spaces of two structured sets $(X, \mathcal{S})$ and $(Y, \mathcal{T})$. The function f then can be understood as a triple (X, Y, Γ_f), where Γ_f denotes the graph of the function. Now we replace the domain X in this triple by the structured set $(X, \mathcal{S})$, and the range Y by the structured set $(Y, \mathcal{T})$, and obtain the triple $\big((X, \mathcal{S}), (Y, \mathcal{T}), \Gamma_f\big)$. We shortly denote this new triple by $f : (X, \mathcal{S}) \to (Y, \mathcal{T})$ and call it the *morphism from* $(X, \mathcal{S})$ *to* $(Y, \mathcal{T})$ *induced by the map* $f : X \to Y$. In other words, $f : X \to Y$ has been *enriched* by the structures on X and Y to give the morphism $f : (X, \mathcal{S}) \to (Y, \mathcal{T})$. Often one still writes $f : X \to Y$ for the resulting morphism, as long as it is clear that it is regarded as a morphism in a category of structured sets.

(b) Even when the category under consideration does not come from one of structured sets and structure preserving maps, we shall write $f : X \to Y$ to denote an element of $\mathrm{Mor}_{\mathsf{C}}(X, Y)$. Moreover, if the context indicates which underlying category is meant, we usually write $\mathrm{Mor}(X, Y)$ instead of $\mathrm{Mor}_{\mathsf{C}}(X, Y)$. Likewise, and as already practiced in the preceding definition, we abbreviate $\circ_{(X,Y,Z)}$ by $\circ$ because this keeps notation clear and will not lead to confusion.

(c) A morphism f of a category C uniquely determines a pair of objects (X, Y) such that $f \in \mathrm{Mor}_{\mathsf{C}}(X, Y)$. One calls X the *source* or *domain* of f, and Y the *target*, *range* or *codomain* of f.

(d) Unless stated differently, categories in this book are assumed to be locally small which means that the collection of morphisms between two objects forms a set, or in other words, using language by (Bourbaki, 2004, Chap. II), that the relation of being a morphism between two given objects is *collectivizing*.

Here is a simple list of examples.

1.2.5 Examples (Categories of structured sets) (a) Sets as objects together with functions between them as morphisms form a category which is denoted by Ens.

(b) Groups together with (group) homomorphisms as morphisms form a category denoted by Grp.

(c) Topological spaces and continuous maps between them form the category Top.

(d) Given a field $\Bbbk$, the vector spaces over $\Bbbk$ together with the $\Bbbk$-linear maps between them as morphisms form a category which we denote by $\mathsf{Vect}_{\Bbbk}$.

(e) The objects of the category $\mathsf{LieAlg}_{\Bbbk}$ are the Lie algebras over the field $\Bbbk$, its morphisms are Lie algebras homomorphisms, i.e. $\Bbbk$-linear maps which preserve the Lie brackets.

1.2.6 Example This example is slightly more subtle. Here the category has objects consisting of topological spaces, but the morphisms between two topological spaces X, Y are the *homotopy classes* of continuous maps $X \to Y$. Since composition respects homotopy classes, the composition of homotopy classes of maps is well-defined. The identity morphisms in this category are obviously the homotopy classes of the identity maps. The resulting category is called the *homotopy category of topological spaces* and is denoted by hTop. See Section 3.1 for further details.

1.2.7 Remark In general, the objects of a category do not have to form a set; they can be too large for that. For instance, the collection of objects in Ens does not form a set.

1.2.8 Definition A category is called *small* if the collection of objects is a set.

The standard examples of categories are the ones above: structured sets together with structure-preserving maps betwen them. Nonetheless, one can easily give other examples that are not of this form.

1.2.9 Example (Groups as categories) Let G be a group. Then we can make a category B_G where the objects just consist of one element $*$ and the maps $* \to *$ are the elements $g \in G$. The identity is the identity of G and composition is multiplication in the group.

In this case, the category does not represent much of a class of objects, but instead we think of the composition law as the key thing. So a group is a special kind of (small) category.

1.2.10 Example (Monoids as categories) A monoid is precisely a category with one object. Recall that a *monoid* is a set together with an associative and unital multiplication (but which need not have inverses).

1.2.11 Example (Posets as categories) Let $(P, \leqslant)$ be a partially ordered set (i.e. a poset). Then P can be regarded as a (small) category, where the objects are the elements $p \in P$, and

$$\mathrm{Mor}_P(p,q) = \begin{cases} \{(p,q)\}, & \text{if } p \leqslant q, \\ \varnothing, & \text{otherwise.} \end{cases}$$

The composition $(q,r) \circ (p,q)$ of two arrows (q,r) and (p,q), where $p \leqslant q \leqslant r$, is defined as the arrow (p,r). The identity morphism of an object $p \in P$ is the pair (p,p).

1.2.12 Remark There is, however, a major difference between category theory and set theory. There is *nothing* in the language of categories that lets one look *inside* an object. We think of vector spaces having elements, spaces having points, etc. By contrast, categories treat these kinds of things as invisible. There is nothing "inside" of an object $X \in \mathsf{C}$; the only way to understand X is to understand the ways one can map into and out of X. Even if one is working with a category of "structured sets," the underlying set of an object in this category is not part of the categorical data. However, there are instances in which the "underlying set" can be recovered as a Mor-set.

1.2.13 Example In the category Top of topological spaces, one can in fact recover the "underlying set" of a topological space via the Mor-sets. Namely, for each topological space X, the points of X are the same thing as the mappings from a one-point space into X. That is, we have

$$X = \mathrm{Mor}_{\mathsf{Top}}(1, X),$$

or more precisely

$$X = \mathrm{Mor}_{\mathsf{Top}}\big((1, \{\varnothing, 1\}), (X, \mathcal{T})\big),$$

where 1 denotes the one-point space $\{\varnothing\}$, $\{\varnothing, 1\}$ the discrete topology on 1, and $\mathcal{T}$ is the topology on X.

Later we will say that the functor assigning to each space its underlying set is *corepresentable.*

1.2.14 Example Let Ab be the category of abelian groups and group homomorphisms. Again, the claim is that using only this category, one can recover the underlying set of a given abelian group A. This is because the elements of A can be canonically identified with *morphisms* $\mathbb{Z} \to A$ (based on where $1 \in \mathbb{Z}$ maps).

1.2.15 Definition We say that C is a *subcategory* of the category D if the collection of objects of C is a subclass of the collection of objects of D, and if whenever X, Y are objects of C, we have

$$\mathrm{Mor}_{\mathsf{C}}(X,Y) \subset \mathrm{Mor}_{\mathsf{D}}(X,Y)$$

with the laws of composition in C induced by that in D.

C is called a *full subcategory* if $\mathrm{Mor}_{\mathsf{C}}(X,Y) = \mathrm{Mor}_{\mathsf{D}}(X,Y)$ whenever X, Y are objects of C.

1.2.16 Example The category of abelian groups is a full subcategory of the category of groups.

The language of commutative diagrams

While the language of categories is, of course, purely algebraic, it will be convenient for psychological reasons to visualize categorical arguments through diagrams. We shall introduce this notation here.

Let C be a category, and let X, Y be objects in C. If $f \in \mathrm{Mor}(X,Y)$, we shall sometimes write f as an arrow

$$f : X \to Y$$

or

$$X \xrightarrow{f} Y$$

as if f were an actual function. If $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ are morphisms, composition $g \circ f : X \to Z$ can be visualized by the picture

$$X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Finally, when we work with several objects, we shall often draw collections of morphisms into diagrams, where arrows indicate morphisms between two objects.

1.2.17 Convention A diagram will be said to *commute* if whenever one goes from one object in the diagram to another by following the arrows in the right order, one obtains the same morphism. For instance, the commutativity of the diagram

$$\begin{tikzcd}
X \arrow[r, "f'"] \arrow[d, "f"'] & W \arrow[d, "g"] \\
Y \arrow[r, "g'"'] & Z
\end{tikzcd}$$

is equivalent to the assertion that

$$g \circ f' = g' \circ f \in \mathrm{Mor}(X,Z) \ .$$

As an example, the assertion that the associative law holds in a category C can be stated as follows. For every quadruple $X, Y, Z, W \in \mathsf{C}$, the following diagram (of sets) commutes:

$$\begin{tikzcd}
\mathrm{Mor}(X,Y) \times \mathrm{Mor}(Y,Z) \times \mathrm{Mor}(Z,W) \arrow[r] \arrow[d] & \mathrm{Mor}(X,Z) \times \mathrm{Mor}(Z,W) \arrow[d] \\
\mathrm{Mor}(X,Y) \times \mathrm{Mor}(Y,W) \arrow[r] & \mathrm{Mor}(X,W).
\end{tikzcd}$$

Here the maps are all given by the composition laws in C. For instance, the downward map to the left is the product of the identity on $\mathrm{Mor}(X,Y)$ with the composition law $\mathrm{Mor}(Y,Z) \times \mathrm{Mor}(Z,W) \to \mathrm{Mor}(Y,W)$.

Isomorphisms

Classically, one can define an isomorphism of groups as a bijection that preserves the group structure. This does not generalize well to categories, as we do not have a notion of "bijection," as there is no way (in general) to talk about the "underlying set" of an object. Moreover, this definition does not generalize well to topological spaces: there, an isomorphism should not just be a bijection, but something which preserves the topology (in a strong sense), i.e. a homeomorphism.

Thus we make:

1.2.18 Definition An *isomorphism* between objects X, Y in a category C is a morphism $f : X \to Y$ such that there exists $g : Y \to X$ with

$$g \circ f = \mathrm{id}_X \quad \text{and} \quad f \circ g = \mathrm{id}_Y \ .$$

Such a g is called an *inverse* to f. An isomorphism of the form $f : X \to X$ that means an isomorphisms where the source and target coincide is called an *automorphism* (*of* X).

1.2.19 Lemma *The inverse of an isomorphism $f : X \to Y$ in a category C is uniquely determined.*

Proof. It is easy to check that the inverse g is unique. Indeed, suppose g, g' both were inverses to f. Then

$$g' = g' \circ \mathrm{id}_Y = g' \circ (f \circ g) = (g' \circ f) \circ g = \mathrm{id}_X \circ g = g.$$ □

1.2.20 Remark The above notion of an isomorphism is more correct than the idea of being one-to-one and onto. For instance, a bijection, even a continuous one, of topological spaces is not necessarily a homeomorphism, i.e. an isomorphism in the category of topological spaces.

1.2.21 Example It is easy to check that an isomorphism in the category Grp is an isomorphism of groups, that an isomorphism in the category Ens is a bijection, and so on.

1.2.22 Remarks (a) We are supposed to be able to identify isomorphic objects. In the categorical sense, this means mapping into X should be the same as mapping into Y, if X, Y are isomorphic, via an isomorphism $f : X \to Y$. Indeed, let Z be another object of C. Then we can define a map

$$f^* : \mathrm{Mor}_{\mathsf{C}}(Z, X) \to \mathrm{Mor}_{\mathsf{C}}(Z, Y)$$

given by post-composition with f. This is a *bijection* if f is an isomorphism (the inverse is given by postcomposition with the inverse to f). Similarly, one can easily see that mapping *out of* X is essentially the same as mapping out of Y. Anything in general category theory that is true for X should be true for Y (as general category theory can only try to understand X in terms of morphisms into or out of it!).

(b) The relation “X, Y are isomorphic” is an equivalence relation on the class of objects of a category C.

(c) Let P be a partially ordered set, and make P into a category as in Example 1.2.11. Then P is a poset if and only if two isomorphic objects are equal.

1.2.23 Definition A *groupoid* is a category where every morphism is an isomorphism.

1.2.24 Remark If G is a groupoid and x an object of G, the set $\mathsf{G}_x := \mathrm{Mor}_{\mathsf{G}}(x, x)$ is a group. It is called the *isotropy group* of G at x. A group is essentially the same as a groupoid with one object.

1.2.25 Example Let X be a topological space, and let $\pi_1(X)$ be the category defined as follows: the objects are elements of X, and morphisms $x \to y$ (for $x, y \in X$) are homotopy classes of maps $\gamma : [0,1] \to X$ (i.e. paths) that send $0 \mapsto x$ and $1 \mapsto y$. Composition of maps is given by concatenation of paths. Because one is working with homotopy classes of paths, composition is associative, indeed. The identity at $x \in X$ is given by the constant path $\varepsilon_x : [0,1] \to X$, $t \mapsto x$. The inverse of a path γ in X is obtained by “going the path backwards” which means by the path $\gamma^- : [0,1] \to X$, $t \mapsto \gamma(1-t)$. The groupoid $\pi_1(X)$ is called the *fundamental groupoid* of X. Note that $\mathrm{Mor}_{\pi_1(X)}(x, x)$ is the *fundamental group* $\pi_1(X, x)$. For details and proofs of this example see (Brown, 2006, Chap. 6).

Monomorphisms and epimorphisms

Besides isomorphisms, one can also charaterize monomorphisms and epimorphisms in a purely categorical setting. That is what we wish to do now. In categories where there is an underlying set the notions of injectivity and surjectivity makes sense but in category theory, one does not in a sense have “access” to the internal structure of objects. In this light, we make the following definition.

1.2.26 Definition A morphism $f : X \to Y$ is a *monomorphism* if for any two morphisms $g_1 : X' \to X$ and $g_2 : X' \to X$ the relation $fg_1 = fg_2$ implies $g_1 = g_2$. A morphism $f : X \to Y$ is an *epimorphism* if for any two maps $g_1 : Y \to Y'$ and $g_2 : Y \to Y'$ the equality $g_1 f = g_2 f$ implies $g_1 = g_2$.

So $f : X \to Y$ is a monomorphism if and only if whenever X' is another object in C, the map

$$f^* : \mathrm{Mor}_{\mathsf{C}}(X', X) \to \mathrm{Mor}_{\mathsf{C}}(X', Y),\ g \mapsto f \circ g$$

is an injection (of sets). Similarly, $f : X \to Y$ is an epimorphisms if and only if for every object Y' in C the map

$$f_* : \mathrm{Mor}_{\mathsf{C}}(Y, Y') \to \mathrm{Mor}_{\mathsf{C}}(X, Y'),\ g \mapsto g \circ f$$

is injective. Note that neither of these statements makes any reference to *surjections* of sets.

1.2.27 Proposition *The composite of two monomorphisms is a monomorphism, as is the composite of two epimorphisms.*

Proof. Assume that $f : X \to Y$ and $g : Y \to Z$ are both monomorphisms in a category C. Let $h_1, h_2 : X' \to X$ be two morphisms in C and assume that $(g \circ f) \circ h_1 = (g \circ f) \circ h_2$. By associativity of composition the equality $g \circ (f \circ h_1) = g \circ (f \circ h_2)$ follows, hence $f \circ h_1 = f \circ h_2$ since g is a monomorphism. But then $h_1 = h_2$ since f is a monomorphism. So $g \circ f$ is a monomorphism as well.

Now assume that $f : X \to Y$ and $g : Y \to Z$ are epimorphisms. Let $h_1, h_2 : Z \to Z'$ be morphisms in C such that $h_1 \circ (g \circ f) = h_2 \circ (g \circ f)$. By associativity of composition $(h_1 \circ g) \circ f = (h_2 \circ g) \circ f$, hence $h_1 \circ g = h_2 \circ g$ since f is an epimorphism. So $h_1 = h_2$ since g is an epimorphism. This proves that $g \circ f$ is an epimorphism, too, and the proof is finished. $\square$

1.3. Functors

A functor is a way of mapping from one category to another: each object is sent to another object, and each morphism is sent to another morphism. We shall study many functors in the sequel: localization, the tensor product, Mor, and fancier ones like $\mathrm{Tor}, \mathrm{Ext}$, and local cohomology functors. The main benefit of a functor is that it doesn't simply send objects to other objects, but also morphisms to morphisms: this allows one to get new commutative diagrams from old ones. This will turn out to be a powerful tool.

Covariant functors

Let C, D be categories. If C, D are categories of structured sets (of possibly different types), there may be a way to associate objects in D to objects in C. For instance, to every group G we can associate its *group ring* $\mathbb{Z}[G]$; to each topological space we can associate its *singular chain complex*, and so on. In many cases, given a map between objects in C preserving the relevant structure, there will be an induced map on the corresponding objects in D. It is from here that we define a *functor*.

1.3.1 Definition A *functor* $F : \mathsf{C} \to \mathsf{D}$ consists of a function $F : \mathsf{C} \to \mathsf{D}$ (that is, a rule that assigns to each object in C an object of D) and, for each pair $X, Y \in \mathsf{C}$, a map $F : \mathrm{Mor}_{\mathsf{C}}(X, Y) \to \mathrm{Mor}_{\mathsf{D}}(FX, FY)$, which preserves the identity maps and composition.

In detail, the last two conditions state the following.

(Func1) If $X \in \mathsf{C}$, then $F(\mathrm{id}_X)$ is the identity morphism $\mathrm{id}_{F(X)} : F(X) \to F(X)$.

(Func2) If $X \xrightarrow{f} Y \xrightarrow{g} Z$ are morphisms in C, then $F(g \circ f) = F(g) \circ F(f)$ as morphisms $F(X) \to F(Z)$. Alternatively, we can say that F *preserves commutative diagrams.*

In the last statement of the definition, note that if

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "h"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

is a commutative diagram in C, then the diagram obtained by applying the functor F, namely

$$\begin{tikzcd} F(X) \arrow[r, "F(f)"] \arrow[dr, "F(h)"'] & F(Y) \arrow[d, "F(g)"] \\ & F(Z) \end{tikzcd}$$

also commutes. It follows that applying F to more complicated commutative diagrams also yields new commutative diagrams.

Let us give a few examples of functors.

1.3.2 Example There is a functor from $\mathsf{Ens} \to \mathsf{Ab}$ sending a set S to the free abelian group $\mathbb{Z}[S] = \mathbb{Z}^{(S)}$ on the set. For the definition of a free abelian group, or more generally a free R-module over a ring R, see Definition 2.8.1.

1.3.3 Example Let X be a topological space. Then to it we can associate the set $\pi_0(X)$ of *connected components* of X.

Recall that the continuous image of a connected set is connected, so if $f : X \to Y$ is a continuous map and $X' \subset X$ connected, $f(X')$ is contained in a connected component of Y. It follows that π_0 is a functor $\mathsf{Top} \to \mathsf{Ens}$. In fact, it is a functor on the *homotopy category* as well, because homotopic maps induce the same maps on π_0.

1.3.4 Example Fix $n \in \mathbb{N}$. There is a functor from $\mathsf{Top} \to \mathsf{Ab}$ (categories of topological spaces and abelian groups) sending a space X to its n-th *singular homology* group $H_n(X)$. We know that given a map of spaces $f : X \to Y$, we get a map of abelian groups $f_* : H_n(X) \to H_n(Y)$. See (Dold, 1995, Sec. VI. 7) or (Hatcher, 2002, Chap. 2), for instance.

We shall often need to compose functors. For instance, we will want to see, for instance, that the *tensor product* (to be defined later, see Section 4.3) is associative, which is really a statement about composing functors. The following (mostly self-explanatory) definition elucidates this.

1.3.5 Definition If $\mathsf{C}, \mathsf{D}, \mathsf{E}$ are categories, and $F : \mathsf{C} \to \mathsf{D}$, $G : \mathsf{D} \to \mathsf{E}$ are covariant functors, then one defines the *composite functor*

$$G \circ F : \mathsf{C} \to \mathsf{E}$$

as the functor which sends an object X of C to the object $G(F(X))$ of E. Similarly, a morphism $f : X \to Y$ is sent to $G(F(f)) : G(F(X)) \to G(F(Y))$.

The composite functor $G \circ F$ is well-defined. To see this observe that for an object X of C the identity morphism id_X is mapped to

$$G \circ F(\mathrm{id}_X) = G(F(\mathrm{id}_X)) = G(\mathrm{id}_{F(X)}) = \mathrm{id}_{G(F(X))}.$$

Moreover, if $f : X \to Y$ and $g : Y \to Z$ are morphisms in C, then

$$\begin{aligned} G \circ F(g \circ f) &= G(F(g \circ f)) = G(F(f) \circ F(f)) = G(F(g)) \circ G(F(f)) = \\ &= \big((G \circ F)(g)\big) \circ \big((G \circ F)(f)\big), \end{aligned}$$

hence conditions $(\mathsf{Func1})$ and $(\mathsf{Func2})$ are both fulfilled for $G \circ F$.

1.3.6 Example (Category of categories) In fact, because we can compose functors, there is a *category of categories*. Let Cat have as objects the small categories and as morphisms the functors between them. Composition is defined as in Definition 1.3.5.

1.3.7 Example (Group actions) Fix a group G. Let us understand what a functor $\mathsf{B}_G \to \mathsf{Ens}$ is. Here B_G is the category of Example 1.2.9. The unique object $*$ of B_G goes to some set X. For each element $g \in G$, we get a morphism $g : * \to *$ and thus a map $\varphi_g : X \to X$. This is supposed to preserve the composition law (which in G is just multiplication), as well as identities. That means that the following diagram commutes for each $g, h \in G$:

$$\begin{tikzcd} X \arrow[r, "\varphi_h"] \arrow[dr, "\varphi_{gh}"'] & X \arrow[d, "\varphi_g"] \\ & X . \end{tikzcd}$$

Moreover, if $e \in G$ is the identity, then $\varphi_e = \mathrm{id}_X$. So a functor $\mathsf{B}_G \to \mathsf{Ens}$ is just a left G-action on a set X.

1.3.8 Example (Forgetful functors) An important example of functors is given by the following. Let C be a "category of structured sets", see Remark 1.2.4 (a) . Then, there is a functor $U : \mathsf{C} \to \mathsf{Ens}$ that sends a structured set to the underlying set. For instance, there is the functor from groups to sets that forgets the group structure or the functor from topological spaces to sets that associates to a topological space its underlying set. More generally, suppose given two categories C, D, such that C can be regarded as "structured objects in D". Then there is a functor $U : \mathsf{C} \to \mathsf{D}$ that forgets the structure. Such functors are called *forgetful functors*.

Contravariant functors

Sometimes what we have described above are called *covariant functors*. Indeed, we shall also be interested in similar objects that reverse the arrows, such as duality functors:

1.3.9 Definition A *contravariant functor* $\mathsf{C} \xrightarrow{F} \mathsf{D}$ (between categories C and D) is similar data as in Definition 1.3.1 except that a morphism $X \xrightarrow{f} Y$ now goes to a morphism $F(Y) \xrightarrow{F(f)} F(X)$. Composites are required to be preserved, albeit in the other direction. In other words, one requires (Func1) to hold true and

(Func 2)$^\circ$ If $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ are morphisms, then $F(g \circ f) = F(f) \circ F(g)$ as morphisms $F(Z) \to F(X)$.

We shall sometimes say just "functor" for *covariant functor*. When we are dealing with a contravariant functor, we will always say the word "contravariant."

A contravariant functor also preserves commutative diagrams, except that the arrows have to be reversed. For instance, if $F : \mathsf{C} \to \mathsf{D}$ is contravariant and the diagram

$$\begin{tikzcd} A \arrow[r] \arrow[d] & C \\ B \arrow[ur] & \end{tikzcd}$$

is commutative in C, then the diagram

$$\begin{tikzcd} F(A) & F(C) \arrow[l] \arrow[dl] \\ F(B) \arrow[u] & \end{tikzcd}$$

commutes in D.

1.3.10 Remark One can, of course, compose contravariant functors as in Definition 1.3.5. But the composition of two contravariant functors will be *covariant.* So there is no "category of categories" where the morphisms between categories are contravariant functors.

Similarly as in Example 1.3.7, we have:

1.3.11 Example A *contravariant* functor from B_G (defined as in Example 1.2.9) to Ens corresponds to a set with a *right* G-action.

1.3.12 Example (Singular cohomology) In algebraic topology, one encounters contravariant functors on the homotopy category of topological spaces via the *singular cohomology* functors $X \mapsto H^n(X;\mathbb{Z})$, see (Dold, 1995, Sec. VI. 7). Given a continuous map $f : X \to Y$, there is a homomorphism of groups

$$f^* : H^n(Y;\mathbb{Z}) \to H^n(X;\mathbb{Z}) .$$

1.3.13 Example (Duality for vector spaces) On the category $\mathsf{Vect}_{\Bbbk}$ of vector spaces over a field $\Bbbk$, we have the contravariant functor

$$V \mapsto V^\vee$$

sending a vector space V to its dual $V^\vee := \mathrm{Hom}(V,\Bbbk) := \mathrm{Mor}_{\mathsf{Vect}_{\Bbbk}}(V,\Bbbk)$. Given a linear map $f : V \to W$ of vector spaces, there is the induced map

$$f^\vee : W^\vee \to V^\vee, \quad \mu \mapsto \mu \circ f$$

which is called the *transpose* of f.

1.3.14 Example If we map $\mathsf{B}_G \to \mathsf{B}_G$ sending $* \mapsto *$ and $g \mapsto g^{-1}$, we get a contravariant functor.

We now give a useful (linguistic) device for translating between covariance and contravariance.

1.3.15 Definition (The opposite category) Let C be a category. Define the *opposite category* C^{op} of C to have the same objects as C but such that the morphisms between X, Y in C^{op} are those between Y and X in C.

There is a contravariant functor $\mathsf{C} \to \mathsf{C}^{\mathrm{op}}$. In fact, contravariant functors out of C are the *same* as covariant functors out of C^{op}.

As a result, when results are often stated for both covariant and contravariant functors, for instance, we can often reduce to the covariant case by using the opposite category.

1.3.16 Remark A map that is an isomorphism in C corresponds to an isomorphism in C^{op}.

Functors and isomorphisms

Now we want to prove a simple and intuitive fact: if isomorphisms allow one to say that one object in a category is "essentially the same" as another, functors should be expected to preserve this.

1.3.17 Proposition *If $f : X \to Y$ is an isomorphism in C, and $F : \mathsf{C} \to \mathsf{D}$ a functor, then $F(f) : FX \to FY$ is an isomorphism.*

The proof is quite straightforward, though there is an important point here. Note that the analogous result holds for *contravariant* functors too.

Proof. If we have maps $f : X \to Y$ and $g : Y \to X$ such that the composites both ways are identities, then we can apply the functor F to this, and we find that since

$$f \circ g = \mathrm{id}_Y, \quad g \circ f = \mathrm{id}_X,$$

it must hold that

$$F(f) \circ F(g) = \mathrm{id}_{F(Y)}, \quad F(g) \circ F(f) = \mathrm{id}_{F(X)}.$$

We have used the fact that functors preserve composition and identities. This implies that $F(f)$ is an isomorphism, with inverse $F(g)$. □

1.3.18 Remark Categories have a way of making things so general that they are trivial. Hence, this material is called general abstract nonsense. Moreover, there is another philosophical point about category theory to be made here: often, it is the definitions, and not the proofs, that matter. For instance, what matters here is not the theorem, but the *definition of an isomorphism.* It is a categorical one, and much more general than the usual notion via injectivity and surjectivity.

1.3.19 Examples (a) As a simple example, $\{0, 1\}$ and $I := [0, 1]$ are not isomorphic in the homotopy category of topological spaces (i.e. are not homotopy equivalent) because $\pi_0([0, 1]) = \{[0_I]\}$ while $\pi_0(\{0, 1\})$ has two elements, namely (the equivalence classes of) the constant maps 0_I and 1_I mapping I to 0 and 1, respectively.

(b) More generally, the higher homotopy group functors π_n, see Hatcher (2002), can be used to show that the n-sphere $\mathbb{S}^n$ is not homotopy equivalent to a point. For then $\pi_n(\mathbb{S}^n, *)$ would be trivial, and it is not.

1.4. Natural transformations

Definition and some examples

1.4.1 There is room, nevertheless, for something else. Instead of having something that sends objects to other objects, one could have something that sends an object to a map. This leads us to the following.

1.4.2 Definition Suppose $F, G : \mathsf{C} \to \mathsf{D}$ are functors. A *natural transformation* $\eta : F \to G$ consists of the following data:

(NTrans) For each object X in C, one has been given a morphism $\eta_X : FX \to GX$ in D such that for every morphism $f : X \to Y$ in C the diagram

$$\begin{tikzcd} FX \arrow[r, "F(f)"] \arrow[d, "\eta_X"'] & FY \arrow[d, "\eta_Y"] \\ GX \arrow[r, "G(f)"'] & GY \end{tikzcd}$$

commutes.

If η_X is an isomorphism for each object X, then we shall say that η is a *natural isomorphism.*

It is similarly possible to define the notion of a natural transformation between *contravariant* functors.

When we say that things are "natural" in the future, we will mean that the transformation between functors is natural in this sense. We shall use this language to state theorems conveniently.

1.4.3 Example (The double dual) Here is the canonical example of "naturality." Let $\mathsf{Vec}_{\Bbbk}^{\mathrm{fd}}$ be the category of finite-dimensional vector spaces over a given field $\Bbbk$, $\operatorname{char} \Bbbk = 0$. Let us further restrict the category such that the only morphisms are the isomorphisms of vector spaces. Denote the resulting category by C. For each object V of C, we know that there is an isomorphism

$$V \simeq V^\vee = \mathrm{Mor}_{\Bbbk}(V, \Bbbk),$$

because both have the same dimension.

Moreover, the maps $V \mapsto V$ and $V \mapsto V^\vee$ can both be extended to covariant functors on C. (Note that the dual $^\vee$ was defined as a *contravariant* functor in Example 1.3.13.) The first is the identity functor. For the second, if $f : V \to W$ is an isomorphism, then there is induced a transpose map $f^\vee : W^\vee \to V^\vee$ (defined by sending a map $W \to \Bbbk$ to the precomposition $V \xrightarrow{f} W \to \Bbbk$), which is an isomorphism. We can take its inverse. So we have two functors from C to itself, the identity and the " inverse dual", and we know that $V \simeq V^\vee$ for each V (though we have not chosen any particular set of isomorphisms).

However, the isomorphism $V \simeq V^\vee$ cannot be made natural. That is, there is no way of choosing isomorphisms

$$T_V : V \simeq V^\vee$$

such that, whenever $f : V \to W$ is an isomorphism of vector spaces, the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "T_V"'] & W \arrow[d, "T_W"] \\ V^\vee \arrow[r, "(f^\vee)^{-1}"'] & W^\vee. \end{tikzcd}$$

Indeed, fix $d > 1$, and choose $V = \Bbbk^d$. Identify $V^\vee$ with $\Bbbk^d$, and so the map T_V can be identified with a $d \times d$ matrix A with coefficients in $\Bbbk$. The requirement is that for each invertible $d \times d$ matrix B, we have

$$(B^t)^{-1}A = AB,$$

by considering the above diagram with $V = W = \Bbbk^d$, and f corresponding to the matrix B. This is impossible unless $A = 0$, by elementary linear algebra. Namely let $B = cI_d$, where $c \in \Bbbk\backslash\{0\}$ and I_d is the identity matrix on $\Bbbk^d$. Then check that $\det(A - \lambda I_d) = 0$ holds true if and only if $0 = \det B^t(A - \lambda I_d)B = \det(A - \lambda c^{2d} I_d)$. But this means that if A has eigenvalue λ, then A has also the eigenvalues λc^{2d} for all $c \in \Bbbk\backslash\{0\}$. Since A can have at most d different eigenvalues, this implies that A has 0 as its only eigenvalue which means $A = 0$.

Nonetheless, it is possible to choose for every finite dimensional $\Bbbk$-vector space V a natural isomorphism

$$V \simeq V^{\vee\vee} .$$

To do this, given V, recall that $V^{\vee\vee}$ is the collection of linear maps $V^\vee \to \Bbbk$. To give a linear map $V \to V^{\vee\vee}$ is thus the same as giving functions $l_V(v) : V^\vee \to \Bbbk$, $v \in V$ such that $l_V(v)$ is linear in v. We can do this by letting $l_V(v)$ be “evaluation at” v. That is, $l_V(v)$ sends a linear functional $\mu : V \to \Bbbk$ to $\mu(v) \in \Bbbk$.

First let us check that $l_V : V \to V^{\vee\vee}$, $v \mapsto l_V(v)$ is linear by computing for $v, w \in V$, $c \in \Bbbk$, and $\mu \in V^\vee$:

$$\begin{aligned} l_V(v + w)(\mu) &= \mu(v + w) = \mu(v) + \mu(w) = l_V(v)(\mu) + l_V(w)(\mu) = \big(l_V(v) + l_V(w)\big)(\mu), \\ l_V(cv)(\mu) &= \mu(cv) = c\,\mu(v) = c\,l_V(v)(\mu) . \end{aligned}$$

The linear map $l_V : V \to V^{\vee\vee}$ has trivial kernel since for each $v \in V\backslash\{0\}$ there exists a linear map $\mu : V \to \Bbbk$ such that $\mu(v) \neq 0$. Because V is finite dimensional, l_V therefore is an isomorphism of vector spaces.

Finally, $V \to V^{\vee\vee}$ is natural in V. To verify this observe first that for $f : V \to W$ linear, the map $f^{\vee\vee} : V^{\vee\vee} \to W^{\vee\vee}$ is defined by

$$\varrho \mapsto \Big(W^\vee \ni \mu \mapsto f^{\vee\vee}(\varrho)(\mu) = \varrho(f^*\mu) = \varrho(\mu \circ f) \in \Bbbk\Big) .$$

Now the diagram

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "l_V"'] & W \arrow[d, "l_W"] \\ V^{\vee\vee} \arrow[r, "f^{\vee\vee}"'] & W^{\vee\vee} \end{tikzcd}$$

commutes, since one has for $v \in V$ and $\mu \in W^\vee$:

$$\big(f^{\vee\vee} \circ l_V\big)(v)(\mu) = \big(f^{\vee\vee}(l_V(v))\big)(\mu) = l_V(v)(\mu \circ f) = \mu(f(v)) = l_W(f(v))(\mu) = \big(l_W \circ f\big)(\mu) .$$

In case V is an arbitrary, possibly infinite dimensional vector space, the natural transformation $l_V : V \to V^{\vee\vee}$ can still be defined, but need not always be an isomorphism.

1.4.4 Example Suppose there are two functors $\mathsf{B}_G \to \mathsf{Ens}$, i.e. two G-sets X and Y. Denote by $\varphi : G \to \mathrm{Mor}(X, X)$ and $\psi : G \to \mathrm{Mor}(Y, Y)$ the corresponding left G-actions. A natural transformation between them then is a G-equivariant map $F : X \to Y$ which means that the following diagram commutes for all $g \in G$:

$$\begin{tikzcd} X \arrow[r, "F"] \arrow[d, "\varphi_g"'] & Y \arrow[d, "\psi_g"] \\ X \arrow[r, "F"'] & Y . \end{tikzcd}$$

1.4.5 Natural transformations can be *composed*. Suppose given functors $F, G, H : \mathsf{C} \to \mathsf{D}$, a natural transformation $T : F \to G$, and a natural transformation $U : G \to H$. Then, for each $X \in \mathsf{C}$, we have maps $TX : FX \to GX, UX : GX \to HY$. We can compose U with T to get a natural transformation $U \circ T : F \to H$.

In fact, we can thus define a *category of functors* $\mathsf{Func}(\mathsf{C}, \mathsf{D})$ (at least if C, D are small). The objects of this category are the functors $F : \mathsf{C} \to \mathsf{D}$. The morphisms are natural transformations between functors. Composition of morphisms is as above.

Equivalences of categories

Often we want to say that two categories C, D are "essentially the same." One way of formulating this precisely is to say that C, D are *isomorphic* in the category of categories. Unwinding the definitions, this means that there exist functors

$$F : \mathsf{C} \to \mathsf{D}, \quad G : \mathsf{D} \to \mathsf{C}$$

such that $F \circ G = \mathrm{id}_{\mathsf{D}}, G \circ F = \mathrm{id}_{\mathsf{C}}$. This notion, of *isomorphism* of categories, is generally far too restrictive.

For instance, we could consider the category of all finite-dimensional vector spaces over a given field k, and we could consider the full subcategory of vector spaces of the form k^n. Clearly both categories encode essentially the same mathematics, in some sense, but they are not isomorphic: one has a countable set of objects, while the other has an uncountable set of objects. Thus, we need a more refined way of saying that two categories are "essentially the same."

1.4.6 Definition Two categories C, D are called *equivalent* if there are functors

$$F : \mathsf{C} \to \mathsf{D}, \quad G : \mathsf{D} \to \mathsf{C}$$

and natural isomorphisms

$$FG \simeq \mathrm{id}_{\mathsf{D}}, \quad GF \simeq \mathrm{id}_{\mathsf{C}}.$$

For instance, the category of all vector spaces of the form k^n is equivalent to the category of all finite-dimensional vector spaces. One functor is the inclusion from vector spaces of the form k^n; the other functor maps a finite-dimensional vector space V to $k^{\dim V}$. Defining the second functor properly is, however, a little more subtle. The next criterion will be useful.

1.4.7 Definition A covariant functor $F : \mathsf{C} \to \mathsf{D}$ is called *fully faithful* if for each pair of objects $X, Y \in \mathsf{C}$ the map $F : \mathrm{Mor}_{\mathsf{C}}(X, Y) \to \mathrm{Mor}_{\mathsf{D}}(FX, FY)$ is a bijection. The functor F is called *essentially surjective* if every object of D is isomorphic to an object of the form FX for some object X of C.

1.4.8 Example So, for instance, the inclusion of a full subcategory is fully faithful (by definition). The forgetful functor from groups to sets is not fully faithful, because not all functions between groups are automatically homomorphisms.

1.4.9 Theorem *A functor $F : \mathsf{C} \to \mathsf{D}$ between categories C and D induces an equivalence of categories if and only if it is fully faithful and essentially surjective.*

Proof. Let us first show that the condition is sufficient, and assume that F is fully faithful and essentially surjective. By essentially surjectivity we can then fix for any $Y \in \mathrm{Ob}(\mathsf{D})$ some $X_Y \in \mathrm{Ob}(\mathsf{C})$ and an isomorphism $\tau_Y : Y \to F(X)$. The fact that F is fully faithful means that for any $g \in \mathrm{Mor}_{\mathsf{D}}(Y_1, Y_2)$, there exists a unique $f_g \in \mathrm{Mor}_{\mathsf{C}}(X_{Y_1}, X_{Y_2})$ satisfying $F(f_g) = \tau_{Y_2} \circ g \circ \tau_{Y_1}^{-1}$. So define $G : \mathsf{D} \to \mathsf{C}$ by $G(Y) = X_Y$ and $G(g) = f_g$. To verify that G is a functor, first note that on an identity morphism we have $F(id_{X_Y}) = \tau_Y \circ id_Y \circ \tau_Y^{-1}$ so it must be that $G(id_Y) = id_{X_Y}$. Next consider the composition of morphisms: $Y_1 \xrightarrow{g_1} Y_2 \xrightarrow{g_2} Y_3$. Since $F(f_{g_2} \circ f_{g_1}) = F(f_{g_2}) \circ F(f_{g_1}) = (\tau_{Y_3} \circ g_2 \circ \tau_{Y_2}^{-1}) \circ (\tau_{Y_2} \circ g_1 \circ \tau_{Y_1}^{-1}) = \tau_{Y_3} \circ (g_2 \circ g_1) \circ \tau_{Y_1}^{-1} = F(f_{g_2 \circ g_1})$ we have that $G(g_2 \circ g_1) = G(g_2) \circ G(g_1)$ implying G is indeed a functor.

Now take a morphism $Y_1 \xrightarrow{g} Y_2$ in order to check commutativity of the diagram in D from Definition 1. Using the τ_Y's that are already defined makes commutativity clear; the bottom of the diagram can be expanded by recalling that $G(g)$ is defined so that $(F \circ G)(g) = \tau_{Y_2} \circ g \circ \tau_{Y_1}^{-1}$.

$$\begin{tikzcd}
 & Y_1 \arrow[r, "g"] \arrow[dl, "\tau_{Y_1}"'] \arrow[d, dotted, "id_{Y_1}"] & Y_2 \arrow[dr, "\tau_{Y_2}"] & \\
(F \circ G)(Y_1) \arrow[r, "\tau_{Y_1}^{-1}"] \arrow[rrr, bend right, "(F \circ G)(g)"'] & Y_1 \arrow[r, "g"] & Y_2 \arrow[r, "\tau_{Y_2}"] & (F \circ G)(Y_2)
\end{tikzcd}$$

For commutativity of the diagram in C, we must first define η_X's. For $X \in \mathrm{Ob}(\mathsf{C})$ we already have an isomorphism $\tau_{F(X)} : F(X) \to (F \circ G \circ F)(X)$. Since F is fully faithful, we may take $\eta_X \in \mathrm{Mor}_{\mathsf{C}}(X, (G \circ F)(X))$ to be the morphism satisfying $F(\eta_X) = \tau_{F(X)}$. Note that taking η_X^{-1} satisfying $F(\eta_X^{-1}) = \tau_{F(X)}^{-1}$ gives $\eta_X^{-1} \circ \eta_X = id_X$ and $\eta_X \circ \eta_X^{-1} = id_{(G \circ F)(X)}$ implying η_X is an isomorphism. So take some morphism $X_1 \xrightarrow{f} X_2$ and apply F to the diagram in C from Definition 1. Again to make commutativity clear the bottom is expanded by recalling that $(G \circ F)(f)$ is defined so that $(F \circ G \circ F)(f) = \tau_{F(X_2)} \circ F(g) \circ \tau^{-1}_{F(X_1)} = F(\eta_{X_2}) \circ F(g) \circ F(\eta_{X_1})^{-1}$.

$$\begin{tikzcd}
 & F(X_1) \arrow[r, "F(f)"] \arrow[dl, "F(\eta_{X_1})"'] \arrow[d, dotted, "id_{F(X_1)}"] & F(X_2) \arrow[dr, "F(\eta_{X_2})"] & \\
(F \circ G \circ F)(X_1) \arrow[r, "F(\eta_{X_1})^{-1}"] \arrow[rrr, bend right, "(F \circ G \circ F)(f)"'] & F(X_1) \arrow[r, "F(f)"] & F(X_2) \arrow[r, "F(\eta_{X_2})"] & (F \circ G \circ F)(X_2)
\end{tikzcd}$$

But F is faithful, so $F((G \circ F)(f) \circ \eta_{X_1}) = F(\eta_{X_2} \circ f)$ implies $(G \circ F)(f) \circ \eta_{X_1} = \eta_{X_2} \circ f$ as desired.

Next we show the condition to be necessary. So suppose that F induces an equivalence of categories and let G be its quasi-inverse. For any $Y \in \mathrm{Ob}(\mathsf{D})$ the isomorphism $\tau_Y : Y \to (F \circ G)(Y)$ shows that F is essentially surjective. To see that F is faithful suppose $F(f_1) = F(f_2)$ for some $f_1, f_2 \in \mathrm{Mor}_{\mathsf{C}}(X_1, X_2)$. Then commutativity of the diagram in C from Definition 1 gives $f_1 = \eta^{-1}_{X_2} \circ (G \circ F)(f_1) \circ \eta_{X_1} = \eta^{-1}_{X_2} \circ (G \circ F)(f_2) \circ \eta_{X_1} = f_2$. Note here that an analogous argument shows that G is faithful as well. Finally, take some $X_1, X_2 \in \mathrm{Ob}(\mathsf{C})$ and $g \in \mathrm{Mor}_{\mathsf{D}}(F(X_1), F(X_2))$. Set $f = \eta^{-1}_{X_2} \circ G(g) \circ \eta_{X_1}$. Using the diagram in C from Definition 1 again, we see that $(G \circ F)(f) = \eta_{X_2} \circ f \circ \eta^{-1}_{X_1}$ which is $G(g)$ by definition of f. Since G is faithful, it must be that $F(f) = g$ implying F is full and completing the proof. $\square$

1.4.10 Remark In the proof of the preceding theorem a strong version of the axiom of choice has been assumed. That is, we have assumed that for every class of nonempty sets there is choice function C on this class satisfying $C(x) \in x$ for each set x. This axiom is an extension of the Neumann-Bernays-Godel (NGB) axioms which, unlike the Zermelo-Fraenkel (ZF) axioms, make a distinction between a set and a proper class. Just as the consistency of (ZF) is independent of the truth or falsity of the axiom of choice for sets, the consistency of (NGB) is independent of the truth or falsity of the strong axiom of choice. For our purposes the axiom was required in order to simultaneously select objects and morphisms in one category corresponding to those in another category; the collections of eligible objects and morphisms may be proper classes.

1.5. Various universal constructions

Now that we have introduced the idea of a category and showed that a functor takes isomorphisms to isomorphisms, we shall take various steps to characterize objects in terms of maps (the most complete of which is the Yoneda lemma, **??**). In general category theory, this is generally all we *can* do, since this is all the data we are given. We shall describe objects satisfying certain "universal properties" here.

As motivation, we first discuss the concept of the "product" in terms of a universal property.

Products and coproducts

1.5.1 Recall that if we have two sets X and Y, the (cartesian) product $X \times Y$ is the set of all pairs (x, y) where $x \in X$ and $y \in Y$. The product is also equipped with natural projections $p_1 : X \times Y \to X$ and $p_2 : X \times Y \to Y$ that take (x, y) to x and y respectively. Thus any element of $X \times Y$ is uniquely determined by where it projects to on X and Y. In fact, this is the case more generally; if we have an index set J and a family of sets $(X_j)_{j \in J}$, then the product $X = \prod_{j \in J} X_j$ of $(X_j)_{j \in J}$ consist of all functional graphs $x \subset J \times \Big(\bigcup\limits_{j \in J} X_j \Big)$ with domain J such that $x(j) \in X_j$ for all $j \in J$; see (Bourbaki, 2004, II. §5.3). An element $x \in X$ therefore is determined uniquely by where the projections $p_j(x) := x_j := x(j)$ land in X_j.

To get into the categorical spirit, we should speak not of elements of X but of maps or better morphisms to X. Here is the general observation: if we have any other set Y with maps $f_j : Y \to X_j$ for all $j \in J$ then there is a unique map $f : Y \to X = \prod_{j \in J} X_j$ such that $p_j \circ f = f_j$ for all $j \in J$. The map f is given by sending $y \in Y$ to the element $f(y) := (f_j(y))_{j \in J}$. This leads to the following characterization of a product using only "mapping properties."

1.5.2 Definition Let $(X_j)_{j \in J}$ be a family of objects in some category C. Then an object X of C equipped with morphisms $p_j : X \to X_j$, $j \in J$, called (*canonical*) *projections* is said to be the *product* of the objects X_j, $j \in J$, if the following universal property holds:

(Prod) Let Y be any other object in C with a family of maps $f_j : Y \to X_j$, $j \in J$. Then there is a unique morphism $f : Y \to X$ such that $p_j \circ f = f_j$ for all $j \in J$ or in other words such that the diagram

$$\begin{tikzcd} Y \arrow[d, "f"'] \arrow[dr, "f_j"] & \\ X \arrow[r, "p_j"'] & X_j \end{tikzcd}$$

commutes for every $j \in J$.

One usually denotes the uniquely determined map $f : Y \to Y$ by $(f_j)_{j \in J}$ or $(f_1, \ldots, f_n)$ when $J = \{1, \ldots, n\}$ and calls it the *pairing* of f_1 and f_2 in the special case where $J = \{1, 2\}$. The product of the family $(X_j)_{j \in J}$ is denoted $\prod_{j \in J} X_j$.

1.5.3 Remarks (a) Proposition 1.5.5 below tells that the product of a family $(X_j)_{j \in J}$ of objects of C is unique up to unique isomorphism. This observation justifies our agreement to denote the product of $(X_j)_{j \in J}$ by a unique symbol.

(b) Rephrasing the defining property of the product, to map into $X = \prod_{j \in J} X_j$ is the same as mapping into all the X_j at once. Note that, however, the product need not exist! More abstractly, the meaning of condition (Prod) can be expressed equivalently by the existence of a natural isomorphism of contravariant functors

$$\mathrm{Mor}\Big(-, \prod_{j \in J} X_j\Big) \simeq \prod_{j \in J} \mathrm{Mor}(-, X_j)$$

which associates to every object Y of C the bijection

$$\left((p_j)_*\right)_{j\in J} : \mathrm{Mor}\left(Y, \prod_{j\in J} X_j\right) \to \prod_{j\in J} \mathrm{Mor}(Y, X_j),\ f \mapsto \left(p_j \circ f\right)_{j\in J}\ .$$

Note that this observation shows that products in the category of sets are really fundamental to the idea of products in any category.

1.5.4 Example One of the benefits of the preceding definition is that an actual category is not specified; thus when we take C to be Ens, we recover the cartesian product of sets by 1.5.1, but if we take C to be Grp or Top, we achieve the regular notion of the product of groups, see **??**, respectively of topological spaces, see **??**.

The categorical product is not unique, but it is as close to being so as possible.

1.5.5 Proposition (Uniqueness of products) *Any two products of a family $(X_j)_{j\in J}$ of objects in C are isomorphic by a unique isomorphism commuting with the projections.*

1.5.6 Remark This is a special case of a general "abstract nonsense" type result that we shall see many more of in the sequel. The precise statement is the following: let X be a product of the family $(X_j)_{j\in J}$ with projections $p_j : X \to X_j$, and let Y be a product of them too, with projections $q_j : Y \to X_j$. Then the claim is that there is a *unique* isomorphism

$$f : X \to Y$$

such that the diagram below commutes for each $j \in J$:

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr, "p_j"'] & & Y \arrow[dl, "q_j"] \\ & X_j\ . & \end{tikzcd} \tag{1.5.1}$$

Proof of the Proposition. Indeed, note that the projections $p_j : X \to X_j$ and the fact that mapping into Y is the same as mapping into all the X_j give a unique map $f : X \to Y$ making the diagram (1.5.1) commute for every $j \in J$. The same reasoning (applied to the $q_j : Y \to X_j$) gives a map $g : Y \to X$ making the diagram

$$\begin{tikzcd} Y \arrow[rr, "g"] \arrow[dr, "q_j"'] & & X \arrow[dl, "p_j"] \\ & X_j & \end{tikzcd} \tag{1.5.2}$$

commute for every $j \in J$. By piecing the two diagrams together, it follows that the composite $g \circ f$ makes the diagrams

$$\begin{tikzcd} X \arrow[rr, "g\circ f"] \arrow[dr, "p_j"'] & & X \arrow[dl, "p_j"] \\ & X_j & \end{tikzcd} \tag{1.5.3}$$

commute, too. But the identity $\mathrm{id}_X : X \to X$ also would make (1.5.3) commute for every $j \in J$, so the *uniqueness* assertion in the definition of the product shows that $g \circ f = \mathrm{id}_X$. Similarly, $f \circ g = \mathrm{id}_Y$. We are done. ☐

1.5.7 Example Let P be a poset, and make P into a category as in Example 1.2.11. Fix $p, q \in P$. The product of p, q then is the greatest lower bound of $\{p, q\}$ (if it exists). This claim holds more generally for arbitrary subsets of P. In particular, consider the poset of subsets of a given set S. Then the product in this category corresponds to the intersection of subsets.

1.5.8 If we reverse the arrows in the above construction, the universal property obtained (known as the "coproduct") characterizes disjoint unions in the category of sets and free products in the category of groups. That is, to map *out* of a coproduct of objects $(X_j)_{j \in J}$ is the same as mapping out of each of these. Let us make this precise.

1.5.9 Definition Let $(X_j)_{j\in J}$ be a family of objects in a category C. Then an object $X \in \mathsf{C}$ together with morphisms $i_j : X_j \to X$ called (*canonical*) *injections* is said to be the *coproduct* of the objects X_j, $j \in J$, if the following universal property holds:

(Coprod) Let Y be any other object in C with a family of maps $g_j : X_j \to Y$, $j \in J$. Then there is a unique morphism $g : X \to Y$ such that $g \circ i_j = g_j$ for all $j \in J$ or in other words such that the diagram

$$\begin{tikzcd} X_j \arrow[r, "i_j"] \arrow[d, "g_j"'] & X \arrow[dl, "g"] \\ Y & \end{tikzcd}$$

commutes for every $j \in J$.

One usually writes $\langle g_j \rangle_{j\in J}$ or, when $J = \{1, \ldots, n\}$, $\langle g_1, \ldots, g_n \rangle$ for the uniquely determined map $g : X \to Y$. Sometimes the notation $(g_j)_{j\in J}$ respectively $(g_1, \ldots, g_n)$ is used. In the special case where $J = \{1, 2\}$, $g = \langle g_1, g_2 \rangle$ is called the *copairing* of g_1 and g_2. The coproduct of the family $(X_j)_{j\in J}$ is denoted $\coprod_{j\in J} X_j$.

1.5.10 Remarks (a) Despite the name indicates otherwise, the canonical injections $i_j : X_j \to X$, $j \in J$ of a coproduct of a family $(X_j)_{j\in J}$ need not be injective when the underlying category consist of structered sets and maps between them. In general, the canonical injections even need not be monomorphisms.

(b) Like for the product, the use of a particular symbol for the coproduct of a family $(X_j)_{j\in J}$ is justified by the fact that that coproduct are uniquely determined up to unique isomorphism, see Proposition 1.5.11.

(c) Analogously as for products, condition (Coprod) can be expressed equivalently by requiring that there exists a natural isomorphism of covariant functors

$$\mathrm{Mor}\Big(\coprod_{j\in J} X_j, -\Big) \simeq \prod_{j\in J} \mathrm{Mor}(X_j, -)$$

which associates to every object Y of C the bijection

$$\big((i_j)^*\big)_{j\in J} : \mathrm{Mor}\Big(\prod_{j\in J} X_j, Y\Big) \to \prod_{j\in J} \mathrm{Mor}(X_j, Y),\ g \mapsto \big(g \circ i_j\big)_{j\in J} .$$

1.5.11 Proposition (Uniqueness of coproducts) *The coproduct of a family $(X_j)_{j\in J}$ of objects in C is uniquely determined up to unique isomorphisms commuting with the canonical injections.*

Proof. Dually to the product case we have to show that for a coproduct X of the family $(X_j)_{j\in J}$ with canonical injections $i_j : X_j \to X$ and another coproduct Y with canonical injections $k_j : X_j \to Y$ there is a unique isomorphism

$$f : X \to Y$$

such that the diagram below commutes for each $j \in J$:

$$\begin{tikzcd} & X_j \arrow[dl, "i_j"'] \arrow[dr, "k_j"] & \\ X \arrow[rr, "f"'] & & Y \end{tikzcd} . \tag{1.5.4}$$

By the universal property of X together with the canonical injections i_j there exists a morphisms f making the diagram 1.5.4 commute. Likewise, by the universal property of Y together with the canonical injections k_j there exists a morphisms g such that the following diagram commutes:

$$\begin{tikzcd} & X_j \arrow[dl, "k_j"'] \arrow[dr, "i_j"] & \\ Y \arrow[rr, "g"'] & & X \end{tikzcd} . \tag{1.5.5}$$

Combining Diagrams 1.5.4 and 1.5.5 gives rise to the following commutative diagram:

$$\begin{tikzcd} & X_j \arrow[dl, "i_j"'] \arrow[dr, "i_j"] & \\ Y \arrow[rr, "g\circ f"'] & & X \end{tikzcd} . \tag{1.5.6}$$

Since replacing $g\circ f$ by the identity morphism id_X in this diagram results in another commutative diagram, the universal property of the coproduct X entails $g \circ f = \mathrm{id}_X$. By switching f and g one obtains $f \circ g = \mathrm{id}_Y$, and the claim follows. $\square$

1.5.12 Examples (a) The coproduct in the category of sets is the disjoint union of sets. More precisely, let $(X_i)_{j\in J}$ be a family of sets. Recall that the *disjoint union* of $(X_i)_{j\in J}$ is defined to be the set

$$\bigsqcup_{j\in J} X_j := \bigcup_{j\in J} X_j \times \{j\} = \Big\{(x,j) \in \Big(\bigcup_{j\in J} X_j\Big) \times J \mid x \in X_j\Big\} .$$

Define the canonical injections $i_j : X_j \to \bigsqcup_{i\in J} X_i$ by $i_j(x) = (x,j)$ for $x \in X_j$. Note that the i_j are injective maps when the underlying category is Ens, but that need not hold for other categories. By construction, the image of i_j coincides with $X_j \times \{j\}$. Moreover $\mathrm{Im}(i_j) \cap \mathrm{Im}(i_{j'}) = \varnothing$ if $j \neq j'$, so the name *disjoint union* for the set $\bigsqcup_{j\in J} X_j$ is justified. Now, if $(g_j)_{j\in J}$ is a family of maps $g_j : X_j \to Y$, then define $g : \bigsqcup_{j\in J} X_j \to Y$ by $g(x,j) := g_j(x)$ for $j \in J$ and $x \in X_j$. By construction, $g_j = g \circ i_J$ for every $j \in J$. Since $\bigsqcup_{j\in J} X_j$ is the union of the images of the canonical injections, g is uniquely determined. The claim folows.

(b) The coproduct in the category of groups is given by the *free product* of groups, see **??**. In the category of abelian groups, though, the coproduct of a family $(A_j)_{j\in J}$ of abelian groups coincides with the direct sum $A := \bigoplus_{j\in J} A_j$, see **??**.

The product and coproduct are, if they exist, functorial in the following sense.

1.5.13 Proposition *Let $(X_j)_{j\in J}$ and $(Y_j)_{j\in J}$ be families of objects in a category C, and $(f_j)_{j\in J}$ a family of morphisms $f_j : X_j \to Y_j$.*

(i) *If the products of both $(X_j)_{j\in J}$ and $(Y_j)_{j\in J}$ exist in C, then there is a unique morphism $f : \prod_{j\in J} X_j \to \prod_{j\in J} Y_j$, which usually will be denoted $\prod_{j\in J} f_j$, such that the diagram*

$$\begin{tikzcd} \prod_{i\in J} X_i \arrow[r, "f"] \arrow[d, "p_j"'] & \prod_{i\in J} Y_i \arrow[d, "q_j"] \\ X_j \arrow[r, "f_j"'] & Y_j \end{tikzcd}$$

commutes for every $j \in J$. Hereby, the p_j and q_j are the canonical projections of $\prod_{j\in J} X_j$ and $\prod_{j\in J} Y_j$, respectively. If $(Z_j)_{j\in J}$ is a third family of objects in C for which the product $\prod_{j\in J} Z_j$ exists and if $(g_j)_{j\in J}$ is a family of morphisms $g_j : Y_j \to Z_j$, then

$$\prod_{j\in J} (g_j \circ f_j) = \Big(\prod_{j\in J} g_j\Big) \circ \Big(\prod_{j\in J} f_j\Big) .$$

Moreover,

$$\prod_{j\in J} \mathrm{id}_{X_j} = \mathrm{id}_{\prod_{j\in J} X_j} .$$

(ii) *If the coproducts of both $(X_j)_{j\in J}$ and $(Y_j)_{j\in J}$ exist in C, then there is a unique morphism $f : \coprod_{j\in J} X_j \to \coprod_{j\in J} Y_j$, which usually will be denoted $\coprod_{j\in J} f_j$, such that the diagram*

$$\begin{tikzcd} X_j \arrow[r, "f_j"] \arrow[d, "i_j"] & Y_j \arrow[d, "k_j"] \\ \coprod_{i\in J} X_i \arrow[r, "f"'] & \coprod_{i\in J} Y_i \end{tikzcd}$$

commutes for every $j \in J$. Hereby, the i_j and k_j are the canonical injections of $\coprod_{j\in J} X_j$ and $\coprod_{j\in J} Y_j$, respectively. If $(Z_j)_{j\in J}$ is a third family of objects in C for which the coproduct $\coprod_{j\in J} Z_j$ exists and if $(g_j)_{j\in J}$ is a family of morphisms $g_j : Y_j \to Z_j$, then

$$\coprod_{j\in J} (g_j \circ f_j) = \Big(\coprod_{j\in J} g_j\Big) \circ \Big(\coprod_{j\in J} f_j\Big) .$$

Finally,

$$\coprod_{j\in J} \mathrm{id}_{X_j} = \mathrm{id}_{\coprod_{j\in J} X_j} .$$

Proof. We only show the claim in the product case; the proof in the coproduct case is completely dual.

The existence of $f : \prod_{j \in J} X_j \to \prod_{j \in J} Y_j$ follows by the universal property of the product $\prod_{j \in J} Y_j$. Let $r_j : Z \to Z_j$, $j \in J$ be the canonical projections of the product $Z := \prod_{j \in J} Z_j$. Then, for all $j \in J$,

$$r_j \circ \prod_{i \in J} (g_i \circ f_i) = g_j \circ f_j \circ p_j = g_j \circ q_j \circ \Big(\prod_{i \in J} f_i\Big) = r_j \circ \Big(\prod_{i \in J} g_i\Big) \circ \Big(\prod_{i \in J} f_i\Big) ,$$

which entails

$$\prod_{j \in J} (g_j \circ f_j) = \Big(\prod_{j \in J} g_j\Big) \circ \Big(\prod_{j \in J} f_j\Big)$$

by the universal property of the product Z. Denoting by X the product $\prod_{j \in J} X_j$, the equality

$$\prod_{j \in J} \mathrm{id}_{X_j} = \mathrm{id}_X$$

holds true, since $p_j \circ \mathrm{id}_X = \mathrm{id}_{X_j} \circ p_j$ for all $j \in J$. So the claim is proved. ☐

We shall, in this section, further investigate the notion of "universality".

Initial and terminal objects

We now introduce another example of universality, which is simpler but also more abstract than the products and coproducts introduced in the previous section.

1.5.14 Definition Let C be a category. An *initial object* in C is an object X of C with the property that $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ has exactly one element for every object Y of C.

1.5.15 Remark So there is a unique morphism out of an initial object X of C into an object Y of that category. Note that this idea is faithful to the categorical spirit of describing objects in terms of their mapping properties.

1.5.16 Examples (a) If C is the category of sets, then the empty set $\varnothing$ is an initial object. There is a unique map from the empty set into any other set X; one has to make no decisions about where elements are to map from when constructing a map $\varnothing \to X$. The resulting map is unique and is the empty map $\varnothing \to X$ with domain $\varnothing$, range X, and graph $\varnothing = \varnothing \times X$.

(b) In the category of groups, the group $\{1\}$ consisting of one element, namely the neutral element, is an initial object.

Note that the initial object in Grp is not that in Ens. This should not be too surprising, because $\varnothing$ cannot be a group.

(c) Let P be a poset, and make it into a category as in Example 1.2.11. Then it is easy to see that an initial object of P is the smallest object in P (if it exists). Note that this is equivalently the product of all the objects in P. In general, the initial object of a category is not the product of all objects in C (this does not even make sense for a large category).

1.5.17 There is a dual notion, called a *terminal object*, where every object can map into it in precisely one way.

1.5.18 Definition A *terminal object* in a category C is an object Y of C such that $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ has exactly one element for every object X of C.

1.5.19 Remark Note that an initial object in C is the same as a terminal object in C^{op}, and vice versa. As a result, it suffices to prove results about initial objects, and the corresponding results for terminal objects will follow formally.

But there is a fundamental difference between initial and terminal objects. Initial objects are characterized by how one maps *out of* them, while terminal objects are characterized by how one maps *into* them.

1.5.20 Example Any one point set is a terminal object in Ens.

The important point about the next result is the conceptual framework it entails.

1.5.21 Proposition (Uniqueness of initial and terminal objects) *Any two initial (respectively terminal) objects in a category* C *are isomorphic by a unique isomorphism.*

Proof. The proof is easy. Assume that Y, Y' are both initial or both terminal objects. Then $\mathrm{Mor}(Y, Y')$ and $\mathrm{Mor}(Y', Y)$ are one-point sets. So there are unique morphisms $f : Y \to Y'$, $g : Y' \to Y$, whose composites must be the identities: we know that $\mathrm{Mor}(Y, Y)$ and $\mathrm{Mor}(Y', Y')$ are one-point sets, too, so the composites have no other choice than to be the identities. This means that the maps $f : Y \to Y'$ and $g : Y' \to Y$ are isomorphisms. □

1.5.22 Remark There is a philosophical point to be made here. We have characterized an object uniquely in terms of mapping properties. More precisely, we have characterized it "uniquely up to unique isomorphism", which is really the best one can do in mathematics. Often in a categorical setting one encounters the following situation: two structured sets represent two objects in a category and these objects are isomorphic up to unique isomorphism, but the underlying sets are different.

Note also that the argument was essentially similar to that of Proposition 1.5.5.

In fact, we could interpret Proposition 1.5.5 as a special case of Proposition 1.5.21. If C is a category and $(X_j)_{j \in J}$ is a family of objects in C, then we can define a category D as follows. An object of D is the data of an object $Y \in \mathsf{C}$ and morphisms $f_j : Y \to X_j$ for all $j \in J$. A morphism between objects $\big(Y, (f_j : Y \to X_j)_{j \in J}\big)$ and $\big(Z, (g_j : Z \to X_j)_{j \in J}\big)$ is a map $Y \to Z$ making the obvious diagrams commute. Then a product $\prod_{j \in J} X_j$ in C is the same thing as a terminal object in D, as one easily checks from the definitions.

Pushouts and pullbacks

Now we are going to talk about more examples of universal constructions, which can all be phrased via initial or terminal objects in some category. This, therefore, is the proof for the uniqueness up to unique isomorphism of *everything* we will do in this section. Later we will present these in more generality.

Like always in this chapter C denotes a category.

1.5.23 Definition Suppose we have objects A, B, C, X of C. A commutative square in C

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \\ C \arrow[r] & X \end{tikzcd}$$

is called *cocartesian* or a *pushout square* (and X is called the *pushout* of $A \to B$ and $A \to C$) if it satisfies the following universal property.

(Cocar) Given a commutative diagram

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \\ C \arrow[r] & Y \end{tikzcd}$$

there is a unique map $X \to Y$ making the following diagram commute.

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \arrow[ddr] & \\ C \arrow[r] \arrow[drr] & X \arrow[dr] & \\ & & Y \end{tikzcd}$$

Sometimes pushouts are also called *fibered coproducts*. We shall also write $X = C \sqcup_A B$.

1.5.24 Remark In other words, to map out of $X = C \sqcup_A B$ into some object Y is to give maps $B \to Y$ and $C \to Y$ whose restrictions to A are the same.

The next few examples will rely on notions to be introduced later.

1.5.25 Example The following is a pushout square in the category of abelian groups:

$$\begin{tikzcd} \mathbb{Z}/2 \arrow[r] \arrow[d] & \mathbb{Z}/4 \arrow[d] \\ \mathbb{Z}/6 \arrow[r] & \mathbb{Z}/12 \end{tikzcd}$$

In the category of groups, the pushout is actually $\mathrm{SL}_2(\mathbb{Z})$, though we do not prove it. The point is that the property of a square's being a pushout is actually dependent on the category.

In general, to construct a pushout of abelian groups $C \sqcup_A B$, one constructs the direct sum $C \oplus B$ and quotients by the subgroup generated by (a, a) (where $a \in A$ is identified with its image in $C \oplus B$). We shall discuss this later, more thoroughly, for modules over a ring.

1.5.26 Example Let R be a commutative ring and let S and Q be two commutative R-algebras. In other words, suppose we have two maps of rings $s : R \to S$ and $q : R \to Q$. Then we can fit this information together into a pushout square:

$$\begin{tikzcd} R \arrow[r] \arrow[d] & S \arrow[d] \\ Q \arrow[r] & X \end{tikzcd}$$

It turns out that the pushout in this case is the tensor product of algebras $S \otimes_R Q$ (see Section 4.3 for the construction). This is particularly important in algebraic geometry as the dual construction will give the correct notion of "products" in the category of "schemes" over a field.

1.5.27 Proposition *If the pushout of the diagram*

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \\ C & \end{tikzcd}$$

in C *exists, it is unique up to unique isomorphism.*

Proof. We can prove this in two ways. One is to suppose that there were two pushout squares:

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \\ C \arrow[r] & X \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \\ C \arrow[r] & X' \end{tikzcd}$$

Then there are unique maps $f : X \to X', g : X' \to X$ from the universal property. In detail, these maps fit into commutative diagrams

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \arrow[ddr] & \\ C \arrow[r] \arrow[drr] & X \arrow[dr, "f"] & \\ & & X' \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \arrow[ddr] & \\ C \arrow[r] \arrow[drr] & X' \arrow[dr, "g"] & \\ & & X \end{tikzcd}$$

Then $g \circ f$ and $f \circ g$ are the identities of X, X' again by *uniqueness* of the map in the definition of the pushout.

Alternatively, we can phrase pushouts in terms of initial objects. We could consider the category of all diagrams as above,

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \\ C \arrow[r] & D \end{tikzcd},$$

where $A \to B, A \to C$ are fixed and D varies. The morphisms in this category of diagrams consist of commutative diagrams. Then the initial object in this category is the pushout, as one easily checks. $\square$

1.5.28 Often when studying categorical constructions, one can create a kind of "dual" construction by reversing the direction of the arrows. This is exactly the relationship between the pushout construction and the pullback construction to be described below.

1.5.29 Definition So suppose we have two morphisms $A \to C$ and $B \to C$, forming a diagram

$$\begin{tikzcd} & B \arrow[d] \\ A \arrow[r] & C \end{tikzcd}.$$

A commutative square in C

$$\begin{tikzcd} Y \arrow[r] \arrow[d] & B \arrow[d] \\ A \arrow[r] & C \end{tikzcd}$$

is called *cartesian* or a *pullback square* (and Y is called the *pullback* or *fibered product* of $A \to C$ and $B \to C$) if it satisfies the following universal property:

(Car) Given a commutative diagram

$$\begin{tikzcd} Y' \arrow[r] \arrow[d] & B \arrow[d] \\ A \arrow[r] & C \end{tikzcd}$$

there is a unique map $Y' \to Y$ making the following diagram commute:

$$\begin{tikzcd} Y' \arrow[dr] \arrow[drr] \arrow[ddr] & & \\ & Y \arrow[r] \arrow[d] & B \arrow[d] \\ & A \arrow[r] & C \end{tikzcd}.$$

We shall also write $Y = B \times_C A$.

1.5.30 Example In the category Ens of sets, if we have sets A, B, C with maps $f : A \to C$ and $g : B \to C$, then the fibered product $A \times_C B$ consists of pairs $(a, b) \in A \times B$ such that $f(a) = g(b)$.

The next example requires prerequisites not developed yet and may be omitted without loss of continuity.

1.5.31 Example As said above, the fact that the tensor product of algebras is a pushout in the category of commutative R-algebras allows for the correct notion of the "product" of schemes. We now elaborate on this example: naively one would think that we could pick the underlying space of the product scheme to just be the topological product of two Zariski topologies. However, it is an easy exercise to check that the product of two Zariski topologies in general is not Zariski! This motivates the need for a different concept.

Suppose we have a field k and two k-algebras A and B and let $X = \mathrm{Spec}(A)$and $Y = \mathrm{Spec}(B)$ be the affine k-schemes corresponding to A and B. Consider the following pullback diagram:

$$\begin{tikzcd} X \times_{\mathrm{Spec}(k)} Y \arrow[r] \arrow[d] & X \arrow[d] \\ Y \arrow[r] & \mathrm{Spec}(k) \end{tikzcd}$$

Now, since Spec is a contravariant functor, the arrows in this pullback diagram have been flipped; so in fact, $X \times_{\mathrm{Spec}(k)} Y$ is actually $\mathrm{Spec}(A \otimes_k B)$. This construction is motivated by the following example: let $A = k[x]$ and $B = k[y]$. Then $\mathrm{Spec}(A)$ and $\mathrm{Spec}(B)$ are both affine lines $\mathbb{A}^{\mathrm{id}}_k$ so we want a suitable notion of product that makes the product of $\mathrm{Spec}(A)$ and $\mathrm{Spec}(B)$ the affine plane. The pullback construction is the correct one since $\mathrm{Spec}(A) \times_{\mathrm{Spec}(k)} \mathrm{Spec}(B) = \mathrm{Spec}(A \otimes_k B) = \mathrm{Spec}(k[x, y]) = \mathbb{A}^2_k$.

1.5.32 Remark
to be added: fill in details The notion of "monomorphism" can be detected using only the notions of fibered product and isomorphism. To see this, suppose $i : X \to Y$ is a monomorphism. Show that the diagonal

$$X \to X \times_Y X$$

is an isomorphism. (The diagonal map is such that the two projections to X both give the identity.) Conversely, show that if $i : X \to Y$ is any morphism such that the above diagonal map is an isomorphism, then i is a monomorphism.

Deduce the following consequence: if $F : \mathsf{C} \to \mathsf{D}$ is a functor that commutes with fibered products, then F takes monomorphisms to monomorphisms.

Diagram schemes and diagrams

1.5.33 When defining pushouts the initially given objects and morphisms have an underlying "diagram scheme" which can be depicted as follows.

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \\ \bullet & \end{tikzcd} \tag{1.5.7}$$

As we will explain, not only pushouts but also further "colimits" defined by a universal property such as coproducts or coequalizers have an underyling diagram scheme which, up to isomorphism, uniquely characterizes the pushout, coproduct or coequalizer, respectively. Likewise, diagram schemes underly and characterize universal limit constructions like pullbacks, products and equalizers. Let us now describe precisely what diagram schemes are.

1.5.34 Definition A *diagram scheme* Σ consists of two sets I, A and two maps $s : A \to I$ and $t : A \to I$. The elements of I are called *indices* or *vertices*, the elements of A *arrows*. For an arrow $a \in A$ one calls $s(a)$ the *source* of a and $t(a)$ its *target*. The diagram scheme Σ is called *finite* if both I and A are finite sets. We sometimes denote a diagram scheme as a quadruple $\Sigma = (I, A, s, t)$.

If $\Sigma = (I, A, s, t)$ and $\Omega = (J, B, \sigma, \tau)$ are both diagram schemes, a *morphism of diagram schemes* from Σ to Ω consist of a pair (φ, f) of maps $\varphi : I \to J$ and $f : A \to B$ such that the following two diagrams commute.

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "s"'] & B \arrow[d, "\sigma"] \\ I \arrow[r, "\varphi"'] & J \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "f"] \arrow[d, "t"'] & B \arrow[d, "\tau"] \\ I \arrow[r, "\varphi"'] & J \end{tikzcd}$$

1.5.35 Remark Using language from graph theory and mathematical physics, a diagram scheme is nothing else than a directed pseudograph (Harary, 1969, Chap. 2) or a quiver (Savage, 2006).

1.5.36 Example Finite diagram schemes are often represented by pictures as the one in Fig. (1.5.7) above. Further examples of graphically represented diagram schemes are the following. In parentheses we mention when a particular diagram scheme leads to one of the standard universal objects.

(a) (product and coproduct)

$$\bullet \qquad \bullet \qquad \bullet \qquad \dots \qquad \bullet$$

(b) (equalizer and coequalizer)

$$\bullet \rightrightarrows \bullet$$

(c)

$$\begin{tikzcd} \bullet \arrow[r] & \bullet \arrow[r] \arrow[dr] & \bullet \\ & & \bullet \arrow[u] \end{tikzcd}$$

(d)

$$\begin{tikzcd} \bullet \arrow[loop above] \arrow[r] & \bullet \arrow[loop above] \arrow[r] & \bullet \arrow[loop above] \end{tikzcd}$$

1.5.37 Given a diagram scheme $\Sigma = (I, A, s, t)$ the identity $\mathrm{id}_\Sigma := (\mathrm{id}_I, \mathrm{id}_A)$ is obviously a morphism of diagram schemes. Moreover, if (φ, f) and (ψ, g) are a morphisms of diagram schemes from $\Sigma = (I, A, s, t)$ to $\Omega = (J, B, \sigma, \tau)$ and from $\Omega = (J, B, \sigma, \tau)$ to $\Xi = (K, C, \alpha, \beta)$, respectively, then $(\psi, g) \circ (\varphi, f) := (\psi \circ \varphi, g \circ f)$ is a morphism of diagram schemes from $\Sigma = (I, A, s, t)$ to $\Xi = (K, C, \alpha, \beta)$ as one checks by commutativity of the following two diagrams.

$$
\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "s"'] & B \arrow[r, "g"] \arrow[d, "\sigma"] & C \arrow[d, "\alpha"] \\
I \arrow[r, "\varphi"'] & J \arrow[r, "\psi"'] & K
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "t"'] & B \arrow[r, "g"] \arrow[d, "\tau"] & C \arrow[d, "\beta"] \\
I \arrow[r, "\varphi"'] & J \arrow[r, "\psi"'] & K
\end{tikzcd}
$$

Since composition of functions is associative we therefore obtain a category denoted DSch and called the *category of diagram schemes.*

1.5.38 Example Each small category can be regarded as a diagram scheme. Actually, one has a forgetful functor U from the category Cat of small categories to the category DSch of diagram schemes which associates to each small category C the diagram scheme $U(\mathsf{C})$ having as index set the set of objects of C and as arrow set the set of morphisms. The source and target maps of $U(\mathsf{C})$ are the same as in C.

1.5.39 Definition Let Σ be a diagram scheme and C a category. By a *diagram* in C of *type* or *shape* Σ one understands a map D which assigns to each index $i \in I$ an object $D(i)$ of C and to each arrow $a \in A$ a morphism $D(a) \in \mathrm{Mor}_{\mathsf{C}}\big(D(s(a)), D(t(a))\big)$. We will use the notation $D : \Sigma \to \mathsf{C}$ or $D : I \to \mathsf{C}$ to express that D is a diagram in the category C of type $\Sigma = (I, A, s, t)$.

1.5.40 Remark If the category C is small, a diagram in C of type Σ is the same thing as a morphism of diagram schemes $\Sigma \to U(\mathsf{C})$.

1.5.41 If $D : \Sigma \to \mathsf{C}$ is a diagram defined on a diagram scheme $\Sigma = (I, A, s, t)$ and $F : \mathsf{C} \to \mathsf{D}$ a covariant functor, then the *composition* $F \circ D : \Sigma \to \mathsf{D}$ is defined as the diagram which associates to each index $i \in I$ the object $F(D(i))$ and to each arrow $a \in A$ the morphism $F(D(a))$. So in other words diagrams in C can be left composed with functors defined on C. Similarly, diagrams defined on a diagram scheme Σ can be right composed with morphisms of diagram schemes having Σ as target.

1.5.42 Theorem *Let $\Sigma = (I, A, s, t)$ be a diagram scheme. Then there exists a small category P_Σ together with a morphism of diagram schemes $(\iota, e) : \Sigma \to \mathsf{P}_\Sigma$ such that the following universal property holds true.*

(FCat) *For each category C and diagram $D : \Sigma \to \mathsf{C}$ there exists a unique (covariant) functor $\overline{D} : \mathsf{P}_\Sigma \to \mathsf{C}$ such that the following diagram of diagrams and functors commutes.*

$$
\begin{tikzcd}
\Sigma \arrow[r, "D"] \arrow[d, "(\iota,e)"'] & \mathsf{C} \\
\mathsf{P}_\Sigma \arrow[ur, "\overline{D}"'] &
\end{tikzcd}
$$

Proof. Let i, j be elements of the index set I. A *path* from i to j of *length* $n \in \mathbb{N}$ in the diagram scheme Σ then is defined as an $n+1$-tuple $p = (a_n, \ldots, a_1, i)$ with $a_k \in A$ for all $k \in [1, n]_{\mathbb{N}}$ such that the following conditions hold true:

(i) If $n \geqslant 1$, then $s(a_1) = i$, $s(a_n) = j$ and $s(a_{k+1}) = t(a_k)$ for all $k \in [1, n-1]_{\mathbb{N}}$.

(ii) If $i \neq j$, then there are no paths of length $n = 0$.

(iii) If $i = j$, then there is exactly one path of length $n = 0$, namely $(i) := \{i\}$. It is called the *identity path* at x.

The index i is called the *source* or *origin* of the path p, the index j is *target* or *end*. A path having the same source and target is called *closed* or a *loop*. The small category P_Σ associated to the diagram scheme Σ is now defined as the category having object set I and morphism sets $\mathrm{Mor}_{\mathsf{P}_\Sigma}(i,j)$ for $i,j \in I$ consisting of all paths in Σ from i to j. Composition of paths $(b_m,\ldots,b_1,j) \in \mathrm{Mor}_{\mathsf{P}_\Sigma}(j,k)$ and $(a_n,\ldots,a_1,i) \in \mathrm{Mor}_{\mathsf{P}_\Sigma}(i,j)$ for $i,j,k \in I$ is defined as

$$(b_m,\ldots,b_1,j) \circ (a_n,\ldots,a_1,i) := \begin{cases} (b_m,\ldots,b_1,a_n,\ldots,a_1,i) & \text{if } m,n \geqslant 1, \\ (a_n,\ldots,a_1,i) & \text{if } k = j,\ m = 0 \text{ and } n \geqslant 1, \\ (b_m,\ldots,b_1,j) & \text{if } j = i,\ n = 0 \text{ and } m \geqslant 1, \\ (i) & \text{if } k = j = i,\ m = n = 0. \end{cases}$$

By definition, the identity path at i serves as identity morphism at i. Associativity of the composition $\circ$ is straightforward, so P_Σ is a category indeed. We call it the *category of paths* in Σ.

There is a canonical embedding (ι, e) of Σ into P_Σ which is the identity map on I that is $\iota = \mathrm{id}_I$ and which maps an arrow $a \in A$ to the path $e(a) = \big(a, s(a)\big) \in \mathrm{Mor}_{\mathsf{P}_\Sigma}\big(s(a), t(a)\big)$.

Now let $D : \Sigma \to \mathsf{C}$ be a diagram. We define the *extension* $\overline{D} : \mathsf{P}_\Sigma \to \mathsf{C}$ by putting $\overline{D}(i) := D(i)$ for all $i \in I$ and, for every path $p = (a_n,\ldots,a_1,i)$ in Σ,

$$\overline{D}(p) := \begin{cases} D(a_n) \circ \ldots \circ D(a_n) & \text{if } n \geqslant 1, \\ \mathrm{id}_{D(i)} & \text{if } n = 0. \end{cases}$$

Then $\overline{D}$ is a functor, indeed, since for every path $q = (b_m,\ldots,b_1,j) \in \mathrm{Mor}_{\mathsf{P}_\Sigma}(j,k)$ composable with $p = (a_n,\ldots,a_1,i) \in \mathrm{Mor}_{\mathsf{P}_\Sigma}(i,j)$ the equality

$$\overline{D}(q) \circ \overline{D}(p) = \begin{cases} D(b_m) \circ \ldots \circ D(b_1) \circ D(a_n) \circ \ldots \circ D(a_1) & \text{if } m,n \geqslant 1, \\ D(a_n) \circ \ldots \circ D(a_1) & \text{if } k = j,\ m = 0 \text{ and } n \geqslant 1, \\ D(b_m) \circ \ldots \circ D(b_1) & \text{if } j = i,\ n = 0 \text{ and } m \geqslant 1, \\ \mathrm{id}_{D(i)} & \text{if } k = j = i,\ m = n = 0 \end{cases}$$

holds true and since the the right hand side of this equality coincides with $\overline{D}(q \circ p)$ by definition of $q \circ p$.

Since for every arrow a in Σ the relation $\overline{D}\big(e(a)\big) = \overline{D}\big(a, s(a)\big) = D(a)$ holds true, the diagram in the universal property (FCat) commutes.

It remains to show that $\overline{D}$ is uniquely determined. But that is immediate after one observes that for any functor $\widetilde{D} : \mathsf{P}_\Sigma \to \mathsf{C}$ fulfilling $\widetilde{D} \circ (\iota, e) = D$ the identities $\widetilde{D}\big((i)\big) = \mathrm{id}_{D(i)} = \overline{D}\big((i)\big)$ and $\widetilde{D}(p) = D(a_n) \circ \ldots \circ D(a_1) = \overline{D}(p)$ hold true for every path $p = (a_n,\ldots,a_1,i)$ in Σ of positive length. $\square$

1.5.43 Remark The category P_Σ constructed in the theorem will be called the *path category* or the *free category* generated by the diagram scheme Σ. By its universal property, diagrams in C of type Σ are in bijective correspondence with functors from the path category P_Σ to C. One therefore generalizes the notion of a diagram in C as follows.

1.5.44 Definition Let J be a small category and C an arbitrary category. By a *diagram* in C of *type* or *shape* J one understands a (covariant) functor $D : \mathsf{J} \to \mathsf{C}$. The category J is also termed the *index category* of the diagram D.

Colimits

1.5.45 We now want to generalize the pushout. Start with a small category J which we will regard as our index category. Recall that *smallness* means that the objects of J form a set. Initially, one is supposed to picture J as the path category generated by something like the diagram scheme

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \\ \bullet & \end{tikzcd}$$

or the diagram scheme:

$$\begin{tikzcd} \bullet \arrow[r, shift left] \arrow[r, shift right] & \bullet \end{tikzcd}$$

Later we will see examples where J can be a more general not necessarily free category. The construction of a *colimit* which we now formulate will work in either case. It will specialize to the pushout when J is generated by the first diagram scheme above.

So we will look at functors

$$F : \mathsf{J} \to \mathsf{C} \ ,$$

which in the case of the three-element diagram scheme will just correspond to diagrams:

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \\ C & \end{tikzcd}$$

We will call a *cocone* on F an object X of C equipped with a family $\varphi = (\varphi_j)_{j \in \mathsf{J}}$ of morphisms $\varphi_j \colon Fj \to X$ such that for all morphisms $i \to j \in \mathsf{J}$ the following diagram commutes.

$$\begin{tikzcd}
Fi \arrow[dd, "F(i\to j)"'] \arrow[dr, "\varphi_i"] & \\
& X \\
Fj \arrow[ur, "\varphi_j"'] &
\end{tikzcd}$$

We will write $\varphi\colon F \to X$, $F \xrightarrow{\varphi} X$ or briefly $F \to X$ to denote that X (together with the family φ) is a cocone on F.

An example would be a cocone on the three-element category above. Then this is just a commutative diagram

$$\begin{tikzcd}
A \arrow[r] \arrow[d] \arrow[dr] & B \arrow[d] \\
C \arrow[r] & X
\end{tikzcd}$$

which can be reduced to a commutative diagram

$$\begin{tikzcd}
A \arrow[r] \arrow[d] & B \arrow[d] \\
C \arrow[r] & X
\end{tikzcd}$$

with the understanding that the diagonal arrow is the composition of two composable outer arrows. There are two such compositions and they result in the same arrow when the diagram commutes.

1.5.46 Definition The *colimit* of the diagram $F : \mathsf{J} \to \mathsf{C}$, written as $\operatorname{colim} F$ or $\operatorname{colim}_I F$ or $\varinjlim_I F$, if it exists, is a cocone $F \xrightarrow{\varphi} X$ with the property that if $F \xrightarrow{\psi} Y$ is any other cocone, then there is a unique morphism $f : X \to Y$ making the diagram

$$\begin{tikzcd}
& X \arrow[dd, "f"] \\
F \arrow[ur, "\varphi"] \arrow[dr, "\psi"'] & \\
& Y
\end{tikzcd}$$

commute. More precisely, this means that for each $j \in \mathsf{J}$ the following diagram commutes.

$$\begin{tikzcd} & X \arrow[dd, "f"] \\ Fj \arrow[ur, "\varphi_j"] \arrow[dr, "\psi_j"'] & \\ & Y \end{tikzcd} \tag{1.5.8}$$

We could form a category $\mathsf{Cocone}(F)$ where the objects are the cocones $F \xrightarrow{\varphi} X$, and the morphisms from $F \xrightarrow{\varphi} X$ to $F \xrightarrow{\psi} Y$ are the maps $f : X \to Y$ that make all the obvious diagrams commute that is all diagrams of the form (1.5.8). In this case, it is easy to see that a *colimit* of the diagram is just an initial object in $\mathsf{Cocone}(F)$.

In any case, we see:

1.5.47 Proposition *The colimit* $\operatorname{colim}_{\mathsf{J}} F$ *of a diagram* $F : \mathsf{J} \to \mathsf{C}$*, if it exists, is unique up to unique isomorphism.*

Motivated by the above remarks which introduced cocones and colimits, Definition 1.5.23 can now be recast as follows.

1.5.48 Definition Let the index category J be generated by the diagram scheme:

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \\ \bullet & \end{tikzcd}$$

Then the colimit of a functor $F : \mathsf{J} \to \mathsf{C}$ is called the *pushout* of F.

Let us go through some further examples. We already looked at pushouts.

1.5.49 Examples (a) Consider the index category J visualized by the following diagram scheme:

$$\bullet \qquad \bullet \qquad \bullet \qquad \bullet$$

So J consists of four objects with no non-identity morphisms. A functor $F : \mathsf{J} \to \mathsf{Ens}$ is just a list of four sets A, B, C, D. The colimit then is the disjoint union $A \sqcup B \sqcup C \sqcup D$. Its universal property is described by the fact that to map out of the disjoint union is the same thing as mapping out of each piece.

(b) Suppose we had the same index category J but the functor F took values in the category of abelian groups. Then F corresponds, again, to a list of four abelian groups. The colimit is the direct sum. The direct sum is characterized by the same universal property.

(c) Still working with the same index category J suppose the functor F took its values in the category of groups. Then the colimit is the free product of the four groups.

(d) Finally assume that C is the category of commutative rings with unit. Then the colimit of a functor $F\colon \mathsf{J} \to \mathsf{C}$ is the tensor product of the four commutative rings.

Coproducts can also be reformulated within the concept of colimits of functors.

1.5.50 Definition When J is the index category with underlying diagram scheme a family of points

$$\bullet \quad \bullet \quad \bullet \quad \dots \quad \bullet \quad \dots$$

with no non-identity morphisms, then the colimit over J is called the *coproduct*.

As explained in Examples 1.5.49, the coproduct means things like direct sums, disjoint unions, and tensor products. If $(A_i)_{i\in I}$ is a family of objects in some category, then we find the universal property of the coproduct can be stated succinctly:

$$\mathrm{Mor}_{\mathsf{C}}\Big(\coprod_{i\in I} A_i, B\Big) = \prod_{i\in I} \mathrm{Mor}_{\mathsf{C}}\left(A_i, B\right) .$$

So the idea of a colimit unifies a whole bunch of constructions. Now let us take a different example.

1.5.51 Example Take J to be the index category with diagram scheme

$$\bullet \rightrightarrows \bullet$$

So a functor $F : \mathsf{J} \to \mathsf{Ens}$ is a diagram

$$A \rightrightarrows B .$$

Call the two arrows $f, g : A \to B$. To get the colimit, we take B and mod out by the equivalence relation generated by $f(a) \sim g(a)$. To map out of this is the same thing as mapping out of B such that the pullbacks to A are the same.

1.5.52 Definition If the index category J is generated by the diagram scheme

$$\bullet \rightrightarrows \bullet$$

the colimit of a functor $F : \mathsf{J} \to \mathsf{C}$ is called the *coequalizer* of F.

1.5.53 Theorem *If the category C has all coproducts and coequalizers, then it has all colimits.*

Proof. Let $F : I \to \mathsf{C}$ be a functor, where I is a small category. We need to obtain an object X with morphisms

$$Fi \to X, \quad i \in I$$

such that for each $f : i \to i'$, the diagram below commutes:

$$\begin{tikzcd} Fi \arrow[r] \arrow[d] & Fi' \arrow[dl] \\ X & \end{tikzcd}$$

and such that X is universal among such diagrams.

To give such a diagram, however, is equivalent to giving a collection of maps

$$Fi \to X$$

that satisfy some conditions. So X should be thought of as a quotient of the coproduct $\sqcup_i Fi$. Let us consider the coproduct $\sqcup_{i \in I, f} Fi$, where f ranges over all morphisms in the category I that start from i. We construct two maps

$$\sqcup_f Fi \rightrightarrows \sqcup_f Fi,$$

whose coequalizer will be that of F. The first map is the identity. The second map sends a factor ∎

Limits

As in the example with pullbacks and pushouts and products and coproducts, one can define a limit by using the exact same universal property above just with all the arrows reversed. Let us explain this in some more detail.

1.5.54 Example The product is an example of a limit where the indexing category is a small category I with no morphisms other than the identity. This example shows the power of universal constructions; by looking at colimits and limits, a whole variety of seemingly unrelated mathematical constructions are shown to be in the same spirit.

Filtered colimits

Filtered colimits are colimits over special indexing categories I which look like totally ordered sets. These have several convenient properties as compared to general colimits. For instance, in the category of *modules* over a ring (to be studied in **??**), we shall see that filtered colimits actually preserve injections and surjections. In fact, they are *exact*. This is not true in more general categories which are similarly structured.

1.5.55 Definition An indexing category is *filtered* if the following hold:

1. Given $i_0, i_1 \in I$, there is a third object $i \in I$ such that both i_0, i_1 map into i. So there is a diagram
$$\begin{tikzcd} i_0 \arrow[dr] & \\ & i \\ i_1 \arrow[ur] & \end{tikzcd}.$$

2. Given any two maps $i_0 \rightrightarrows i_1$, there exists i and $i_1 \to i$ such that the two maps $i_0 \rightrightarrows i$ are equal: intuitively, any two ways of pushing an object into another can be made into the same eventually.

1.5.56 Example If I is the category
$$* \to * \to * \to \dots,$$
i.e. the category generated by the poset $\mathbb{Z}_{\geqslant 0}$, then that is filtered.

1.5.57 Example If G is a torsion-free abelian group, the category I of finitely generated subgroups of G and inclusion maps is filtered. We don't actually need the lack of torsion.

1.5.58 Definition Colimits over a filtered category are called *filtered colimits.*

1.5.59 Example Any torsion-free abelian group is the filtered colimit of its finitely generated subgroups, which are free abelian groups.

This gives a simple approach for showing that a torsion-free abelian group is flat.

1.5.60 Proposition *If I is filtered[1] and $\mathsf{C} = \mathsf{Ens}, \mathsf{Ab}, \mathsf{Grp}$, etc., and $F : I \to \mathsf{C}$ is a functor, then $\mathrm{colim}_I F$ exists and is given by the disjoint union of $F_i, i \in I$ modulo the relation $x \in F_i$ is equivalent to $x' \in F_{i'}$ if x maps to x' under $F_i \to F_{i'}$. This is already an equivalence relation.*

The fact that the relation given above is transitive uses the filtering of the indexing set. Otherwise, we would need to use the relation generated by it.

1.5.61 Example Take $\mathbb{Q}$. This is the filtered colimit of the free submodules $\mathbb{Z}(1/n)$.

Alternatively, choose a sequence of numbers $m_1, m_2, \dots,$ such that for all p, n, we have $p^n \mid m_i$ for $i \gg 0$. Then we have a sequence of maps
$$\mathbb{Z} \stackrel{m_1}{\to} \mathbb{Z} \stackrel{m_2}{\to} \mathbb{Z} \to \dots.$$
The colimit of this is $\mathbb{Q}$. There is a quick way of seeing this, which is left to the reader.

[1]Some people say filtering.

When we have a functor $F : I \to \mathsf{Ens}, \mathsf{Grp}, R - \mathsf{Mod}$ taking values in a "nice" category (e.g. the category of sets, (left-) modules over a ring R, etc.), one can construct the colimit by taking the union of the $F_i, i \in I$ and quotienting by the equivalence relation $x \in F_i \sim x' \in F_{i'}$ if $f : i \to i'$ sends x into x'. This is already an equivalence relation, as one can check.

Another way of saying this is that we have the disjoint union of the F_i modulo the relation that $a \in F_i$ and $b \in F_{i'}$ are equivalent if and only if there is a later i'' with maps $i \to i'', i' \to i''$ such that a, b both map to the same thing in $F_{i''}$.

One of the key properties of filtered colimits is that, in "nice" categories they commute with finite limits.

1.5.62 Proposition *In the category of sets, filtered colimits and finite limits commute with each other.*

The reason this result is so important is that, as we shall see, it will imply that in categories such as the category of R-modules, filtered colimits preserve *exactness*.

Proof. Let us show that filtered colimits commute with (finite) products in the category of sets. The case of an equalizer is similar, and finite limits can be generated from products and equalizers.

So let I be a filtered category, and $\{A_i\}_{i\in I}, \{B_i\}_{i\in I}$ be functors from $I \to \mathsf{Ens}$. We want to show that

$$\varinjlim_I (A_i \times B_i) = \varinjlim_I A_i \times \varinjlim_I B_i.$$

To do this, note first that there is a map in the direction $\to$ because of the natural maps $\varinjlim_I (A_i \times B_i) \to \varinjlim_I A_i$ and $\varinjlim_I (A_i \times B_i) \to \varinjlim_I B_i$. We want to show that this is an isomorphism.

Now we can write the left side as the disjoint union $\bigsqcup_I (A_i \times B_i)$ modulo the equivalence relation that (a_i, b_i) is related to (a_j, b_j) if there exist morphisms $i \to k, j \to k$ sending $(a_i, b_i), (a_j, b_j)$ to the same object in $A_k \times B_k$. For the left side, we have to work with pairs: that is, an element of $\varinjlim_I A_i \times \varinjlim_I B_i$ consists of a pair (a_{i_1}, b_{i_2}) with two pairs $(a_{i_1}, b_{i_2}), (a_{j_1}, b_{j_2})$ equivalent if there exist morphisms $i_1, j_1 \to k_1$ and $i_2, j_2 \to k_2$ such that both have the same image in $A_{k_1} \times A_{k_2}$. It is easy to see that these amount to the same thing, because of the filtering condition: we can always modify an element of $A_i \times B_j$ to some $A_k \times B_k$ for k receiving maps from i, j. ∎

1.5.63 Remark Let A be an abelian group, $e : A \to A$ an *idempotent* operator, i.e. one such that $e^2 = e$. Show that eA can be obtained as the filtered colimit of

$$A \xrightarrow{e} A \xrightarrow{e} A \dots.$$

The initial object theorem

We now prove a fairly nontrivial result, due to Freyd. This gives a sufficient condition for the existence of initial objects. We shall use it in proving the adjoint functor theorem below.

Let C be a category. Then we recall that $A \in \mathsf{C}$ if for each $X \in \mathsf{C}$, there is a *unique* $A \to X$. Let us consider the weaker condition that for each $X \in \mathsf{C}$, there exists *a* map $A \to X$.

1.5.64 Definition Suppose C has equalizers. If $A \in \mathsf{C}$ is such that $\mathrm{Mor}_{\mathsf{C}}(A, X) \neq \varnothing$ for each $X \in \mathsf{C}$, then X is called *weakly initial.*

We now want to get an initial object from a weakly initial object. To do this, note first that if A is weakly initial and B is any object with a morphism $B \to A$, then B is weakly initial too. So we are going to take our initial object to be a very small subobject of A. It is going to be so small as to guarantee the uniqueness condition of an initial object. To make it small, we equalize all endomorphisms.

1.5.65 Proposition *If A is a weakly initial object in C, then the equalizer of all endomorphisms $A \to A$ is initial for C.*

Proof. Let A' be this equalizer; it is endowed with a morphism $A' \to A$. Then let us recall what this means. For any two endomorphisms $A \rightrightarrows A$, the two pullbacks $A' \rightrightarrows A$ are equal. Moreover, if $B \to A$ is a morphism that has this property, then B factors uniquely through A'.

Now $A' \to A$ is a morphism, so by the remarks above, A' is weakly initial: to each $X \in \mathsf{C}$, there exists a morphism $A' \to X$. However, we need to show that it is unique.

So suppose given two maps $f, g : A' \rightrightarrows X$. We are going to show that they are equal. If not, consider their equalizer O. Then we have a morphism $O \to A'$ such that the post-compositions with f, g are equal. But by weak initialness, there is a map $A \to O$; thus we get a composite

$$A \to O \to A'.$$

We claim that this is a *section* of the embedding $A' \to A$. This will prove the result. Indeed, we will have constructed a section $A \to A'$, and since it factors through O, the two maps

$$A \to O \to A' \rightrightarrows X$$

are equal. Thus, composing each of these with the inclusion $A' \to A$ shows that f, g were equal in the first place.

Thus we are reduced to proving:

1.5.66 Lemma *Let A be an object of a category C. Let A' be the equalizer of all endomorphisms of A. Then any morphism $A \to A'$ is a section of the inclusion $A' \to A$.*

Proof. Consider the canonical inclusion $i : A' \to A$. We are given some map $s : A \to A'$; we must show that $si = \mathrm{id}_{A'}$. Indeed, consider the composition

$$A' \xrightarrow{i} A \xrightarrow{s} A' \xrightarrow{i} A.$$

Now i equalizes endomorphisms of A; in particular, this composition is the same as

$$A' \xrightarrow{i} A \xrightarrow{\mathrm{id}} A; \qquad \square$$

that is, it equals i. So the map $si : A' \to A$ has the property that $isi = i$ as maps $A' \to A$. But i being a monomorphism, it follows that $si = \mathrm{id}_{A'}$. $\square$

1.5.67 Theorem (Freyd) *Let* C *be a category admitting all small limits.*[2] *Then* C *has an initial object if and only if the following* solution set condition holds: *there is a set* $\{X_i, i \in I\}$ *of objects in* C *such that any* $X \in \mathsf{C}$ *can be mapped into by one of these.*

The idea is that the family $\{X_i\}$ is somehow weakly universal *together.*

Proof. If C has an initial object, we may just consider that as the family $\{X_i\}$: we can hom out (uniquely!) from a universal object into anything, or in other words a universal object is weakly universal.

Suppose we have a "weakly universal family" $\{X_i\}$. Then the product $\prod X_i$ is weakly universal. Indeed, if $X \in \mathsf{C}$, choose some i' and a morphism $X_{i'} \to X$ by the hypothesis. Then this map composed with the projection from the product gives a map $\prod X_i \to X_{i'} \to X$. Proposition 1.5.65 now implies that C has an initial object. □

Completeness and cocompleteness

1.5.68 Definition A category C is said to be *complete* if for every functor $F : I \to \mathsf{C}$ where I is a small category, the limit $\lim F$ exists (i.e. C has all small limits). If all colimits exist, then C is said to be *cocomplete.*

If a category is complete, various nice properties hold.

1.5.69 Proposition *If* C *is a complete category, the following conditions are true:*

1. *all (finite) products exist*
2. *all pullbacks exist*
3. *there is a terminal object*

Proof. The proof of the first two properties is trivial since they can all be expressed as limits; for the proof of the existence of a terminal object, consider the empty diagram $F : \varnothing \to \mathsf{C}$. Then the terminal object is just $\lim F$. □

Of course, if one dualizes everything we get a theorem about cocomplete categories which is proved in essentially the same manner. More is true however; it turns out that finite (co)completeness are equivalent to the properties above if one requires the finiteness condition for the existence of (co)products.

[2] We shall later call such a category *complete*.

Continuous and cocontinuous functors

1.6. Yoneda's lemma

add this section is barely fleshed out

Let C be a category. In general, we have said that there is no way to study an object in a category other than by considering maps into and out of it. We will see that essentially everything about $X \in \mathsf{C}$ can be recovered from these hom-sets. We will thus get an embedding of C into a category of functors.

The functors h_X

We now use the structure of a category to construct hom functors.

1.6.1 Definition Let $X \in \mathsf{C}$. We define the contravariant functor $h_X : \mathsf{C} \to \mathsf{Ens}$ via

$$h_X(Y) = \mathrm{Mor}_{\mathsf{C}}(Y, X).$$

This is, indeed, a functor. If $g : Y \to Y'$, then precomposition gives a map of sets

$$h_X(Y') \to h_X(Y), \quad f \mapsto f \circ g$$

which satisfies all the usual identities.

As a functor, h_X encodes *all* the information about how one can map into X. It turns out that one can basically recover X from h_X, though.

The Yoneda lemma

Let $X \xrightarrow{f} X'$ be a morphism in C. Then for each $Y \in \mathsf{C}$, composition gives a map

$$\mathrm{Mor}_{\mathsf{C}}(Y, X) \to \mathrm{Mor}_{\mathsf{C}}(Y, X').$$

It is easy to see that this induces a *natural* transformation

$$h_X \to h_{X'}.$$

Thus we get a map of sets

$$\mathrm{Mor}_{\mathsf{C}}(X, X') \to \mathrm{Mor}(h_X, h_{X'}),$$

where $h_X, h_{X'}$ lie in the category of contravariant functors $\mathsf{C} \to \mathsf{Ens}$. In other words, we have defined a *covariant functor*

$$\mathsf{C} \to \mathsf{Fun}(\mathsf{C}^{\mathrm{op}}, \mathsf{Ens}).$$

This is called the *Yoneda embedding.* The next result states that the embedding is fully faithful.

1.6.2 Theorem (Yoneda's lemma) *If $X, X' \in \mathsf{C}$, then the map $\mathrm{Mor}_{\mathsf{C}}(X, X') \to \mathrm{Mor}(h_X, h_{X'})$ is a bijection. That is, every natural transformation $h_X \to h_{X'}$ arises in one and only one way from a morphism $X \to X'$.*

1.6.3 Theorem (Strong Yoneda lemma)

Representable functors

We use the same notation of the preceding secction for a category C and $X \in \mathsf{C}$, we let h_X be the contravariant functor $\mathsf{C} \to \mathsf{Ens}$ given by $Y \mapsto \mathrm{Mor}_{\mathsf{C}}(Y, X)$.

1.6.4 Definition A contravariant functor $F : \mathsf{C} \to \mathsf{Ens}$ is *representable* if it is naturally isomorphic to some h_X.

The point of a representable functor is that it can be realized as maps into a specific object. In fact, let us look at a specific feature of the functor h_X. Consider the object $\alpha \in h_X(X)$ that corresponds to the identity. Then any morphism

$$Y \to X$$

factors *uniquely* as

$$Y \to X \xrightarrow{\alpha} X$$

(this is completely trivial!) so that any element of $h_X(Y)$ is a $f^*(\alpha)$ for precisely one $f : Y \to X$.

1.6.5 Definition Let $F : \mathsf{C} \to \mathsf{Ens}$ be a contravariant functor. A *universal object* for C is a pair (X, α) where $X \in \mathsf{C}, \alpha \in F(X)$ such that the following condition holds: if Y is any object and $\beta \in F(Y)$, then there is a unique $f : Y \to X$ such that α pulls back to β under f.

In other words, $\beta = f^*(\alpha)$.

So a functor has a universal object if and only if it is representable. Indeed, we just say that the identity $X \to X$ is universal for h_X, and conversely if F has a universal object (X, α), then F is naturally isomorphic to h_X (the isomorphism $h_X \simeq F$ being given by pulling back α appropriately).

The article **?** by Vistoli contains a good introduction to and several examples of this theory. Here is one of them:

1.6.6 Example Consider the contravariant functor $F : \mathsf{Ens} \to \mathsf{Ens}$ that sends any set S to its power set $\mathcal{P}(S)$ (i.e. its collection of subsets). This is a contravariant functor: if $f : S \to T$, there is a morphism

$$\mathcal{P}(T) \to \mathcal{P}(S), \quad T' \mapsto f^{-1}(T').$$

This functor is representable. Indeed, the universal object can be taken as the pair

$$(\{0, 1\}, \{1\}).$$

To understand this, note that a subset S; of S determines its *characteristic function* $\chi_{S'} : S \to \{0, 1\}$ that takes the value 1 on S and 0 elsewhere. If we consider $\chi_{S'}$ as a morphism $S \to \{0, 1\}$, we see that

$$S' = \chi_{S'}^{-1}(\{1\}).$$

Moreover, the set of subsets is in natural bijection with the set of characteristic functions, which in turn are precisely *all* the maps $S \to \{0, 1\}$. From this the assertion is clear.

We shall meet some elementary criteria for the representability of contravariant functors in the next subsec. For now, we note[3] that in algebraic topology, one often works with the *homotopy category* of pointed CW complexes (where morphisms are pointed continuous maps modulo homotopy), any contravariant functor that satisfies two relatively mild conditions (a Mayer-Vietoris condition and a condition on coproducts), is automatically representable by a theorem of Brown. In particular, this implies that the singular cohomology functors $H^n(-, G)$ (with coefficients in some group G) are representable; the representing objects are the so-called Eilenberg-MacLane spaces $K(G, n)$. See Hatcher (2002).

Limits as representable functors

add

Criteria for representability

Let C be a category. We saw in the previous subsec that a representable functor must send colimits to limits. We shall now see that there is a converse under certain set-theoretic conditions. For simplicity, we start by stating the result for corepresentable functors.

1.6.7 Theorem ((Co)representability theorem) *Let C be a complete category, and let $F : \mathsf{C} \to \mathsf{Ens}$ be a covariant functor. Suppose F preserves limits and satisfies the solution set condition: there is a set of objects $\{Y_\alpha\}$ such that, for any $X \in \mathsf{C}$ and $x \in F(X)$, there is a morphism*

$$Y_\alpha \to X$$

carrying some element of $F(Y_\alpha)$ onto x.

Then F is corepresentable.

Proof. To F, we associate the following *category* D. An object of D is a pair (x, X) where $x \in F(X)$ and $X \in \mathsf{C}$. A morphism between (x, X) and (y, Y) is a map

$$f : X \to Y$$

that sends x into y (via $F(f) : F(X) \to F(Y)$). It is easy to see that F is corepresentable if and only if there is an initla object in this category; this initial object is the "universal object."

We shall apply the initial object theorem, Theorem 1.5.67. Let us first verify that D is complete; this follows because C is and F preserves limits. So, for instance, the product of (x, X) and (y, Y) is $((x, y), X \times Y)$; here (x, y) is the element of $F(X) \times F(Y) = F(X \times Y)$. The solution set condition states that there is a weakly initial family of objects, and the initial object theorem now implies that there is an initial object. ∎

[3] The reader unfamiliar with algebraic topology may omit these remarks.

1.7. Adjoint functors

1.7.1 According to MacLane, “Adjoint functors arise everywhere.” We shall see several examples of adjoint functors in this book such as Hom and the tensor product. The fact that a functor has an adjoint often immediately implies useful properties about it like that it commutes with either limits or colimits. This will lead, for instance, to a conceptual argument proving right-exactness of the tensor product later on.

For the whole section we suppose that C, D are categories and $F : \mathsf{C} \to \mathsf{D}$, $G : \mathsf{D} \to \mathsf{C}$ two (covariant) functors.

Definition

1.7.2 Definition The functors F, G are *adjoint* if there is a natural isomorphism

$$\alpha_{c,d} : \mathrm{Mor}_{\mathsf{D}}(Fc, d) \xrightarrow{\sim} \mathrm{Mor}_{\mathsf{C}}(c, Gd)$$

whenever c is an object of C and d an object of D. The functor F is said to be the *right adjoint* and G is the *left adjoint.*

Here, naturality means that for every morphism $f : c_1 \to c_2$ in C and every morphism $g : d_1 \to d_2$ in D the square

$$\begin{tikzcd} \mathrm{Mor}_{\mathsf{D}}(Fc_2, d_1) \arrow[r, "\alpha_{c_2,d_1}"] \arrow[d, "\mathrm{Mor}_{\mathsf{D}}(Ff,g)=(Ff)^* g_*"'] & \mathrm{Mor}_{\mathsf{C}}(c_2, Gd_1) \arrow[d, "\mathrm{Mor}_{\mathsf{C}}(f,Gg)=f^*(Gg)_*"] \\ \mathrm{Mor}_{\mathsf{D}}(Fc_1, d_2) \arrow[r, "\alpha_{c_1,d_2}"'] & \mathrm{Mor}_{\mathsf{C}}(c_1, Gd_2) \end{tikzcd} \tag{1.7.1}$$

commutes. In other words this means that α is a natural isomorphism between the two functors $\mathrm{Mor}_{\mathsf{D}}(F-,-), \mathrm{Mor}_{\mathsf{C}}(-,G-) : \mathsf{C}^{\mathrm{op}} \times \mathsf{D} \to \mathsf{Ens}$.

1.7.3 Examples (a) There is a simple pair of adjoint functors between Ens and Ab. Here, the first functor sends a set S to the free abelian group $\mathbb{Z}[S] = \mathbb{Z}^{(S)}$ (see Definition 2.8.1 for a discussion of free modules over arbitrary rings), while the second, U, is the “forgetful” functor that sends an abelian group to its underlying set. Then $\mathbb{Z}[-]$ and U are adjoints. That is, to give a group-homomorphism $\mathbb{Z}^{(S)} \to A$ for some abelian group A is the same as giving a map of sets $S \to A$. This is precisely the defining property of the free abelian group.

(b) In fact, most “free” constructions are just left adjoints. For instance, recall the universal property of the free group $F(S)$ on a set S (see (**?**, I. §12)): to give a group-homomorphism $F(S) \to G$ for G any group is the same as choosing an image in G of each $s \in S$. That is,

$$\mathrm{Mor}_{\mathsf{Grp}}(F(S), G) = \mathrm{Mor}_{\mathsf{Ens}}(S, U(G)).$$

This states that the free functor $S \mapsto F(S)$ is left adjoint to the forgetful functor U from Grp to Ens.

(c) The abelianization functor $G \mapsto G^{\mathrm{ab}} = G/[G,G]$ from $\mathsf{Grp} \to \mathsf{Ab}$ is left adjoint to the inclusion $\mathsf{Ab} \to \mathsf{Grp}$. That is, if G is a group and A an abelian group, there is a natural correspondence between homomorphisms $G \to A$ and $G^{\mathrm{ab}} \to A$. Note that Ab is a subcategory of Grp such that the inclusion admits a left adjoint; in this situation, the subcategory is called *reflective*.

Adjunctions

1.7.4 The fact that two functors are adjoint is encoded by a simple set of algebraic data between them. To see this, suppose $F : \mathsf{C} \to \mathsf{D}$, $G : \mathsf{D} \to \mathsf{C}$ are adjoint functors. For any object c of the category C we know that

$$\mathrm{Mor}_{\mathsf{D}}(Fc, Fc) \xrightarrow{\simeq} \mathrm{Mor}_{\mathsf{C}}(c, GFc),$$

so that the identity morphism $Fc \to Fc$, which is natural in c, corresponds to a map $c \to GFc$ also natural in c. In other words we obtain a natural transformation

$$\eta : \mathrm{id}_{\mathsf{C}} \to GF$$

by mapping the object c to the morphism $\eta_c = \alpha_{c,Fc}(\mathrm{id}_{Fc})$. This assignment is natural indeed since for a morphism $f : c_1 \to c_2$ in C the equality

$$\begin{aligned} GFf \circ \big(\alpha_{c_1,Fc_1}(\mathrm{id}_{Fc_1})\big) &= \big((GFf)_* \circ \alpha_{c_1,Fc_1}\big)(\mathrm{id}_{Fc_1}) = \big(\alpha_{c_1,Fc_2} \circ (Ff)_*\big)(\mathrm{id}_{Fc_1}) = \\ &= \big(\alpha_{c_1,Fc_2} \circ (Ff)^*\big)(\mathrm{id}_{Fc_2}) = \big(f^* \circ \alpha_{c_2,Fc_2}\big)(\mathrm{id}_{Fc_2}) = \big(\alpha_{c_2,Fc_2}(\mathrm{id}_{Fc_2})\big) \circ f \end{aligned}$$

holds true by commutativity of Diagram 1.7.1, hence the square

$$\begin{tikzcd} c_1 \arrow[r, "\eta_{c_1}"] \arrow[d, "f"'] & GFc_1 \arrow[d, "GFf"] \\ c_2 \arrow[r, "\eta_{c_2}"'] & GFc_2 \end{tikzcd}$$

commutes.

Similarly, we get a natural transformation

$$\varepsilon : FG \to \mathrm{id}_{\mathsf{D}}$$

by mapping an object d of the category D to the morphism $\alpha^{-1}_{Gd,d}(\mathrm{id}_{Gd})$. So the morphism $\varepsilon_d : FGd \to d$ corresponds to the identity $Gd \to Gd$ under the adjoint correspondence. Given a morphism $g : d_1 \to d_2$ in D the square

$$\begin{tikzcd} FGd_1 \arrow[r, "\varepsilon_{d_1}"] \arrow[d, "FGg"'] & d_1 \arrow[d, "g"] \\ FGd_2 \arrow[r, "\varepsilon_{d_2}"'] & d_2 \end{tikzcd}$$

commutes since by commutativity of Diagram 1.7.1

$$g \circ \big(\alpha^{-1}_{Gd_1,d_1}(\mathrm{id}_{Gd_1})\big) = \big(g_* \circ \alpha^{-1}_{Gd_1,d_1}\big)(\mathrm{id}_{Gd_1}) = \big(\alpha^{-1}_{Gd_1,d_2} \circ (Gg)_*\big)(\mathrm{id}_{Gd_1}) =$$
$$= \big(\alpha^{-1}_{Gd_1,d_2} \circ (Gg)^*\big)(\mathrm{id}_{Gd_2}) = \big((FGg)^* \circ \alpha^{-1}_{Gd_2,d_2}\big)(\mathrm{id}_{Gd_2}) = \big(\alpha^{-1}_{Gd_2,d_2}(\mathrm{id}_{Gd_2})\big) \circ FGg \ .$$

This proves naturality of ε.

One calls η the *unit* and ε the *counit* of the pair of adjoint functors F, G. The unit and counit are not simply arbitrary. We are, in fact, going to show that they determine the isomorphisms $\alpha_{c,d} : \mathrm{Mor}_{\mathsf{D}}(Fc, d) \xrightarrow{\sim} \mathrm{Mor}_{\mathsf{C}}(c, Gd)$. This will be a little bit of diagram-chasing.

We know that the isomorphism $\mathrm{Mor}_{\mathsf{D}}(Fc, d) \xrightarrow{\sim} \mathrm{Mor}_{\mathsf{C}}(c, Gd)$ is natural. In fact, this is the key point. Let $\phi : Fc \to d$ be any map. Then there is a morphism $(c, Fc) \to (c, d)$ in the product category $\mathsf{C}^{\mathrm{op}} \times \mathsf{D}$; by naturality of the adjoint isomorphism, we get a commutative square of sets

$$\begin{tikzcd}
\mathrm{Mor}_{\mathsf{D}}(Fc, Fc) \arrow[r, "\mathrm{adj}"] \arrow[d, "\phi_*"] & \mathrm{Mor}_{\mathsf{C}}(c, GFc) \arrow[d, "G(\phi)_*"] \\
\mathrm{Mor}_{\mathsf{D}}(Fc, d) \arrow[r, "\mathrm{adj}"] & \mathrm{Mor}_{\mathsf{C}}(c, Gd)
\end{tikzcd}$$

Here the mark adj indicates that the adjoint isomorphism is used. If we start with the identity id_{Fc} and go down and right, we get the map $c \to Gd$ that corresponds under the adjoint correspondence to $Fc \to d$. However, if we go right and down, we get the natural unit map $\eta(c) : c \to GFc$ followed by $G(\phi)$.

Thus, we have a *recipe* for constructing a map $c \to Gd$ given $\phi : Fc \to d$:

1.7.5 Proposition (The unit and counit determines everything) *Let (F, G) be a pair of adjoint functors with unit and counit transformations η, ε.*

Then given $\phi : Fc \to d$, the adjoint map $\psi : c \to Gd$ can be constructed simply as follows. Namely, we start with the unit $\eta(c) : c \to GFc$ and take

$$\psi = G(\phi) \circ \eta(c) : c \to Gd \qquad (1.7.2)$$

(here $G(\phi) : GFc \to Fd$).

In the same way, if we are given $\psi : c \to Gd$ and want to construct a map $\phi : Fc \to d$, we construct

$$\varepsilon(d) \circ F(\psi) : Fc \to FGd \to d. \qquad (1.7.3)$$

In particular, we have seen that the *unit and counit morphisms determine the adjoint isomorphisms.*

Since the adjoint isomorphisms $\mathrm{Mor}_{\mathsf{D}}(Fc, d) \to \mathrm{Mor}_{\mathsf{C}}(c, Gd)$ and $\mathrm{Mor}_{\mathsf{C}}(c, Gd) \to \mathrm{Mor}_{\mathsf{D}}(Fc, d)$ are (by definition) inverse to each other, we can determine conditions on the units and counits.

For instance, the natural transformation $F \circ \eta$ gives a natural transformation $F \circ \eta : F \to FGF$, while the natural transformation $\varepsilon \circ F$ gives a natural transformation $FGF \to F$. (These are slightly different forms of composition!)

1.7.6 Lemma *The composite natural transformation $F \to F$ given by $(\varepsilon \circ F) \circ (F \circ \eta)$ is the identity. Similarly, the composite natural transformation $G \to GFG \to G$ given by $(G \circ \varepsilon) \circ (\eta \circ G)$ is the identity.*

Proof. We prove the first assertion; the second is similar. Given $\phi : Fc \to d$, we know that we must get back to ϕ applying the two constructions above. The first step (going to a map $\psi : c \to Gd$) is by (1.7.2) $\psi = G(\phi) \circ \eta(c)$; the second step sends ψ to $\varepsilon(d) \circ F(\psi)$, by (1.7.3). It follows that

$$\phi = \varepsilon(d) \circ F(G(\phi) \circ \eta(c)) = \varepsilon(d) \circ F(G(\phi)) \circ F(\eta(c)).$$

Now suppose we take $d = Fc$ and $\phi : Fc \to Fc$ to be the identity. We find that $F(G(\phi))$ is the identity $FGFc \to FGFc$, and consequently we find

$$\mathrm{id}_{F(c)} = \varepsilon(Fc) \circ F(\eta(c)).$$

This proves the claim. □

1.7.7 Definition Let $F : \mathsf{C} \to \mathsf{D}, G : \mathsf{D} \to \mathsf{C}$ be covariant functors. An *adjunction* is the data of two natural transformations

$$\eta : 1 \to GF, \quad \varepsilon : FG \to 1,$$

called the *unit* and *counit*, respectively, such that the composites $(\varepsilon \circ F) \circ (F \circ \varepsilon) : F \to F$ and $(G \circ \varepsilon) \circ (\eta \circ G)$ are the identity (that is, the identity natural transformations of F, G).

We have seen that a pair of adjoint functors gives rise to an adjunction. Conversely, an adjunction between F, G ensures that F, G are adjoint, as one may check: one uses the same formulas (1.7.2) and (1.7.3) to define the natural isomorphism.

For any set S, let $F(S)$ be the free group on S. So, for instance, the fact that there is a natural map of sets $S \to F(S)$, for any set S, and a natural map of groups $F(G) \to G$ for any group G, determines the adjunction between the free group functor from Ens to Grp, and the forgetful functor $\mathsf{Grp} \to \mathsf{Ens}$.

As another example, we give a criterion for a functor in an adjunction to be fully faithful.

1.7.8 Proposition *Let F, G be a pair of adjoint functors between categories C, D. Then G is fully faithful if and only if the unit maps $\eta : 1 \to GF$ are isomorphisms.*

Proof. We use the recipe (1.7.2). Namely, we have a map $\mathrm{Mor}_{\mathsf{D}}(Fc, d) \to \mathrm{Mor}_{\mathsf{C}}(c, Gd)$ given by $\phi \mapsto G(\phi) \circ \eta(c)$. This is an isomorphism, since we have an adjunction. As a result, composition with η is an isomorphism of hom-sets if and only if $\phi \mapsto G(\phi)$ is an isomorphism. From this the result is easy to deduce. □

1.7.9 Example For instance, recall that the inclusion functor from Ab to Grp is fully faithful (clear). This is a right adjoint to the abelianization functor $G \mapsto G^{ab}$. As a result, we would expect the unit map of the adjunction to be an isomorphism, by Proposition 1.7.8.

The unit map sends an abelian group to its abelianization: this is obviously an isomorphism, as abelianizing an abelian group does nothing.

Adjoints and (co)limits

One very pleasant property of functors that are left (resp. right) adjoints is that they preserve all colimits (resp. limits).

1.7.10 Proposition *A left adjoint $F : \mathsf{C} \to \mathsf{D}$ preserves colimits. A right adjoint $G : \mathsf{D} \to \mathsf{C}$ preserves limits.*

As an example, the free functor from Ens to Ab is a left adjoint, so it preserves colimits. For instance, it preserves coproducts. This corresponds to the fact that if A_1, A_2 are sets, then $\mathbb{Z}[A_1 \sqcup A_2]$ is naturally isomorphic to $\mathbb{Z}[A_1] \oplus \mathbb{Z}[A_2]$.

Proof. Indeed, this is mostly formal. Let $F : \mathsf{C} \to \mathsf{D}$ be a left adjoint functor, with right adjoint G. Let $f : I \to \mathsf{C}$ be a "diagram" where I is a small category. Suppose $\operatorname{colim}_I f$ exists as an object of C. The result states that $\operatorname{colim}_I F \circ f$ exists as an object of D and can be computed as $F(\operatorname{colim}_I f)$. To see this, we need to show that mapping out of $F(\operatorname{colim}_I f)$ is what we want—that is, mapping out of $F(\operatorname{colim}_I f)$ into some $d \in \mathsf{D}$—amounts to giving compatible $F(f(i)) \to d$ for each $i \in I$. In other words, we need to show that $\operatorname{Mor}_{\mathsf{D}}(F(\operatorname{colim}_I f), d) = \lim_I \operatorname{Mor}_{\mathsf{D}}(F(f(i)), d)$; this is precisely the defining property of the colimit.

But we have

$$\operatorname{Mor}_{\mathsf{D}}(F(\operatorname*{colim}_I f), d) = \operatorname{Mor}_{\mathsf{C}}(\operatorname*{colim}_I f, Gd) = \lim_I \operatorname{Mor}_{\mathsf{C}}(fi, Gd) = \lim_I \operatorname{Mor}_{\mathsf{D}}(F(fi), d),$$

by using adjointness twice. This verifies the claim we wanted. ∎

The idea is that one can easily map *out* of the value of a left adjoint functor, just as one can map out of a colimit.

0.2. Number systems

2.1. Natural numbers

Peano structures

2.1.1 Definition (Peano) A triple $(\mathbb{P}, 0, s)$ consisting of a set $\mathbb{P}$, an element $0 \in \mathbb{P}$ and a map $s : \mathbb{P} \to \mathbb{P}$ is called a *Peano structure*, if the following axioms hold true:

(P1) 0 is not in the image of s.

(P2) s is injective.

(P3) (Induction Axiom) Every inductive subset of $\mathbb{P}$ coincides with $\mathbb{P}$, where by an *inductive subset of* $\mathbb{P}$ one understands a set $I \subset \mathbb{P}$ having the following properties:

- (I1) 0 is an element of I.
- (I2) If $n \in I$, then $s(n) \in I$.

The element 0 is called *zero* or *zero element* of the Peano structure, the map $s : \mathbb{P} \to \mathbb{P}$ the *successor map*.

By Axiom (P1), 0 is not in the image of the successor map. But all other elements of the Peano structure are, as our first result tells.

2.1.2 Proposition *Let* $(\mathbb{P}, 0, s)$ *be a Peano structure. Then the image of* s *coincides with the set* $\mathbb{P}_{\neq 0} := \{n \in \mathbb{P} \mid n \neq 0\}$ *of all non-zero elements, in signs* $s(\mathbb{P}) = \mathbb{P}_{\neq 0}$.

Proof. Put $I := \{0\} \cup s(\mathbb{P})$. We show that I is an inductive set. By definition, $0 \in I$. Assume that $n \in I$. Then $s(n) \in s(\mathbb{P}) \subset I$, so I is an inductive set indeed. By Axiom (P3), I coincides with $\mathbb{P}$, which entails the claim. □

2.1.3 Definition If $(\mathbb{P}, 0, s)$ and $(\mathbb{P}', 0', s')$ are two Peano structures, a *morphism* from $(\mathbb{P}, 0, s)$ to $(\mathbb{P}', 0', s')$ is a map $f : \mathbb{P} \to \mathbb{P}'$ with the following properties:

(P4) $f(0) = 0'$,

(P5) $f \circ s = s' \circ f$.

One denotes such a morphism by $f : (\mathbb{P}, 0, s) \to (\mathbb{P}', 0', s')$.

2.1.4 Theorem *The Peano structures as objects together with their morphisms form a category.*

Proof. For each Peano structure $(\mathbb{P}, 0, s)$ the identity map $\mathrm{id}_{\mathbb{P}}$ is obviously a morphism from $(\mathbb{P}, 0, s)$ to $(\mathbb{P}, 0, s)$. Moreover, if $f : (\mathbb{P}, 0, s) \to (\mathbb{P}', 0', s')$ and $g : (\mathbb{P}', 0', s') \to (\mathbb{P}'', 0'', s'')$ are two morphisms of Peano structures, their composition as mappings $g \circ f$ is a morphism from $(\mathbb{P}, 0, s)$ to $(\mathbb{P}'', 0'', s'')$, because $g \circ f(0) = g(f(0)) = g(0') = 0''$ and $g \circ f \circ s = g \circ s' \circ f = s'' \circ g \circ f$. We denote by $g \circ f : (\mathbb{P}, 0, s) \to (\mathbb{P}'', 0'', s'')$ the resulting morphism and call it the *composition* of $f : (\mathbb{P}, 0, s) \to (\mathbb{P}', 0', s')$ and $g : (\mathbb{P}', 0', s') \to (\mathbb{P}'', 0'', s'')$.

Since the composition of mappings is associative and the identity maps act as neutral elements with respect to composition of mappings, the claim follows. ☐

2.1.5 Theorem (Dedekind's Iteration Theorem, (Dedekind, 1893, Satz 126))
Assume that $(\mathbb{P}, 0, s)$ is a Peano structure. Let X be a set, x_0 a distinguished element of X, and $t : X \to X$ a function. Then there exists a unique function $f : \mathbb{P} \to X$ such that $f(0) = x_0$ and $f \circ s = t \circ f$.

Proof. Our proof follows (Mendelson, 2008, Proof of the Iteration Theorem). We first introduce some new language. We will call a function $g : A \to X$ defined on a subset $A \subset \mathbb{P}$ *admissible*, if it has the following properties:

(i) $0 \in A$ and $g(0) = x_0$.

(ii) For every $n \in \mathbb{P}$ the relation $s(n) \in A$ entails $n \in A$ and $g(s(n)) = t(g(n))$.

If in addition to these properties a given element $n \in \mathbb{P}$ lies in the domain of g, i.e. if $n \in A$, we say that $g : A \to X$ is *n-admissible*. We now prove a series of claims.

Claim 1. If $g : A \to X$ is $s(n)$-admissible, then it is n-admissible.
By assumption, g is $s(n)$-admissible, hence (ii) entails $n \in A$. So g is n-admissible, too.

Claim 2. For each $n \in \mathbb{P}$ there exists an n-admissible function $g : A \to X$.
We show that the set $I \subset \mathbb{P}$ of all $n \in \mathbb{P}$ for which there exists an n-admissible function is inductive. By the Induction Axiom (P3) this will then entail the claim. Obviously, $0 \in I$, since the function $\{0\} \to X$, $0 \mapsto x_0$ is 0-admissible. Now assume that $n \in I$, and let $g : A \to X$ be an admissible function with $n \in A$. We define an $s(n)$-admissible $g^* : A^* \to X$ as follows, where $A^* := A \cup \{s(n)\}$. Restricted to A, the function g^* is defined to be equal to g. If $s(n) \in A$ we are done, and g^* coincides with g. Otherwise $s(n) \notin A$, and we put $g^*(s(n)) := t(g(n))$. In any case, $A^* \subset \mathbb{P}$, $s(n) \in A^*$, and $g^* : A^* \to X$ satisfies (i) and (ii) by construction.

Claim 3. If $g : A \to X$ and $h : B \to X$ are two n-admissible functions, then $g(n) = h(n)$.
Let $I \subset \mathbb{P}$ be the set of all $n \in \mathbb{P}$ such that for all n-admissible functions $g : A \to X$ and $h : B \to X$ the relation $g(n) = h(n)$ holds true. Obviously, $0 \in I$, since any two admissible functions $g : A \to X$ and $h : B \to X$ satisfy $g(0) = x_0 = h(0)$ by (i). Now assume $n \in I$, and let $g : A \to X$ and $h : B \to X$ be two $s(n)$-admissible functions. Since $s(n) \in A \cap B$, one gets $n \in A \cap B$ by (ii), hence g and h are both n-admissible, too. By using (ii) again one concludes $g(s(n)) = t(g(n)) = t(h(n)) = h(s(n))$. Hence $s(n) \in I$, so one obtains $I = \mathbb{P}$ by the Induction Axiom. The claim follows.

Claim 4. There exists an admissible function $f : \mathbb{P} \to X$.
Given $n \in \mathbb{P}$ choose an n-admissible function $g : A \to X$, and put $f(n) := g(n)$. by the previous claim the value $f(n)$ does not depend on the particular choice of an n-admissible g, hence f is

well-defined. Let us show that f is admissible. Obviously, $f(0) = x_0$ since every admissible g satisfies $g(0) = x_0$ by (i). Now let $n \in \mathbb{P}$ and choose an $s(n)$-admissible $g : A \to X$. Then one concludes by (ii) and the definition of f that $n \in A$ and $f(s(n)) = g(s(n)) = t(g(n)) = t(f(n))$. Hence f is admissible.

Claim 5. Any two admissible functions $f_1 : \mathbb{P} \to X$ and $f_2 : \mathbb{P} \to X$ coincide.
Let I be the set of all $n \in \mathbb{P}$ such that $f_1(n) = f_2(n)$. Obviously, $0 \in I$ since $f_1(0) = x_0 = f_2(0)$. Now let $n \in I$, or in other words assume $f_1(n) = f_2(n)$. Then by (ii) $f_1(s(n)) = t(f_1(n)) = t(f_2(n)) = f_2(s(n))$, which means $s(n) \in I$. Thus I is an inductive set, so coincides with $\mathbb{P}$ by the Induction Axiom.

With the verification of *Claim 4.* and *Claim 5.* the proof is finished. □

By the next two results, Peano structures are unique up isomorphism.

2.1.6 Corollary *If $(\mathbb{P}, 0, s)$ and $(\mathbb{P}', 0', s')$ are two Peano structures, there exists a unique morphism $f : (\mathbb{P}, 0, s) \to (\mathbb{P}', 0', s')$.*

Proof. The claim follows immediately from the preceding theorem when putting $X := \mathbb{P}'$, $x_0 := 0'$ and $t := s'$. □

2.1.7 Theorem *Every morphism $f : (\mathbb{P}, 0, s) \to (\mathbb{P}', 0', s')$ between two Peano structures is an isomorphism.*

Proof. Assume that we can show that f is bijective. Then the inverse map $g := f^{-1}$ satisfies $g(0') = 0$ and $g \circ s' = g \circ s' \circ f \circ g = g \circ f \circ s \circ g = s \circ g$, hence is a morphism of Peano structures as well. So it suffices to show that f is bijective.

By Axiom (P3), surjectivity follows when the image of f is an inductive subset of $\mathbb{P}'$. But that holds true, since $0' = f(0)$ is an element of $f(\mathbb{P})$ and since for each element $n' \in \mathbb{P}'$ for which there exists an $n \in \mathbb{P}$ with $n' = f(n)$ the relation $s'(n') = s'(f(n)) = f(s(n)) \in f(\mathbb{P})$ holds true.

Now let K be the set of all $n \in \mathbb{P}$ for which $\{n\} = f^{-1}(f(n))$. We show that this set is inductive as well, which by Axiom (P3) implies that f is injective. First observe that $0 \in K$. Namely, by Proposition 2.1.2, there exists for every non-zero $k \in \mathbb{P}$ an $l \in \mathbb{P}$ with $k = s(l)$, which entails $f(k) = f(s(l)) = s(f(l)) \neq 0'$. Now let $n \in K$. Assume that $k \in \mathbb{P}$ is an element with $f(k) = f(s(n))$. Then $k \neq 0$ by Axiom (P1), because $f(k) = f(s(n)) = s'(f(n)) \neq 0'$. By Proposition 2.1.2 one can therefore find an $m \in \mathbb{P}$ such that $s(m) = k$. By the equality $s'(f(m)) = f(s(m)) = f(k) = f(s(n)) = s'(f(n))$ and Axiom (P2) one concludes $f(m) = f(n)$. By $n \in K$, the equality $m = n$ follows, hence $k = s(m) = s(n)$ and $s(n) \in K$. The proof is finished. □

2.1.8 So far we know that up to isomorphism there is at most one Peano structure. But we do not yet know whether such a structure exists. We will show existence by a construction going back to John von Neumann (1923). To this end recall the axiom of infinity of Zermelo–Fraenkel set theory which says that there exists a set I with $\emptyset \in I$ and $x \cup \{x\} \in I$ for all $x \in I$. We call a set I with these properties an *inductive set*. Fix an inductive set I and denote by $\mathcal{I} \subset \mathcal{P}(I)$ the set of all inductive subsets of I. Now put

$$\mathbb{N} := \bigcap \mathcal{I}, \quad 0 := \varnothing, \quad \text{and let} \quad s : \mathbb{N} \to \mathbb{N},\ n \mapsto n \cup \{n\} \ .$$

Because $\mathbb{N}$ is inductive by the following proposition, the map s is well-defined, indeed. We call the triple $(\mathbb{N}, 0, s)$ the (*set-theoretic* or *von Neumann*) *system of natural numbers*. The elements of $\mathbb{N}$ are called *natural numbers*.

2.1.9 Proposition *The set $\mathbb{N}$ is the smallest inductive set, i.e. $\mathbb{N}$ is inductive and contained in every inductive set.*

Proof. We first prove that the set $\mathbb{N}$ is inductive. Obviously $\varnothing \in \mathbb{N}$, since $\varnothing$ is an element of each inductive subset of I. If n is an element of $\mathbb{N}$, then it lies in each inductive subset of I, which implies that $n \cup \{n\}$ is an element of each inductive subset of I, too, hence $n \cup \{n\} \in \mathbb{N}$. Because $\mathbb{N}$ is inductive, the map s is well-defined.

It remains to show that $\mathbb{N}$ is contained in every inductive set. To verify this, let J be an arbitrary inductive set and I the inductive set used in the definition of $\mathbb{N}$. Then $\varnothing \in J \cap I$. Moreover, if $x \in J \cap I$, then $x \cup \{x\} \in J \cap I$ as well, since both J and I are inductive. By definition of $\mathbb{N}$ the relation $\mathbb{N} \subset J$ follows, hence $\mathbb{N}$ is the smallest inductive set indeed. $\square$

2.1.10 Remarks (a) The proposition entails in particular that the construction of $\mathbb{N}$ does not depend on the initial choice of the inductive set I.

(b) For later purposes it will be useful to denote the set of all non-zero natural numbers by an individual symbol. We will write $\mathbb{N}_{\neq 0}$ or $\mathbb{N}_{>0}$ for that set.

2.1.11 Lemma *Let I be an inductive set, i an element of I, and $n \in \mathbb{N}$. If $i \in n$, then i is an element of $\mathbb{N}$ as well, and $i \subset n$.*

Proof. Let $J := \{n \in \mathbb{N} \mid \forall i \in I : \ i \in n \implies i \in \mathbb{N} \,\&\, i \subset n\}$. We show that J is an inductive set which by Proposition 2.1.9 will entail the claim. Clearly, $\varnothing \in J$, since $\varnothing$ does not have any elements. Assume that $x \in J$, and consider $x \cup \{x\}$. If $i \in I$ and $i \in x \cup \{x\}$, then $i \in x$ or $i = x$. In the latter case, $i \in J \subset \mathbb{N}$ and $i \subset x \cup \{x\}$. In the first case, $i \in \mathbb{N}$ and $i \subset x \subset x \cup \{x\}$ by the inductive assumption $x \in J$. The proof is finished. $\square$

2.1.12 Theorem (cf. von Neumann (1923)) *The system of natural numbers $(\mathbb{N}, 0, s)$ is a Peano structure.*

Proof. By construction, 0 is an element of $\mathbb{N}$ and $s : \mathbb{N} \to \mathbb{N}$ a function. Since $n \in s(n)$ for every element $n \in \mathbb{N}$, 0 is not in the image of s. This gives Axiom (P1). Now assume that $s(n) = s(m)$. Then $m \cup \{m\} = n \cup \{n\}$. This implies that $m \in n \,\&\, n \in m$ holds true or that $m = n$. In the latter case we are done with proving Axiom (P2). In the first case we are done with this as well, since then $m \subset n$ and $n \subset m$ by Lemma 2.1.11. The Induction Axiom (P3) is an immediate consequence of Proposition 2.1.9. $\square$

Addition of natural numbers

2.1.13 Dedekind's iteration theorem allows the definition of addition for the set of natural numbers $\mathbb{N}$. To this end fix some $m \in \mathbb{N}$ and let $\alpha_m : \mathbb{N} \to \mathbb{N}$ be the unique function which satisfes $\alpha_m(0) = m$ and $\alpha_m(s(n)) = s(\alpha_m(n))$ for all $n \in \mathbb{N}$. Using this notation we introduce *addition of natural numbers* as the function

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N},\ (m, n) \mapsto m + n := \alpha_m(n) \ .$$

In the following proposition we state the fundamental properties of addition of natural numbers.

2.1.14 Theorem *The set* $\mathbb{N}$ *of natural numbers together with addition* $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ *and the element* 0 *becomes an* abelian monoid *which means that the following axioms are satisfied:*

(Grp1) *Addition is* associative *that means*

$$(l + m) + n = l + (m + n) \quad \textit{for all } l, m, n \in \mathbb{N} \ .$$

(Grp2) *The element* 0 *is* neutral *with respect to addition which means that*

$$0 + n = n + 0 = n \quad \textit{for all } n \in \mathbb{N} \ .$$

(Grp4) *Addition is* commutative *that means*

$$m + n = n + m \quad \textit{for all } m, n \in \mathbb{N} \ .$$

Proof. We first show that for all $m, n \in \mathbb{N}$

$$\alpha_{s(m)}(n) = \alpha_m(s(n)) \ . \tag{2.1.1}$$

For $n = 0$ this is clear since then both sides are equal to $s(m)$. So assume that $\alpha_{s(m)}(n) = \alpha_m(s(n))$ for some $n \in \mathbb{N}$. Then

$$\alpha_{s(m)}(s(n)) = s(\alpha_{s(m)}(n)) = s(\alpha_m(s(n))) = \alpha_m(s(s(n))).$$

By the Induction Axiom Equation (2.1.1) therefore holds for all $m, n \in \mathbb{N}$.

Next we prove associativity of $+$. To this end we have to show that $\alpha_{\alpha_l(m)}(n) = \alpha_l(\alpha_m(n))$ for all $l, m, n \in \mathbb{N}$. For $m = n = 0$ we have $\alpha_{\alpha_l(0)}(0) = \alpha_l(0) = \alpha_l(\alpha_0(0))$. Now assume that for some $m \in \mathbb{N}$ the relation $\alpha_{\alpha_l(m)}(0) = \alpha_l(\alpha_m(0))$ holds. Then

$$\begin{aligned} \alpha_{\alpha_l(s(m))}(0) &= \alpha_{s(\alpha_l(m))}(0) = s(\alpha_{\alpha_l(m)}(0)) = \\ &= s(\alpha_l(\alpha_m(0))) = \alpha_l(s(\alpha_m(0))) = \alpha_l(\alpha_m(s(0))) = \alpha_l(\alpha_{s(m)}(0)), \end{aligned}$$

where in the last equality we have used Equation (2.1.1). By the Induction Axiom one concludes that $\alpha_{\alpha_l(m)}(0) = \alpha_l(\alpha_m(0))$ for all $l, m \in \mathbb{N}$. Now assume that for some $n \in \mathbb{N}$ and all $l, m \in \mathbb{N}$ the relation $\alpha_{\alpha_l(m)}(n) = \alpha_l(\alpha_m(n))$ holds true. Then

$$\alpha_{\alpha_l(m)}(s(n)) = s(\alpha_{\alpha_l(m)}(n)) = s(\alpha_l(\alpha_m(n))) = \alpha_l(s(\alpha_m(n))) = \alpha_l(\alpha_m(s(n))).$$

By the Induction Axiom associativity of + follows.

Before we verify commutativity let us first show that $\alpha_0(n) = n$ for all $n \in \mathbb{N}$. Together with the the equality $\alpha_n(0) = n$ this will entail that 0 is neutral with respect to addition. By definition $\alpha_0(0) = 0$. So assume that $\alpha_0(n) = n$ for some $n \in \mathbb{N}$. Then $\alpha_0(s(n)) = s(\alpha(n)) = s(n)$, hence $\alpha_0(n) = n$ for all $n \in \mathbb{N}$ by the Induction Axiom.

In particular we have now proved that $\alpha_0(n) = \alpha_n(0)$ for all $n \in \mathbb{N}$. Next assume that $\alpha_m(n) = \alpha_n(m)$ for some $m \in \mathbb{N}$ and all $n \in \mathbb{N}$. Then, using Equation (2.1.1), $\alpha_{s(m)}(n) = \alpha_m(s(n)) = s(\alpha_m(n)) = s(\alpha_n(m)) = \alpha_n(s(m))$, hence commutativity of addition follows by the Induction Axiom.

We have now finished the proof that $(\mathbb{N}, +, 0)$ is an abelian monoid. □

2.1.15 Definition If one is given a triple $(M, *, e)$ where M is a set, $* : M \times M \to M$ a map and $e \in M$ an element such that the above axioms (Grp1) and (Grp2) are fulfilled with $\mathbb{N}$ replaced by M, + replaced by $*$, and 0 by e, then one calls M (together with $*$ and e) a *monoid*. If in addition Axiom (Grp4) holds true (with the same replacements), $(M, *, e)$ is called an *abelian monoid*. The binary operation $* : M \times M \to M$ of a monoid M is sometimes called its *law of composition* or shorter its *composition law*.

2.1.16 Theorem *The abelian monoid* $(\mathbb{N}, +, 0)$ *has the* cancellation property *that means the following holds true:*

(CancL) *Every element* $l \in \mathbb{N}$ *is* left cancellable*, i.e. for all* $m, n \in \mathbb{N}$ *the relation* $l + m = l + n$ *implies* $m = n$.

(CancR) *Every element* $l \in \mathbb{N}$ *is* right cancellable*, i.e. for all* $m, n \in \mathbb{N}$ *the relation* $m + l = n + l$ *implies* $m = n$.

Proof. By commutativity of + it suffices to show (CancR). Obviously, the relation $m + 0 = n + 0$ implies $m = n$. So assume that for some $l \in \mathbb{N}$ one can conclude from $m + l = n + l$ the equality $m = n$. Now assume $m + s(l) = n + s(l)$. Then

$$s(m + l) = s(\alpha_m(l)) = \alpha_m(s(l)) = m + s(l) = n + s(l) = \alpha_n(s(l)) = s(\alpha_n(l)) = s(n + l),$$

which by injectivity of s entails $m + l = n + l$. Hence $m = n$ by inductive hypthesis. By the Induction Axiom every $l \in \mathbb{N}$ now has to be left cancellable. □

2.1.17 Definition A monoid $(M, *, e)$ is said that to have the *left* (respectively *right*) *cancellation property* if (CancL) (respectively (CancR)) is satisfied when replacing + by $*$ and 0 by e. If $(M, *, e)$ has the left and right cancellation property the monoid is said to have the *cancellation property* or is called a *cancellation monoid*. In other words Theorem 2.1.16 tells that $(\mathbb{N}, +, 0)$ is a cancellation monoid.

2.1.18 Proposition *For all* $m \in \mathbb{N}$ *and all* $n \in \mathbb{N}_{>0}$ *one has*

$$m \neq m + n.$$

Proof. If $m = m + n$, then $n = 0$ since $m = m + 0$ and since m is left cancellable. □

2.1.19 Lemma *For all $k, l \in \mathbb{N}$ the relation $k + l = 0$ is equivalent to $k = l = 0$.*

Proof. Recall that a natural number is non-zero if and only if it is in the image of the successor map. So if k or l is non-zero, that number is a successor, hence the sum of k and l is a successor as well by definition of addition. But then $k + l \neq 0$. If $k = l = 0$, then obviously $k + l = 0$. □

2.1.20 Theorem (Trichotomy law for addition) *For all $m, n \in \mathbb{N}$ exactly one of the following statements holds true:*

(i) $m = n + k$ *for some* $k \in \mathbb{N}_{>0}$.

(ii) $m = n$.

(iii) $n = m + l$ *for some* $l \in \mathbb{N}_{>0}$.

Proof. If $m = n$, then neither (i) nor (iii) can hold true by the preceding Proposition 2.1.18. So assume $m \neq n$. If $m = n + k$ and $n = m + l$ for some $k, l \in \mathbb{N}_{>0}$, then $m = m + (k + l)$. Now $k + l$ is non-zero by Lemma 2.1.19 which contradicts Proposition 2.1.18. Hence one can not have $m = n + k$ and $n = m + l$ with non-zero k, l at the same time. Thus we have shown that at most one of the three statements (i), (ii), and (iii) can be true.

Now fix $m \in \mathbb{N}$ and let I be the set of all $n \in \mathbb{N}$ such that one of (i), (ii), or (iii) holds true. We show that I is inductive which will entail the claim. First observe that $n \in I$ for $n = 0$, because if $m = 0$, then $m = n$ holds, otherwise $m = 0 + m = n + m$ with $m \neq 0$. Assume that $n \in I$ for some $n \in \mathbb{N}$. If $m = n$, then $s(n) = n + 1 = m + 1$, so $s(n) \in I$ in this case, too. If $n = m + l$ for some $l \in \mathbb{N}_{>0}$, then $s(n) = n + 1 = (m + l) + 1 = m + (l + 1)$, so $s(n) \in I$ again, since $l + 1 \in \mathbb{N}_{>0}$. Finally, if $m = n + k$ for some $k \in \mathbb{N}_{>0}$, then $k = s(l)$ for some $l \in \mathbb{N}$, and $m = n + s(l) = s(n) + l$ by Equation (2.1.1). For non-zero l this equation entails $s(n) \in I$. If $l = 0$, we have equality of m and $s(n)$, hence $s(n) \in I$ holds then as well. So we have shown that I is inductive, and the theorem is proved. □

Multiplication of natural numbers

2.1.21 Similarly like for addition, we use Dedekind's iteration theorem to define multiplication of natural numbers. Again fix some $m \in \mathbb{N}$ and let $\mu_m : \mathbb{N} \to \mathbb{N}$ be the unique function which satisfes $\mu_m(0) = 0$ and $\mu_m(s(n)) = \alpha_m(\mu_m(n))$ for all $n \in \mathbb{N}$. *Multiplication of natural numbers* is then defined as the function

$$\cdot : \mathbb{N} \times \mathbb{N} \to \mathbb{N}, \ (m, n) \mapsto m \cdot n := \mu_m(n) \ .$$

The fundamental algebaric properties of natural number are expressed in the following result.

2.1.22 Theorem *The set $\mathbb{N}$ of natural numbers together with addition $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, multiplication $\cdot : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and the elements 0 and $1 := s(0)$ becomes a* commutative semiring *that means the following axioms hold true:*

(SRing1) $\mathbb{N}$ *together with addition $+$ and the element 0 is an abelian monoid.*

(Ring2) $\mathbb{N}$ *together with multiplication $\cdot$ and the element 1 is a monoid.*

(SRing2a) *Multiplication by* 0 annihilates $\mathbb{N}$ *that is*

$$0 \cdot n = n \cdot 0 = 0 \quad \textit{for all } n \in \mathbb{N} \ .$$

(Ring3) *Multiplication* distributes *from the left and the right over addition that means*

$$l \cdot (m+n) = (l \cdot m) + (l \cdot n) \quad \textit{for all } l, m, n \in \mathbb{N}, \textit{ and}$$
$$(m+n) \cdot l = (m \cdot l) + (n \cdot l) \quad \textit{for all } l, m, n \in \mathbb{N}.$$

(Ring4) *Multiplication is* commutative *that is*

$$m \cdot n = n \cdot m \quad \textit{for all } m, n \in \mathbb{N} \ .$$

Proof. By Theorem 2.1.14 Axiom (SRing1) holds true.

Let us show that $0 \cdot m = 0$ for all $m \in \mathbb{N}$. To this end observe first that $0 \cdot 0 = \mu_0(0) = 0$. Assuming that $0 \cdot m = 0$ for some $m \in \mathbb{N}$ we conclude that

$$0 \cdot (s(m)) = \mu_0(s(m)) = \alpha_0(\mu_0(m)) = \mu_0(m) = 0,$$

where we have used that 0 is neutral with respect to addition. By induction, the claimed equality $0 \cdot m = 0$ follows for all $m \in \mathbb{N}$. Since by definition $m \cdot 0 = \mu_m(0) = 0$ for all $m \in \mathbb{N}$, we also have shown Axiom (SRing2a).

Next we verify right distributivity. Obviously $(m+n) \cdot 0 = 0 = (m \cdot 0) + (n \cdot 0)$. Assume that $(m+n) \cdot l = (m \cdot l) + (n \cdot l)$ for some $l \in \mathbb{N}$ and all $m, n \in \mathbb{N}$. Then, by the inductive hypothesis and repeated application of associativity and commutativity of addition,

$$\begin{aligned}
(m+n) \cdot s(l) &= \mu_{\alpha_m(n)}(s(l)) = \alpha_{\alpha_m(n)}(\mu_{\alpha_m(n)}(l)) = (m+n) + ((m+n) \cdot l) = \\
&= (m+n) + ((m \cdot l) + (n \cdot l)) = ((m+n) + (m \cdot l)) + (n \cdot l) = \\
&= (m + (n + (m \cdot l))) + (n \cdot l) = (m + ((m \cdot l) + n)) + (n \cdot l) = \\
&= ((m + (m \cdot l)) + n) + (n \cdot l) = (m + (m \cdot l)) + (n + (n \cdot l)) = \\
&= \alpha_m(\mu_m(l)) + \alpha_n(\mu_n(l)) = \mu_m(s(l)) + \mu_n(s(l)) = (m \cdot s(l)) + (n \cdot s(l)).
\end{aligned}$$

By the Induction Axiom right distributivity follows.

Next observe that $n \cdot 1 = \mu_n(s(0)) = \alpha_n(\mu_n(0)) = n + 0 = n$ for all $n \in \mathbb{N}$, which essentailly says that 1 is right neutral with respect to multiplication. Left multiplicative neutrality of 1 follows by induction on n. By definition of multiplication $1 \cdot 0 = \mu_1(0) = 0$. Under the inductive hypothesis $1 \cdot n = n$ one concludes

$$1 \cdot s(n) = \mu_1(s(n)) = \alpha_1(\mu_1(n)) = \alpha_1(1 \cdot n) = \alpha_1(n) = \alpha_{s(0)}(n) = s(\alpha_0(n)) = s(n) \ ,$$

so 1 is also left neutral with respect to multiplication.

To verify commutativity of $\cdot$ observe that we already proved $m \cdot 0 = 0 = 0 \cdot m$ for all $m \in \mathbb{N}$. Assuming that $m \cdot n = n \cdot m$ for some $n \in \mathbb{N}$ and all $m \in \mathbb{N}$ we conclude, using right distributivity and that 1 is left neutral,

$$\begin{aligned}
m \cdot s(n) &= \mu_m(s(n)) = \alpha_m(\mu_m(n)) = m + (m \cdot n) = m + (n \cdot m) = (1 \cdot m) + (n \cdot m) = \\
&= (1+n) \cdot m = (n+1) \cdot m = \alpha_n(s(0)) \cdot m = s(\alpha_n(0)) \cdot m = s(n) \cdot m.
\end{aligned}$$

By induction, this proves commutativity of multiplication.

Commutativity of multiplication now entails that multiplication also left distributes over addition and that 1 is also left neutral with respect to multiplication.

It remains to show associativity of multiplication. To this end first note that $(l \cdot m) \cdot 0 = 0 = l \cdot 0 = l \cdot (m \cdot 0)$ for all $l, m \in \mathbb{N}$. Assume that $(l \cdot m) \cdot n = l \cdot (m \cdot n)$ for some $n \in \mathbb{N}$ and all $l, m \in \mathbb{N}$. Then

$$\begin{aligned}(l \cdot m) \cdot (s(n)) &= \mu_{\mu_l(m)}(s(n)) = \alpha_{\mu_l(m)}(\mu_{\mu_l(m)}(n)) = (l \cdot m) + ((l \cdot m) \cdot n) = \\ &= (l \cdot m) + (l \cdot (m \cdot n)) = l \cdot (m + (m \cdot n)) = \mu_l(\alpha_m(\mu_m(n))) = \\ &= \mu_l(\mu_m(s(n))) = l \cdot (m \cdot s(n)),\end{aligned}$$

which by induction implies that multiplication is associative.

So all axioms of a semiring have been verified for $\mathbb{N}$, and the proof is finished. □

2.1.23 Definition Assume to be given a quintuple $(R, +, \cdot, 0, 1)$ such that R is a set, $+ : R \times R \to R$ and $\cdot : R \times R \to R$ are maps and $0, 1 \in R$ elements. Then R (together with $+$, $\cdot$, 0 and 1) is called a *semiring* if the above axioms (SRing1), (Ring2), (SRing2a), and (Ring3) are fulfilled with $\mathbb{N}$ replaced by R. If in addition Axiom (Ring4) holds true, the semiring is called *commutative*.

2.1.24 Remark It is generally agreed upon in mathematics, and we follow that here too, that multiplication $\cdot$ in a semiring $(R, +, \cdot, 0, 1)$ takes precedence over addition $+$. This means that for elements $r, s \in R$ an expression of the form $r \cdot s + \ldots$ (respectively of the form $\ldots + r \cdot s$ or $\ldots + r \cdot s + \ldots$) is to be interpreted as an abbreviation for the expression $(r \cdot s) + \ldots$ (respectively for the expression $\ldots + (r \cdot s)$ or $\ldots + (r \cdot s) + \ldots$), where the terms to the left and to the right are left invariant. With this agreement, the left and right distribution laws can now be written more shortly as:

$$q \cdot (r + s) = q \cdot r + q \cdot s \quad \text{and} \quad (r + s) \cdot q = r \cdot q + s \cdot q \quad \text{for all } q, r, s \in R\ .$$

2.1.25 Definition As usual, the first nine non-zero natural numbers are denoted by the following symbols:

$$\begin{aligned}1 := s(0), \quad 2 := s(1), \quad 3 := s(2), \quad 4 := s(3), \quad 5 := s(4), \\ 6 := s(5), \quad 7 := s(6), \quad 8 := s(7), \quad 9 := s(8).\end{aligned}$$

2.1.26 From now one we will avoid using the symbols s, α_n, and μ_m and replace them by the standard notation involving only the addition symbol $+$, the multiplication symbol $\cdot$, and the number symbols. Let us write this down in more detail and rewrite the basic terms involving s, α_n, and μ_m in standard notation.

2.1.27 Lemma *The following equations hold true for all natural numbers m and n:*

$$s(n) = n + 1 = 1 + n\ , \tag{2.1.2}$$

$$\alpha_m(n) = m + n = n + m\ , \tag{2.1.3}$$

$$\mu_m(n) = m \cdot n = n \cdot m\ , \tag{2.1.4}$$

$$\alpha_m(s(n)) = m + (n + 1) = (m + 1) + n\ , \tag{2.1.5}$$

$$\mu_m(s(n)) = (m \cdot n) + m\ . \tag{2.1.6}$$

Proof. Compute $n + 1 = \alpha_n(s(0)) = s(\alpha_n(0)) = s(n)$. Together with commutativity of addition this equality entails the first equation.

Equations (2.1.3) and (2.1.4) are consequences of the definitions of $+$ and $\cdot$ and commutativity of these operations.

Equation (2.1.5) follows from Equations (2.1.3), (2.1.2) and (2.1.1).

The last equation is a rewrite of the equality $\mu_m(s(n)) = \alpha_m(\mu_m(n))$. □

2.1.28 Even though the monoid $(\mathbb{N}, \cdot, 1)$ is not a cancellation monoid since 0 annihilates $\mathbb{N}$, every non-zero natural number can be multiplicatively cancelled from the left and right.

2.1.29 Proposition *Every element $l \in \mathbb{N}_{>0}$ is both left and right cancellable that is for all $m, n \in \mathbb{N}$ the relation $m \cdot l = n \cdot l$ or $l \cdot m = l \cdot n$ implies $m = n$.*

Proof. Let $k, l \in \mathbb{N}$ and assume that $k \cdot (l + 1) = 0$. Then $0 = k \cdot (l + 1) = k \cdot l + k$, hence $k = k \cdot l = 0$ by Lemma 2.1.19. Therefore, the relation $k \cdot l = 0$ implies $k = 0$ if l is non-zero.

Now assume that $m \cdot l = n \cdot l$ for $m, n \in \mathbb{N}$ and some $l \in \mathbb{N}_{>0}$. By Theorem 2.1.20, the trichotomy law for addition, there exists $k \in \mathbb{N}$ such that $n = m + k$ or $m = n + k$. Assume the first. Then $m \cdot l = n \cdot l = (m + k) \cdot l = m \cdot l + k \cdot l$. By the cancellation law for addition, $k \cdot l = 0$ which entails $k = 0$. Analogously one concludes $k = 0$ if $m = n + k$. So in either case we obtain $m = n$. This is what we had to show. □

2.1.30 Let us finally introduce in this section another algebraic operation on $\mathbb{N}$, namely the power operation. We will show that the power operation provides an example of a monoid homomorphism.

2.1.31 Definition For every natural number k let $\mathbb{N} \to \mathbb{N}$, $n \mapsto k^n$ be the map uniquely defined by Dedekind's Iteration Theorem such that

$$k^0 = 1 \quad \text{and} \quad k^{n+1} = k^n \cdot k \quad \text{for all } n \in \mathbb{N} .$$

One calls the thus defined natural number k^n the n-th power of k.

2.1.32 The power operation intertwines the additive and multiplicative monoid structures on $\mathbb{N}$ or in more precise terms it is a monoid homomorphism between those two monoid structures on $\mathbb{N}$. Before we state and prove that result we briefly introduce the concept of a general homomorphism of monoids.

2.1.33 Definition Let $(M, \star, e)$ and $(N, \bullet, \eta)$ be both monoids. A map $f : M \to N$ then is called a *(monoid) homomorphism* from $(M, \star, e)$ to $(N, \bullet, \eta)$ if it has the following properties:

(Grp5) The map f preserves the binary operations that is

$$f(a \star b) = f(a) \bullet f(b) \quad \text{for all } a, b \in G .$$

(Grp6) The map f preserves the neutral elements which means

$$f(e) = \eta \ .$$

2.1.34 Proposition *The power operation on natural numbers is a homomorphism of monoids from $(\mathbb{N}, +, 0)$ to $(\mathbb{N}, \cdot, 1)$ that is $k^0 = 1$ and $k^{(m+n)} = k^m \cdot k^n$ for all $k, m, n \in \mathbb{N}$. Moreover, $k^n \cdot l^n = (k \cdot l)^n$ for all $k, l, n \in \mathbb{N}$.*

Proof. By definition of the power operation, $k^0 = 1$ for all $k \in \mathbb{N}$. Now fix $m \in \mathbb{N}$. To prove the second part of the claim it suffices to show that the set I of all natural n for which the equality $k^{(m+n)} = k^m \cdot k^n$ holds true is an inductive subset of $\mathbb{N}$. Observe that $k^{(m+0)} = k^m = k^m \cdot 1 = k^m \cdot k^0$, so $0 \in I$. Now assume that $n \in \mathbb{N}$. Then, using the definition of the power operation,

$$k^{m+(n+1)} = k^{(m+n)+1} = k^{(m+n)} \cdot k = (k^m \cdot k^n) \cdot k = k^m \cdot (k^n \cdot k) = k^m \cdot k^{n+1} \ .$$

Hence $n + 1 \in I$, so I is inductive indeed and the homomorphism property is proved.

Now fix $k, l \in \mathbb{N}$. Obviously, $k^0 \cdot l^0 = 1 = (k \cdot l)^0$. Assume that $k^n \cdot l^n = (k \cdot l)^n$ for some $n \in \mathbb{N}$. Then

$$k^{n+1} \cdot l^{n+1} = (k^n \cdot k) \cdot (l^n \cdot l) = (k^n \cdot l^n) \cdot (k \cdot l) = (k \cdot l)^n \cdot (k \cdot l) = (k \cdot l)^{n+1} \ .$$

So the set of all $n \in \mathbb{N}$ for which the equality $k^n \cdot l^n = (k \cdot l)^n$ holds true is an inductive subset of $\mathbb{N}$, hence coincides with $\mathbb{N}$. This proves the second claim. □

The order of natural numbers

2.1.35 Definition One calls a number $m \in \mathbb{N}$ *smaller* or *less* than a number $n \in \mathbb{N}$, in signs $m < n$, if there exists an $l \in \mathbb{N}_{>0}$ such that $n = m + l$. The relation that m is less or equal than n will be denoted by $m \leqslant n$.

If n is smaller than m, we call m *greater* than n and denote this by $m > n$. By $m \geqslant n$ we denote the relation that m is greater or equal than n.

2.1.36 Theorem *The relation $\leqslant$ on the set of natural numbers is a total order relation that is the following axioms are satisfied:*

(O1) Reflexivity
For all $n \in \mathbb{N}$ the relation $n \leqslant n$ holds true.

(O2) Antisymmetry
If $m \leqslant n$ and $n \leqslant m$ for some $m, n \in \mathbb{N}$, then $m = n$.

(O3) Transitivity
For all $k, m, n \in \mathbb{N}$ the relations $k \leqslant m$ and $m \leqslant n$ entail $k \leqslant n$.

(O4) Totality
For all $m, n \in \mathbb{N}$ the relation $m \leqslant n$ or the relation $n \leqslant m$ holds true.

In particular, the law of trichotomy *is satisfied for natural numbers which means that for all* $m, n \in \mathbb{N}$ *exactly one of the following holds true:*

$$m < n, \quad m = n, \quad \text{or} \quad n < m \ .$$

Proof. First note that the relation $m \leqslant n$ is equivalent to the existence of an $l \in \mathbb{N}$ such that $n = m + l$.

The relation $\leqslant$ on $\mathbb{N}$ is reflexive by definition. Assume that $m \leqslant n$ and $n \leqslant m$ for two numbers $m, n \in \mathbb{N}$. Then there exist $k, l \in \mathbb{N}$ such that $n = m + l$ and $m = n + k$. This implies $n = (n+k)+l = n+(k+l)$. By Proposition 2.1.18, $k+l = 0$, hence $k = l = 0$ by Lemma 2.1.19. This implies $m = n$, so $\leqslant$ is antisymmetric. Let us show that $\leqslant$ is transitive. To this end assume $k \leqslant m$ and $m \leqslant n$. Then there exist $j, l \in \mathbb{N}$ such that $m = k + j$ and $n = m + l$. This implies $n = (k + j) + l = k + (j + l)$, hence $k \leqslant n$ and transitivity of $\leqslant$ is proved. So we have shown that $\leqslant$ is an order relation on $\mathbb{N}$.

It remains to verify the trichotomy law which also entails Axiom (O4) or in other words that $\leqslant$ is a total order. But the trichotomy law is an immediate consequence of the trichotomoy law for addition, Theorem 2.1.20. □

2.1.37 Definition Let us remind the reader at this point that a set X together with a binary relation $\leqslant$ on it is called an *ordered set*, *partially ordered set* or a *poset* if Axioms (O1) to (O3) are satisfied (obviously after replacing $\mathbb{N}$ by X in the axioms). The relation $\leqslant$ then is called an *order relation* or a *partial order* on X. If in addition Axiom (O4) holds true, $(X, \leqslant)$ is called a *totally ordered set* and $\leqslant$ a *total order* on X. A set X together with a binary relation $\leqslant$ which fulfills the axiom of reflexivity (O1) and of transitivity (O3) is sometimes called a *preordered set*.

Two elements x, y of a partially ordered set $(X, \leqslant)$ are called *comparable* if $x \leqslant y$ or $y \leqslant x$. So X being totally ordered by $\leqslant$ means that any two of its elements are comparable.

For an order relation $\leqslant$ on a set X one usually abbreviates for elements $x, y \in X$ the relation $x \leqslant y$ and $x \neq y$ by $x < y$. Moreover, $x \geqslant y$ and $x > y$ stand for $y \leqslant x$ and $y < x$, respectively.

Let $(X, \leqslant)$ and $(Y, \preceq)$ be two ordered sets. Then a map $f : X \to Y$ is called *order preserving* or *monotone increasing* if $f(x) \preceq f(y)$ for all $x, y \in X$ with $x \leqslant y$. If instead $x \leqslant y$ implies $f(x) \succeq f(y)$ for all $x, y \in X$, then f is called *order reversing* or *monotone decreasing*. If $x < y$ always implies $f(x) \prec f(y)$, then f is said to be *strictly order preserving* or *strictly monotone increasing*. In case $f(x) \succ f(y)$ for all $x, y \in X$ with $x < y$, then f is a *strictly order reversing* or in other words *strictly monotone decreasing* map.

2.1.38 Remark Note that if an order relation $\leqslant$ on a set X is total, then the *trichotomy law* holds true that is that for each pair of elements $x, y \in X$ exactly one of the relations $x < y$, $x = y$, or $y < x$ is true.

2.1.39 Theorem (Monotony laws) *The algebraic operations and the order relation on* $\mathbb{N}$ *are compatible in the following sense:*

(M1) Monotony of addition
For all $m, n \in \mathbb{N}$ *and* $k \in \mathbb{N}$ *the relation* $m < n$ *implies* $m + k < n + k$.

(M2) Monotony of multiplication
For all $m, n \in \mathbb{N}$ and $k \in \mathbb{N}_{>0}$ the relation $m < n$ implies $m \cdot k < n \cdot k$.

Proof. Assume that $m < n$. Then there exists a unique $l \in \mathbb{N}_{>0}$ such that $n = m + l$.

The equality $n = m + l$ implies $(n + k) = (m + k) + l$ for all $k \in \mathbb{N}$ which entails (M1).

By distributivity $(n \cdot k) = (m \cdot k) + (l \cdot k)$ for every $k \in \mathbb{N}_{>0}$, so (M2) follows if we can yet show that $l \cdot k \in \mathbb{N}_{>0}$. To prove this observe that $k = s(i)$ for some $i \in \mathbb{N}$ and that $l = s(j)$ for some $j \in \mathbb{N}$. Therefore, $l \cdot k = l \cdot s(i) = l \cdot i + l = l \cdot i + s(j) = s\big(l \cdot i + j\big)$, which entails $l \cdot k \in \mathbb{N}_{>0}$ and the claim is proved. □

2.1.40 Proposition *Let m, n be natural numbers such that $n \leqslant m \leqslant n + 1$. Then $m = n$ or $m = n+1$. In other words this means that there is no natural number which is larger than n and smaller than the successor $n + 1$.*

Proof. Assume that $n \leqslant m \leqslant n + 1$ holds true. Then there exists some $l \in \mathbb{N}$ such that $m + l = n + 1$. If If $l = 0$, then $m = n + 1$, if $l = 1$, then $m = n$ by the cancellation property. But we can not have $l > 1$, since otherwise $n + 1 \leqslant m + 1 < m + l = n$ by monotony of addition which is a contradiction to $n < n + 1$. The proof is finished. □

2.1.41 Theorem (Archimedean property of natural numbers) *Let m, n be non-zero natural numbers. Then there exists $k \in \mathbb{N}$ such that*

$$k \cdot m > n \ .$$

Proof. If $m > n$, then put $k = 1$ and we are done. So assume $m \leqslant n$. Then there exists an $l \in \mathbb{N}$ such that $n = m + l$. Put $k = l + 2$. Since $m \geqslant 1$, distributivity and monotony of multiplication entail

$$k \cdot m = (1 + l + 1) \cdot m = m + (l + 1) \cdot m \geqslant m + (l + 1) = (m + l) + 1 = n + 1 > n \ .$$

The claim is proved. □

2.1.42 Before we formulate the well ordering principle for natural numbers let us briefly recall the notions of minimal and maximal elements of a subset of an ordered set and related notions.

2.1.43 Definition Let $(X, \leqslant)$ be an ordered set and $Y \subset X$ a non-empty subset.

(i) An element $b \in X$ is called a *lower bound* of Y if $b \leqslant y$ for all $y \in Y$ and an *upper bound* if $y \leqslant b$ for all $y \in Y$.

(ii) A non-empty subset of X having a lower bound is said to be *bounded below.* If it has an upper bound it is called *bounded above.* If the subset is both bounded below and bounded above, then it is called *bounded.* A non-empty subset of X which is not bounded is called *unbounded.*

(iii) A lower bound l of Y is called a *greatest lower bound* or an *infimum* of Y, if $b \leqslant l$ for every lower bound of Y. An upper bound bound u of Y is called a *least upper bound* or a *supremum* of Y, if $u \leqslant b$ for every upper bound of Y.

(iv) An element $m \in Y$ is called a *minimal* element of Y if for every $y \in Y$ with $y \leqslant m$ the equality $y = m$ holds and a *maximal* element of Y if for every $y \in Y$ with $m \leqslant y$ the equality $y = m$ is true.

(v) An element $m \in Y$ is called a *least element* or *minimum* of Y if $m \leqslant y$ for all $y \in Y$ and a *greatest element* or *maximum* of Y if $y \leqslant m$ for all $y \in Y$.

2.1.44 Lemma *If they exist, the greatest lower bound and the least upper bound of a non-empty subset $Y \subset X$ of an ordered set $(X, \leqslant)$ are uniquely determined. Likewise, least and greatest elements are uniquely determined when they exist.*

Proof. Let l and l' be greatest lower bounds of Y. Then $l' \leqslant l$ and $l \leqslant l'$, hence $l = l'$. The same argument works for the least upper bound and least and greatest elements. $\square$

2.1.45 Remarks (a) The notions of a *lower bound* and of an *upper bound* of a subset $Y \subset X$ defined in Definition 2.1.43 (i) make also sense when $(X, \leqslant)$ is just a preordered set. We will follow this convention.

(b) Unless any two elements of Y in Definition 2.1.43 (iv) are comparable, minimal and maximal elements of Y need not be uniquely determined.

2.1.46 Definition If it exists, the greatest lower bound of a subset Y of an ordered set $(X, \leqslant)$ will be denoted by $\inf Y$, the least upper bound of Y by $\sup Y$, the minimum of Y by $\min Y$, and the maximum of Y by $\max Y$.

2.1.47 Lemma *Every non-empty set of natural numbers which is bounded above has a greatest element.*

Proof. By induction on n we show that every non-empty subset $B \subset \mathbb{N}$ having n has upper bound has a greatest element. If 0 is an upper bound of B, then $B = \{0\}$, and 0 is a greatest element. Assume that for some $n \in \mathbb{N}$ every subset of $\mathbb{N}$ having n as upper bound has a greatest element. Let $B \subset \mathbb{N}$ be a subset having $n + 1$ is an upper bound. Then either $n + 1 \in B$ or $n + 1$ is not an element of B. In the first case, $n + 1$ is a greatest element of B. In the second, $m < n + 1$ for all $m \in B$, hence $m \leqslant n$ for all $m \in B$. By inductive hypothesis B therefore has greatest element in this case too. The argument is finished. $\square$

2.1.48 Theorem (Well ordering principle of natural numbers) *Every non-empty set of natural numbers has a least element.*

Proof. Let $M \subset \mathbb{N}$ be non-empty, and $B \subset \mathbb{N}$ the set of lower bounds of $\mathbb{N}$. Since $0 \in B$, B is non-empty. Since M is non-empty, B is bounded above, hence has a greatest element m by the preceding lemma. Now m is an element of M as well. Assume it is not. Then $m < n$ for all $n \in M$, hence $m + 1 \leqslant n$ for all $n \in M$, and $m + 1$ is a lower bound of M which contradicts m being the greatest lower bound of M. So m is an element of M indeed. Since m is also a lower bound of M, it is a least element of M. $\square$

Finite sets

2.1.49 Definition By an *interval* in $\mathbb{N}$ or an *interval of natural numbers* one understands a set I of the form

$$I = \{k \in \mathbb{N} \mid m \leqslant k \leqslant m\} \quad \text{or} \quad I = \{k \in \mathbb{N} \mid m \leqslant k\}$$

for some $m, n \in \mathbb{N}$. One usually denotes the first kind of interval by $[m, n]_{\mathbb{N}}$ or $\{m, \ldots, n\}$, the second by $\mathbb{N}_{\geqslant m}$. Note that for $n < m$ the intervall $[m, n]_{\mathbb{N}}$ is empty.

A set X is called *finite*, if there exists a bijective map $X \to [1, n]_{\mathbb{N}}$ for some $n \in \mathbb{N}$.

2.1.50 Proposition *Let m and n be natural numbers. Then there exists a bijection $f : [1, n]_{\mathbb{N}} \to [1, m]_{\mathbb{N}}$ if and only if $m = n$.*

Proof. We only need to show that the condition $m = n$ is necessary for the existence of a bijection $f : [1, n]_{\mathbb{N}} \to [1, m]_{\mathbb{N}}$, because it is obviously sufficient; take the identity function, for example.

We prove necessity by induction on n. First assume $n = 0$, and let $f : [1, 0]_{\mathbb{N}} \to [1, m]_{\mathbb{N}}$ be a bijection. Since $[1, 0]_{\mathbb{N}} = \varnothing$, the function f is the empty function. The empty function is only bijective if the range is empty as well. Hence $[1, m]_{\mathbb{N}} = \varnothing$, which is the case if and only if $m = 0$, because otherwise $1 \in [1, m]_{\mathbb{N}}$.

Now assume that the claim holds for some $n \in \mathbb{N}$. Let $f : [1, n+1]_{\mathbb{N}} \to [1, m]_{\mathbb{N}}$ be a bijection. Since $n + 1 \geqslant 1$, the interval $[1, n+1]_{\mathbb{N}}$ is not empty, hence $[1, m]_{\mathbb{N}}$ is so, too. This implies that $m = \tilde{m} + 1$ for a unique $\tilde{m} \in \mathbb{N}$. Let $k = f^{-1}(m)$. If $k = n + 1$, let $\tau : [1, n+1]_{\mathbb{N}} \to [1, n+1]_{\mathbb{N}}$ be the identity function, otherwise put $\tau(k) = n + 1$, $\tau(n + 1) = k$, and $\tau(i) = i$ for all $i \in [1, n+1]_{\mathbb{N}} \backslash \{k, n+1\}$. The function $g = f \circ \tau : [1, n+1]_{\mathbb{N}} \to [1, m]_{\mathbb{N}}$ then is a bijection as well, and $g(n+1) = m$. Hence the restriction $g|_{[1,n]_{\mathbb{N}}}$ has image $[1, m]_{\mathbb{N}} \backslash \{m\} = [1, \tilde{m}]_{\mathbb{N}}$; note that for the latter equality we have used Proposition 2.1.40. By inductive assumption one obtains $n = \tilde{m}$, hence $n + 1 = m$, and the proposition is proved. $\square$

2.1.51 Definition For a finite set X the unique natural number n for which there exists a bijection $X \to [1, n]_{\mathbb{N}}$ is called the *cardinality* of X. It is denoted $|X|$.

2.1.52 Proposition *Let X and Y be finite sets. Then the following holds true:*

(i) *The cardinality of the disjoint union $X \sqcup Y$ is $|X| + |Y|$.*

(ii) *The cardinality of the cartesian product $X \times Y$ is $|X| \cdot |Y|$.*

(iii) *The cardinality of the power set $\mathcal{P}(X)$ is $2^{|X|}$.*

Proof. ad (i). It suffices to show that for all $m, n \in \mathbb{N}_{>0}$ the disjoint union $[1, m]_{\mathbb{N}} \sqcup [1, n]_{\mathbb{N}}$ can be bijectively mapped onto $[1, m+n]_{\mathbb{N}}$. But such a bijection is given by

$$[1, m]_{\mathbb{N}} \sqcup [1, n]_{\mathbb{N}} = \{1\} \times [1, m]_{\mathbb{N}} \cup \{2\} \times [1, n]_{\mathbb{N}} \to [1, m+n]_{\mathbb{N}}, \ (i, k) \mapsto \begin{cases} k, & \text{if } i = 1 \\ m + k, & \text{if } i = 2. \end{cases}$$

ad (ii). Now one needs to prove that $|[1,m]_{\mathbb{N}} \times [1,n]_{\mathbb{N}}| = n \cdot m$ for all $m, n \in \mathbb{N}$. Let us show that the set I of all $n \in \mathbb{N}$ for which this equality holds for every $m \in \mathbb{N}$ is inductive. This will prove (ii). Obviously, $0 \in I$ since $[1,m]_{\mathbb{N}} \times \emptyset = \emptyset$. Next observe that if $Z = \{z\}$ is a set having a unique element z, then the map $[1,m]_{\mathbb{N}} \times Z \to [1,m]_{\mathbb{N}}$, $(k,z) \mapsto k$ is a bijection, so $[1,m]_{\mathbb{N}} \times Z$ has cardinality m. Under the assumption $n \in I$ this observation entails

$$\begin{aligned}\big|[1,m]_{\mathbb{N}} \times [1,n+1]_{\mathbb{N}}\big| &= \big|[1,m]_{\mathbb{N}} \times [1,n]_{\mathbb{N}} \cup [1,m]_{\mathbb{N}} \times \{n+1\}\big| = \\ &= \big|[1,m]_{\mathbb{N}} \times [1,n]_{\mathbb{N}}\big| + \big|[1,m]_{\mathbb{N}} \times \{n+1\}\big| = \\ &= m \cdot n + m = m \cdot (n+1) \ ,\end{aligned}$$

where in the second equality we have used (i). Hence $n+1$, and the second claim is proved.

ad (iii). Obviously, if $f : X \to [1,n]_{\mathbb{N}}$ is a bijection, then the induced map $f : \mathcal{P}(X) \to \mathcal{P}([1,n]_{\mathbb{N}})$, $A \mapsto f(A)$ is one, too. This means that we need to verify the claim only for sets of the form $X = [1,n]_{\mathbb{N}}$ with $n \in \mathbb{N}$. Let J be the set of all natural numbers n such that $\big|\mathcal{P}([1,n]_{\mathbb{N}})\big| = 2^n$. Since $\mathcal{P}(\emptyset) = \{\emptyset\}$, one has $|\mathcal{P}(\emptyset)| = 1 = 2^0$, hence $0 \in J$. Assume $n \in J$. Then $\mathcal{P}([1,n+1]_{\mathbb{N}})$ is the disjoint union the set P of all subsets of $[1,n+1]_{\mathbb{N}}$ not containing $n+1$ and the set Q of all subsets containing $n+1$. Then $\big|\mathcal{P}([1,n+1]_{\mathbb{N}})\big| = |P| + |Q|$. Moreover, P coincides with $\mathcal{P}([1,n]_{\mathbb{N}})$, hence $|P| = 2^n$. So the claim is proved if one can yet show that $|Q| = 2^n$, because then $n+1 \in J$ by

$$\big|\mathcal{P}([1,n+1]_{\mathbb{N}})\big| = |P| + |Q| = 2 \cdot 2^n = 2^{n+1} \ .$$

But the map $Q \to P$, $A \mapsto A \backslash \{n+1\}$ is a bijection with inverse $P \to Q$, $B \mapsto B \cup \{n+1\}$, which entails the remaining claim. □

2.1.53 Proposition *Let $(X, \leqslant)$ be a finite ordered set of cardinality $n \in \mathbb{N}_{>0}$. Then X has a minimum and a maximum and there exists a unique strictly order preserving map $[1,n]_{\mathbb{N}} \to X$. This map is a bijection.*

Proof. We use induction by n. For $n = 1$ the claim is trivial because then the unique element of X is both a minimum and a maximum and there exists only one map between $\{1\}$ and X in this situation. That map is obviously order preserving. So assume the claim holds true for some natural $n > 0$ and that X is an ordered set of cardinality $n+1$. Let $x_0 \in X$ be an element of X and consider $X' = X \backslash \{x_0\}$. Then X' has a minium x'_{m} and a maximum x'_{M}. If $x_0 < x'_{\mathrm{m}}$ put $x_{\mathrm{m}} = x_0$, otherwise let $x_{\mathrm{m}} = x'_{\mathrm{m}}$. By construction, x_{m} is the minimum of X. Likewise one constructs the maximum x_{M}: define $x_{\mathrm{M}} = x_0$ if $x'_{\mathrm{M}} < x_0$, otherwise let $x_{\mathrm{M}} = x'_{\mathrm{M}}$. By inductive assumption there exists a unique strict order preserving map $f' : [1,n]_{\mathbb{N}} \to X \backslash \{x_{\mathrm{M}}\}$. Extend f' to a map $f : [1,n+1]_{\mathbb{N}} \to X$ by putting $f(n+1) = x_{\mathrm{M}}$. By constructing f is strictly order preserving. Let $g : [1,n+1]_{\mathbb{N}} \to X$ be another strictly order preserving map. Since g is strictly order preserving, the map $g' : [1,n]_{\mathbb{N}} \to X \backslash \{g(n+1)\}$, $k \mapsto g(k)$ is well-defined and strictly order preserving. By inductive assumption g' is a bijection. Hence $x < g(n+1)$ for all $x \in X \backslash \{g(n+1)\}$, so $g(n+1) = x_{\mathrm{M}} = f(n+1)$. Since f' and g' coincide by inductive assumption, $f = g$ follows. Moreover, $\mathrm{im}(f) = \mathrm{im}(g) = X$, which shows surjectivity of f. As a strictly order preserving map it is also injective, and the claim is proved. □

Recursion and sequences

2.1.54 Recursion is a crucial tool to construct sequences of natural numbers, real numbers or more general objects, and goes back to Dedekind's iteration scheme. We will introduce here a quite broad notion of a sequence and then formulate and prove the recursion theorem. Afterwards we will use it to define further operations on natural numbers like the factorial and some examples of recursively defined sequences.

2.1.55 Definition Let X denote a set. A map $x : I \to X$ from an interval $I \subset \mathbb{N}$ to X is called a *sequence* in X and I its *index set*. A sequence $x : I \to X$ is usually denoted $(x_k)_{k\in I}$, where x_k stands for the value $x(k)$ at $k \in I$. If the set $I \subset \mathbb{N}$ is finite, one calls a sequence $(x_k)_{k\in I}$ *finite* otherwise *infinite*. The cardinality of the index set of a finite set is sometimes called the *length* of the sequence or the index set. In case the index set of $(x_k)_{k\in I}$ is of the form $I = [m,n]_{\mathbb{N}}$, we will usually denote the sequence by $(x_k)_{k=m}^n$ or $(x_k)_{m\leqslant k\leqslant n}$. If the index set of $(x_k)_{k\in I}$ is equal to some $\mathbb{N}_{\geqslant m}$, the sequence is usually written in the form $(x_k)_{k=m}^\infty$.

2.1.56 Theorem (Recursion Theorem) *Let X be a set, $x_0 \in X$ an element, and $t : \mathbb{N}\times X \to X$ a function. Then there exists a unique function $f : \mathbb{N} \to X$ such that $f(0) = x_0$ and such that $f(n+1) = t(n, f(n))$ for all $n \in \mathbb{N}$.*

Proof. Define the map $T : \mathbb{N} \times X \to \mathbb{N} \times X$ by $T(n,x) = (n, t(n,x))$ for all $n \in \mathbb{N}$ and $x \in X$. By Dedekind's Iteration Theorem 2.1.5 there exists a unique function $F : \mathbb{N} \to \mathbb{N} \times X$ such that $F(0) = (0, x_0)$ and $F(n+1) = T(F(n))$ for all $n \in \mathbb{N}$. Put $f := \mathrm{pr}_2 \circ F$, where $\mathrm{pr}_i : \mathbb{N} \times X \to X$ for $i = 1, 2$is projection onto the i-th coordinate. Then $f(0) = x_0$. Let us show that $f(n+1) = t(n, f(n))$ for all $n \in \mathbb{N}$. Since $\mathrm{pr}_1 \circ T = \mathrm{pr}_1$, Dedekind's Iteration Theorem 2.1.5 entails $\mathrm{pr}_1 \circ F = \mathrm{id}_{\mathbb{N}}$, hence

$$f(n+1) = \mathrm{pr}_2(F(n+1)) = \mathrm{pr}_2(T(F(n))) = t\big(\mathrm{pr}_1(F(n)), f(n)\big) = t(n, f(n)) \quad \text{for all } n \in \mathbb{N} \ .$$

So we have shown the existence of a function f with the desired properties. Let $\tilde{f} : \mathbb{N} \to X$ be another one. Then $\tilde{f}(0) = x_0 = f(0)$ and $\tilde{f}(n+1) = t(n, \tilde{f}(n)) = t(n, f(n)) = f(n+1)$ if $\tilde{f}(n) = f(n)$. By induction, the equality $\tilde{f} = f$ follows. The claim is proved. □

2.1.57 Proposition and Definition *Let $(M, \star, e)$ be a monoid. Then the following holds true:*

(i) *For every $a \in M$ there exists a unique sequence $\left(\times^n a\right)_{n\in\mathbb{N}}$ in M such that*

$$\times^0 a = e \quad \textit{and} \quad \times^{n+1} a = (\times^n a) \star a \quad \textit{for all } n \in \mathbb{N} \ .$$

*In case the monoid M is multiplicatively written that is the operation is denoted by $\cdot$ or $\circ$ and the neutral element by 1, the element $\times^n a$ will be denoted a^n and called the n-*th power *of a. If the monoid M is abelian and additively written which means that the operation is denoted by $+$ or $\oplus$ and the neutral element by 0, then $\times^n a$ will be denoted $n \cdot a$ or briefly na.*

(ii) *Let $\mathcal{I} \subset \mathcal{P}(N)$ be the set of finite intervals in $\mathbb{N}$ and $\mathcal{S}$ the set $\bigcup_{I\in\mathcal{I}} M^I$ of all finite sequences in M. Then there exists a unique map $\bigtimes : \mathcal{S} \to M$ such that*

$$\bigtimes \big((a_i)_{i\in\varnothing}\big) = e \textit{ and } \bigtimes \big((a_i)_{i\in I}\big) = \bigtimes \big((a_i)_{i\in I'}\big) \star a_{\max I} \textit{ for all } (a_i)_{i\in I} \in \mathcal{S}, I \neq \varnothing \ .$$

Hereby, $(a_i)_{i\in\varnothing}$ denotes the empty sequence in M, and I' for every non-empty finite interval $I \subset \mathbb{N}$ the interval $I\backslash\{\max I\}$. One usually denotes $\bigtimes\big((a_i)_{i\in I}\big)$ by $\bigtimes_{i\in I} a_i$ or by $\bigtimes_{i=m}^{n} a_i$, when $I = [m,n]$, and calls it the composition *of the finite sequence $(a_i)_{i\in I}$.*

In case the monoid M is multiplicatively written, then one writes $\prod_{i\in I} a_i$ or $\prod_{i=m}^{n} a_i$ for the composition of a finite sequence $(a_i)_{i\in I}$ in M. If the monoid M is abelian and additively written, then one writes $\sum_{i\in I} a_i$ or $\sum_{i=m}^{n} a_i$ instead of $\bigtimes_{i\in I} a_i$ and calls the composition of $(a_i)_{i\in I}$ the sum *of the finite sequence $(a_i)_{i\in I}$ in M.*

Proof. *ad* (*i*). This follows immediately from the Recursion Theorem 2.1.56.

ad (*ii*). For $n \in \mathbb{N}$ let $\mathcal{I}_n$ be the set of all intervals in $\mathbb{N}$ of length n, $\mathcal{S}_n = \bigcup_{I\in\mathcal{I}_n} M^I$ the set of sequences of length n, and X_n be the set $M^{\mathcal{S}_n}$ of all functions $\mathcal{S}_n \to M$. Note that the X_n are pairwise disjoint and put $X = \bigcup_{n\in\mathbb{N}} X_n$. Now let $x_0 \in X$ be the element of X_0 which maps the empty sequence $(a_i)_{i\in\varnothing}$ to 1. Define $t_n : X_n \to X_{n+1}$ for $n \in \mathbb{N}$ by

$$t(x)\big((a_i)_{i\in I}\big) := x\big((a_i)_{i\in I'}\big) \star a_{\max I} \quad \text{where } x \in X_n,\ I \in \mathcal{I}_{n+1},\ \text{ and } (a_i)_{i\in I} \in \mathcal{S}_{n+1}$$

The maps t_n define a unique map $t : X \to X$ such that t_n coincides with the restriction of t to X_n. By the Recursion Theorem 2.1.56 there now exists a map $f : \mathbb{N} \to X$ such that

$$f(0) = x_0 \quad \text{and} \quad f(n+1) = t(f(n)) \quad \text{for all } n \in \mathbb{N}\ .$$

Since $f(0) \in X_0$ and since $f(n) \in X_n$ implies $f(n+1) = t(f(n)) = t_n(f(n)) \in X_{n+1}$ for all $n \in \mathbb{N}$, one has $f(n)) \in X_n$ for all $n \in \mathbb{N}$ which means that $f(n)$ is a map that assigns to each sequence $(a_i)_{i\in I}$ of length n a value $f(n)\big((a_i)_{i\in I}\big) \in M$. The map $\bigtimes : \mathcal{S} \to M$, $(a_i)_{i\in I} \mapsto f(|I|)\big((a_i)_{i\in I}\big)$ then has the claimed properties since

$$\bigtimes\big((a_i)_{i\in\varnothing}\big) = f(0)\big((a_i)_{i\in\varnothing}\big) = 1$$

and since for all $n \in \mathbb{N}$ and sequences $(a_i)_{i\in I} \in \mathcal{S}_{n+1}$

$$\begin{aligned}\bigtimes\big((a_i)_{i\in I}\big) &= f(n+1)\big((a_i)_{i\in I}\big) = t(f(n))\big((a_i)_{i\in I}\big) = \\ &= f(n)\big((a_i)_{i\in I'}\big) \star a_{\max I} = \bigtimes\big((a_i)_{i\in I'}\big) \star a_{\max I}\ .\end{aligned}$$

Assume that $\widetilde{\bigtimes} : \mathcal{S} \to M$ is another such map. Let $J \subset \mathbb{N}$ be the set of all $n \in \mathbb{N}$ such that for every sequence $(a_i)_{i\in I}$ of finite length n the equality $\widetilde{\bigtimes}\big((a_i)_{i\in I}\big) = \bigtimes\big((a_i)_{i\in I}\big)$ holds true. Obviously, 0 is an element of J. If $n \in J$, then one has for every sequence $(a_i)_{i\in I}$ of length $n+1$

$$\widetilde{\bigtimes}\big((a_i)_{i\in I}\big) = \widetilde{\bigtimes}\big((a_i)_{i\in I'}\big) \star a_{\max I} = \bigtimes\big((a_i)_{i\in I'}\big) \star a_{\max I} = \bigtimes\big((a_i)_{i\in I}\big)\ .$$

Hence $n+1 \in J$, too, so J is an inductive subset of $\mathbb{N}$ and coincides with $\mathbb{N}$. Therefore $\widetilde{\bigtimes} = \bigtimes$ and the claim is proved. □

2.1.58 Remark Given an element a of a monoid M, the equality $\times^n a = \bigtimes_{i=1}^{n} a$ holds true for every $n \in \mathbb{N}$, where $(a)_{1\leqslant i\leqslant n}$ denotes the constant family with value a.

2.1.59 Proposition *Let $(M, \star, e)$ be a monoid and $a \in M$. Then the map $\mathbb{N} \to M$, $n \mapsto \times^n a$ is a monoid homomorphism.*

Proof. By definition of the operation, $\times^0 a = e$. Now fix $m \in \mathbb{N}$. To prove that $\times^{m+n} a = (\times^m a) \star (\times^n a)$ for all $n \in \mathbb{N}$ observe first that $\times^{m+0} a = \times^m a = (\times^m a) \star (\times^0 a)$ and then that the validity of $\times^{m+n} a = (\times^m a) \star (\times^n a)$ implies

$$\begin{aligned}\times^{m+(n+1)} a &= (\times^{m+n} a) \star a = \big((\times^m a) \star (\times^n a)\big) \star a \\ &= (\times^m a) \star \big((\times^n a) \star a\big) = (\times^m a) \star (\times^{n+1} a) \ .\end{aligned}$$

By the induction axiom the equality $\times^{m+n} a = (\times^m a) \star (\times^n a)$ therefore holds true for all $m, n \in \mathbb{N}$ and the claim is proved. $\square$

2.1.60 Definition The *factorial* is the function $! : \mathbb{N} \to \mathbb{N}$, $n \to n!$ defined *recursively* by $0! = 1$ and $(n+1)! = n! \cdot (n+1)$.

2.1.61 Definition For all $n, k \in \mathbb{N}$, the *binomial coefficients* $\binom{n}{k}$ are defined recursively as follows:

$$\binom{0}{k} = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{else,} \end{cases} \qquad \text{and} \qquad \binom{n+1}{k} = \begin{cases} 1, & \text{if } k = 0 \text{ or } k = n+1, \\ \binom{n}{k-1} + \binom{n}{k}, & \text{if } 1 \leqslant k \leqslant n, \\ 0, & \text{else .} \end{cases}$$

2.1.62 Remark (Proof by induction) Let $P(x)$ denote a formula or property in one variable x. To prove that $P(n)$ holds true for all natural numbers n greater or equal than a fixed natural number m it suffices to show that the set

$$I = \big\{ n \in \mathbb{N} \mid (n < m) \ \vee \ \big(n \geqslant m \;\&\; P(n)\big) \big\}$$

is inductive. But this set is inductive if and only if one can verify $P(m)$ and that for all natural $n \geqslant m$ the implication $P(n) \Rightarrow P(n+1)$ holds true. When one verifies these two statements one usually says that one proves $P(n)$ (for natural n or natural $n \geqslant m$) by induction on n. From now on we will use this language, too.

2.2. Integers

2.2.1 Even though the semiring of natural numbers has the cancellation property with regard to addition, there does not exist an additive inverse for a non-zero element in $\mathbb{N}$. More precisely, this means that for $n \in \mathbb{N}_{>0}$ there is no element $m \in \mathbb{N}$ such that $n + m = m + n = 0$. Otherwise the trichotomy law for addition would be violated. The lack of additive inverses in $\mathbb{N}$ is a "deficiency" and as usual in mathematics, one heals this by adding what is missing, in this case additive inverses or in other words negative numbers. As a result one obtains the abelian group of integers $(\mathbb{Z}, +, 0)$ which even carries the structure of a commutative ring when extending multiplication from $\mathbb{N}$ to $\mathbb{Z}$.

In modern mathematics, the idea of "supplementing" an abelian monoid with additive inverses led to the invention of K-theory in Grothendieck (1957a), where the K-theory $K_0(X)$ of a quasi-projective variety X has been constructed as the K-group associated to the abelian monoid of isomorphism classes of (algebraic) vector bundles over X. The main ideas of constructing $\mathbb{Z}$ out of $\mathbb{N}$ and of constructing the K-group of an abelian monoid are essentially the same, so we will present the more abstract concept already here.

Interlude on abelian groups and rings

Before we can describe the construction of Grothendieck's K-group, we need to explain what an abelian group and a ring are: the former is an abelian monoid in which every element has an inverse, the latter a semiring which forms an abelian group with respect to addition. Let us first give the full definition of a group which includes the non-abelian case as well.

2.2.2 Definition A monoid $(G, \star, e)$ is called a *group* if it satisfies the additional axiom

(Grp3) For every $a \in G$ exists an *inverse* that is an element $b \in G$ such that

$$a \star b = b \star a = e \ .$$

If in addition to this (Grp4) is satisfied that is if $a \star b = b \star a$ for all $a, b \in G$, then the group is called *abelian*.

2.2.3 Example (a) The set $\mathbb{N}$ of natural numbers together with addition $+$ as binary operation and 0 as neutral element is an abelian monoid by Theorem 2.1.14, but it is not a group since no element of $\mathbb{N}_{>0}$ has an inverse by the trichotomy law for addition, Theorem 2.1.20.

(b) The set $G = \{-1, 1\}$ with operation $\cdot : G \times G \to G$ defined by $1 \cdot 1 = (-1) \cdot (-1) = 1$ and $(-1) \cdot 1 = 1 \cdot (-1) = -1$ and neutral element 1 is an abelian group.

2.2.4 Remarks (a) The binary operation of a group G is also called its *group law*.

(b) The binary operation of an abelian group or monoid A is often denoted by $+$ or $\oplus$, and its neutral element by 0. We sometimes say in this case that the operation is *additively written*. If A is an abelian group with additively written operation, the inverse of an element $a \in A$ is usually denoted by $-a$ and called the *negative* of a.

(c) If the binary operation of a a group or monoid G is denoted by $\cdot$ or $\circ$, then one sometimes says that the operation is *multiplicatively written*. Usually, the identity element of a monoid or group G with multiplicatively written operation is denoted by 1, and the inverse of an element $b \in G$ (if it exists) by b^{-1}.

(d) Even though not every element in a monoid $(G, \star, e)$ needs to have an inverse some still might. One gives those elements $a \in G$ for which there exists an *inverse* that is a $b \in G$ such that $a \star b = b \star a = e$ a name and calls them *invertible*.

2.2.5 Following the categorical paradigm of this book we now introduce structure preserving maps between monoids and groups.

2.2.6 Definition Let $(G, \star, e)$ and $(H, \bullet, \eta)$ be both groups or both monoids. A map $f : G \to H$ then is called a (*group* respectively *monoid*) *homomorphism* if the following holds:

(Grp5) The binary operations are preserved which means that

$$f(a \star b) = f(a) \bullet f(b) \quad \text{for all } a, b \in G \ .$$

(Grp6) The neutral elements are preserved that is

$$f(e) = \eta \ .$$

2.2.7 Proposition *Let $(M, \star, e)$ be a monoid.*

(i) *The neutral element of M is uniquely determined that is if e' is another element neutral with respect to the operation $\star$, then $e = e'$.*

(ii) *If $b \in M$ is a left inverse of some $a \in M$ and $c \in M$ a right inverse that is if $b \star a = e$ and $a \star c = e$, then $b = c$ and a is invertible. In particular inverse elements in M are uniquely determined.*

(iii) *If $a \in M$ is invertible with inverse $b \in M$, then b is invertible with inverse a.*

(iv) *If $a \in M$ is invertible with inverse $b \in M$ and $c \in M$ is invertible with inverse $d \in M$, then ac is invertible with inverse db.*

(v) *If $f : G \to H$ is a homomorphism from $(G, \star, e)$ to a monoid $(H, \bullet, \eta)$, and if $b \in G$ is the inverse of some $a \in G$, then $f(b)$ is the inverse of $f(a)$ in H.*

Proof. *ad (i)*. Both e and e' are neutral, therefore

$$e = e \star e' = e' \ .$$

ad (ii). This follows from

$$b = b \star e = b \star (a \star c) = (b \star a) \star c = e \star c = c \ .$$

ad (iii). This is clear since $ab = e$ and $ba = e$.

ad (iv). The claim is immediate by

$$(ac)(db) = a(cd)b = aeb = ab = e \quad \text{and} \quad (db)(ac) = d(ba)c = dec = dc = e \ .$$

ad (v). Just compute

$$f(a) \bullet f(b) = f(a \star b) = f(e) = \eta \quad \text{and} \quad f(b) \bullet f(a) = f(b \star a) = f(e) = \eta \ . \qquad \square$$

2.2.8 Proposition *Let $(G, \star, e)$ be group and $G \to G$, $a \mapsto a^{-1}$ its inverse map.*

(i) *For given elements $a, b \in G$ the equations $ax = b$ and $ya = b$ have each exactly one solution namely $x = a^{-1}b$ and $y = ba^{-1}$.*

(ii) *The inverse map is an* involution *that is* $\left(a^{-1}\right)^{-1} = a$ *for all* $a \in G$.

(iii) *The inverse map is an* antihomomorphism *that is* $(ab)^{-1} = b^{-1}a^{-1}$ *for all* $a, b \in G$.

(iv) *If* $(H, \bullet, \eta)$ *is a group as well and* $f : G \to H$ *a map which satisfies* (Grp5)*, then* f *is a group homomorphism.*

Proof. *ad* (*i*). Obviously, $x = a^{-1}b$ solves the equation $ax = b$ and $y = ba^{-1}$ the equation $ya = b$. Uniqueness of x and y follows by multiplication of $ax = b$ from the left and $ya = b$ from the right by a^{-1}.

ad (*ii*). Proposition 2.2.7 (iii) says that the inverse map on a group is an involution.

ad (*iii*). This is a consequence of Proposition 2.2.7 (iv).

ad (*iv*). We only have to show that f preserves the identity elements. To this end let ϱ be the inverse of $f(e)$ and compute:

$$\eta = \varrho \bullet f(e) = \varrho \bullet f(e \star e) = \varrho \bullet \big(f(e) \bullet f(e)\big) = \big(\varrho \bullet f(e)\big) \bullet f(e) = \eta \bullet f(e) = f(e) \ . \qquad \square$$

2.2.9 Theorem and Definition (i) *The identity map on a monoid is a homomorphism, likewise the composition of two monoid homomorphisms.*

(ii) *Monoids as objects together with their homomorphisms as morphisms form a category denoted by* Mon*. When restricting objects to abelian monoids one obtains a full subcategory denoted by* AbMon*.*

(iii) *Groups as objects together with their homomorphisms as morphisms form a full subcategory of* Mon *denoted by* Grp*. When restricting objects to abelian groups one obtains the full subcategory* Ab *of abelian groups.*

Proof. *ad* (*i*). The identity map id_G on a monoid $(G, \star, e)$ is obviously a homomorphism since it leaves the product of two elements and the neutral element e invariant. Assume that $(G, \star, e)$, $(H, \bullet, \eta)$ and $(K, \cdot, o)$ are three monoids and let $f : G \to H$ and $g : H \to K$ be a homomorphism. Then the composition $g \circ f : G \to K$ satisfies

$$g \circ f(a \star b) = g(f(a \star b)) = g(f(a) \bullet f(b)) = g(f(a)) \cdot g(f(b)) = (g \circ f(a)) \cdot (g \circ f(b))$$

for all $a, b \in G$ and

$$g \circ f(e) = g(f(e)) = g(\eta) = o \ ,$$

hence $g \circ f$ is a homomorphism of monoids.

ad (*ii*). This follows from (i).

ad (*iii*). This is clear by (ii) and the definition of group homomorphisms. $\square$

2.2.10 Definition A semiring $(R, +, \cdot, 0, 1)$ is called a *ring* if it satisfies

(Ring1) R together with addition $+$ and the element 0 is an abelian group.

If in addition (Ring4) holds true i.e. if multiplication is commutative, then R is called a *commutative ring*.

2.2.11 Remark Note that Axiom (SRing2a), which says that 0 annihilates every element, follows from (Ring1) and the distributivity axiom (Ring3). Namely, if these two axioms hold for $(R, +, \cdot, 0, 1)$, then for all $x \in R$

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

and similarly $0 = x \cdot 0$. This means that in the definition of a ring, unlike for the one of a semiring, Axiom (SRing2a) is actually not necessary because it is a consequence of the other ones. In **??**, where we expand on ring theory, we therefore do not make use of (SRing2a) in the definion of a ring.

2.2.12 Definition Let $(R, +_R, \cdot_R, 0_R, 1_R)$ and $(S, +_S, \cdot_S, 0_S, 1_S)$ be both semirings or both rings. A map $f : R \to S$ then is called a (*semiring* respectively *ring*) *homomorphism* if the following holds:

(Ring5) The addition maps are preserved which means that

$$f(r +_R s) = f(r) +_S f(s) \quad \text{for all } r, s \in R \ .$$

(Ring6) The multiplication maps are is preserved which means

$$f(r \cdot_R s) = f(r) \cdot_S f(s) \quad \text{for all } r, s \in R \ .$$

(Ring7) The zero elements are preserved that is

$$f(0_R) = 0_S \ .$$

(Ring8) The identity elements are preserved that is

$$f(1_R) = 1_S \ .$$

An injective ring or semiring homomorphism $f : R \to S$ is called an embedding of rings respectively semirings.

2.2.13 Proposition *If $(R, +_R, \cdot_R, 0_R, 1_R)$ and $(S, +_S, \cdot_S, 0_S, 1_S)$ are both rings, and $f : R \to S$ is a map which satisfies* (Ring5)*,* (Ring6)*, and* (Ring8)*, then* (Ring7) *is satisfied as well, so f is a ring homomorphism.*

Proof. This follows immediately from Proposition 2.2.8 (iv). □

2.2.14 Theorem and Definition (i) *The identity map on a semiring is a homomorphism, likewise the composition of two semiring homomorphisms.*

(ii) *Semirings as objects together with their homomorphisms as morphisms form a category denoted by* SRing*. When restricting objects to commutative semirings one obtains a full subcategory denoted by* CSRing.

(iii) *Rings as objects together with their homomorphisms as morphisms form a full subcategory of* SRing *denoted by* Ring. *When restricting objects to commutative rings one obtains the full subcategory* CRing *of commutative rings.*

Proof. *ad (i).* The identity map id_R on a semiring $(R, +_R, \cdot_R, 0_R, 1_R)$ is obviously a homomorphism since it leaves the sum and product of two elements invariant and preserves the zero and identity elements. Assume that and $(S, +_S, \cdot_S, 0_S, 1_S)$ and $(T, +_T, \cdot_T, 0_T, 1_T)$ are two more semirings and let $f : R \to S$ and $g : S \to T$ be homomorphisms. Then the composition $g \circ f : R \to T$ satisfies

$$g \circ f(r \cdot_R s) = g(f(r \cdot_R s)) = g(f(r) \cdot_S f(s)) = g(f(r)) \cdot_T g(f(s)) = (g \circ f(r)) \cdot_T (g \circ f(s))$$

for all $r, s \in R$ and

$$g \circ f(1_R) = g(f(1_R)) = g(1_S) = 1_T \ ,$$

hence $g \circ f$ is a homomorphism of semirings.

ad (ii). This follows from (i).

ad (iii). This is clear by (ii) and the definition of ring homomorphisms. □

2.2.15 Proposition *The following algebraic relations hold in any ring* $(R, +, \cdot, 0, 1)$*:*

(i) $(-r)s = r(-s) = -(rs)$ *for all* $r, s \in R$.

(ii) $(-1)^2 = 1$.

(iii) *A multiplicative identity element in* R *is uniquely determined.*

(iv) *Assume that* R *possesses an identity element. Then the inverse for an invertible* $r \in R$ *is uniquely determined. If it exists, the inverse of* r *is denoted by* r^{-1}.

Proof. *ad (i).* Since 0 is an annihilator in R, one obtains

$$0 = 0 \cdot s = (r + (-r)) \cdot s = rs + (-r)s \ ,$$

which entails $(-r)s = -(rs)$ after adding $-(rs)$ on both sides. Similarly, one shows $r(-s) = -(rs)$.

ad (ii). This is a consequence of $0 = 0 \cdot (-1) = (1 + (-1)) \cdot (-1) = (-1) + (-1)^2$.

Properties (iii) and (iv) follow from Proposition 2.2.7 (i) and (ii), respectively. □

The Grothendieck group of an abelian monoid

2.2.16 Definition Let M be an abelian monoid with binary operation $\oplus : M \times M \to M$ and neutral element $0 \in M$. An abelian group K together with a morphism $\kappa : M \to K$ of monoids is called a *Grothendieck group* of M if the following universal property is satisfied:

(Gro) For every abelian group A and every morphism of monoids $f : M \to A$ there exists a unique homomorphism of groups $f_K : K \to A$ such that the diagram

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "\kappa"'] & A \\ K \arrow[ur, "f_K"'] & \end{tikzcd}$$

commutes.

2.2.17 We will now construct a Grothendieck group for an abelian monoid $(M, \oplus, 0)$. To this end define an equivalence relation $\sim$ on the cartesian product $M \times M$ by putting $(m, n) \sim (p, q)$ for $(m, n), (p, q) \in M \times M$ if $m \oplus q \oplus k = p \oplus n \oplus k$ for some $k \in M$. The relation $\sim$ is obviously reflexive and symmetric. To verify transitivity let $(a, b) \sim (m, n)$ and $(m, n) \sim (p, q)$. Then choose $k, l \in M$ such that $a \oplus n \oplus k = m \oplus b \oplus k$ and $m \oplus q \oplus l = p \oplus n \oplus l$ and compute

$$\begin{aligned} (a \oplus q) \oplus (n \oplus k \oplus l) &= (a \oplus n \oplus k) \oplus (q \oplus l) = (m \oplus b \oplus k) \oplus (q \oplus l) = \\ &= (m \oplus q \oplus l) \oplus (b \oplus k) = (p \oplus n \oplus l) \oplus (b \oplus k) = (p \oplus b) \oplus (n \oplus k \oplus l). \end{aligned}$$

So $(a, b) \sim (p, q)$, hence $\sim$ is transitive and an equivalence relation indeed.

The $\sim$ equivalence class of an element $(m, n) \in M \times M$ will be denoted by $[m, n]$.

Next we define a binary operation on the set of equivalence classes

$$K^{\mathrm{Gro}}(M) := (M \times M)/\!\sim \ .$$

For $[m, n], [p, q] \in K^{\mathrm{Gro}}(M)$ put

$$[m, n] \oplus [p, q] := [m \oplus p, n \oplus q].$$

The sum $[m, n] \oplus [p, q]$ is well-defined by this indeed, since for $(m, n) \sim (m', n')$ and $(p, q) \sim (p', q')$ the pairs $(m \oplus p, n \oplus q)$ and $(m' \oplus p', n' \oplus q')$ are equivalent by

$$\begin{aligned} (m \oplus p) \oplus (n' \oplus q') \oplus (k \oplus l) &= (m \oplus n' \oplus k) \oplus (p \oplus q' \oplus l) = \\ &= (m' \oplus n \oplus k) \oplus (p' \oplus q \oplus l) = (m' \oplus p') \oplus (n \oplus q) \oplus (k \oplus l), \end{aligned}$$

where $k, l \in M$ have been chosen so that $m \oplus n' \oplus k = m' \oplus n \oplus k$ and $p \oplus q' \oplus l = p' \oplus q \oplus l$. Hence one obtains a map $\oplus : K^{\mathrm{Gro}}(M) \times K^{\mathrm{Gro}}(M) \to K^{\mathrm{Gro}}(M)$. The element $0 := [0, 0]$ obviously serves as neutral element. Associativity and commutativity of the operation $\oplus$ on $K^{\mathrm{Gro}}(M)$ follow, because they hold componentwise. $(K^{\mathrm{Gro}}(M), \oplus, 0)$ is even a group that means additive inverses exist. Namely, since $0 = [m, m]$ for all $m \in M$, the element $[n, m]$ is an additive inverse for $[m, n] \in K^{\mathrm{Gro}}(M)$:

$$[m, n] \oplus [n, m] = [n, m] \oplus [m, n] = [m \oplus n, m \oplus n] = 0 \ .$$

So $(K^{\mathrm{Gro}}(M), \oplus, 0)$ is an abelian group, and the map

$$\kappa : M \to K^{\mathrm{Gro}}(M), \quad m \to [m] := [m, 0]$$

a morphism of monoids.

2.2.18 Proposition *If the abelian monoid $(M, \oplus, 0)$ has the cancellation property, the canonical map $\kappa : M \to K^{\mathrm{Gro}}(M)$, $m \to [m, 0]$ is injective. Moreover, two pairs $(m, n) \in M \times M$ and $(p, q) \in M \times M$ represent the same elements in $K^{\mathrm{Gro}}(M)$ which in other words means $[m, n] = [p, q]$ if and only if $m \oplus q = p \oplus n$.*

Proof. Assume that $(M, \oplus, 0)$ has the cancellation property. We first show that elements $(m, n) \in M \times M$ and $(p, q) \in M \times M$ are then equivalent if and only if $m \oplus q = p \oplus n$. This will also prove the second claim.

If $m \oplus q = p \oplus n$, then (m, n) and (p, q) are obviously equivalent. Vice versa, if $(m, n) \sim (p, q)$, then $m \oplus q \oplus k = p \oplus n \oplus k$ for some $k \in M$, which implies $m \oplus q = p \oplus n$, since M has the cancellation property.

It remains to prove injectivity of κ. So let $[m, 0] = [n, 0]$. Then by what we have shown already $m = m + 0 = n + 0 = n$. Hence κ is injective. □

2.2.19 Proposition and Definition *The abelian group $K^{\mathrm{Gro}}(M) := (M \times M)/\sim$ associated to an abelian monoid $(M, \oplus, 0)$ is a Grothendieck group for M. It is called* the Grothendieck group *of M. Moreover, assigning to each monoid $(M, \oplus, 0)$ the Grothendieck group $K^{\mathrm{Gro}}(M)$ and to each morphism $f : M \to N$ of abelian monoids $(M, \oplus, 0)$ and $(N, +, 0)$ the unique group homomorphism $K^{\mathrm{Gro}}(f)$ making the diagram*

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d] & N \arrow[d] \\ K^{\mathrm{Gro}}(M) \arrow[r, "K^{\mathrm{Gro}}(f)"'] & K^{\mathrm{Gro}}(N) \end{tikzcd}$$

commute, is a functor from the category AbMon *of abelian monoids to the category* Ab *of abelian groups.*

Proof. Let us show that $K^{\mathrm{Gro}}(M)$ together with the morphism of monoid $\kappa : M \to K^{\mathrm{Gro}}(M)$ satisfies the universal property (Gro). So assume that $(A, +, 0)$ is an abelian group and that $f : M \to A$ is a morphism of monoids. Define $f_{K^{\mathrm{Gro}}(M)} : K^{\mathrm{Gro}}(M) \to A$ by

$$[m, n] \mapsto f(m) - f(n).$$

This map is well-defined. To verify this choose $k \in M$ such that $m \oplus q \oplus k = p \oplus n \oplus k$ and compute

$$f(m) + f(q) = f(m \oplus q \oplus k) - f(k) = f(p \oplus n \oplus k) - f(k) = f(p) + f(n).$$

Hence $f(m) - f(n) = f(p) - f(q)$, so $f_{K^{\mathrm{Gro}}(M)}$ is well-defined indeed. The map $f_{K^{\mathrm{Gro}}(M)}$ is a group homomorphism, since

$$\begin{aligned} f_{K^{\mathrm{Gro}}(M)}([m, n] \oplus [p, q]) &= f_{K^{\mathrm{Gro}}(M)}([m \oplus p, n \oplus q]) = f(m \oplus p) - f(n \oplus q) = \\ &= f(m) + f(p) - (f(n) + f(q)) = (f(m) - f(n)) + (f(p) - f(q)) = \\ &= f_{K^{\mathrm{Gro}}(M)}([m, n]) + f_{K^{\mathrm{Gro}}(M)}([p, q]). \end{aligned}$$

Assume that $f' : K^{\mathrm{Gro}}(M) \to A$ is another group homomorphism such that $f' \circ \kappa = f$. Then, for all $[m,n] \in K^{\mathrm{Gro}}(M)$

$$f_{K^{\mathrm{Gro}}(M)}([m,n]) = f(m) - f(n) = f'(\kappa(m)) - f'(\kappa(n)) = f'(\kappa(m) - \kappa(n)) = f'([m,n]),$$

hence $f_{K^{\mathrm{Gro}}(M)} = f'$. This proves that $f_{K^{\mathrm{Gro}}(M)}$ together with κ is a Grothendieck group of M. The remainder of the claim is now an immediate consequence of the universal property (Gro).□

2.2.20 Remark The group operation on the Grothendieck group of an abelian monoid M will usually be denoted by the same symbol like the binary operation on the monoid. Moreover, the image of an element $m \in M$ under the canonical morphism $M \to K^{\mathrm{Gro}}(M)$ will be denoted by $[m]$. If M has the cancellation property, then one just writes m instead of $[m]$ which will not lead to confusion by injectivity of $M \to K^{\mathrm{Gro}}(M)$ in this case.

2.2.21 Proposition *In case $(R, +, \cdot, 0, 1)$ is a semiring, the Grothendieck group $K^{\mathrm{Gro}}(R)$ of the abelian monoid $(R, +, 0)$ carries a natural multiplicative structure given by*

$$\cdot : K^{\mathrm{Gro}}(R) \times K^{\mathrm{Gro}}(R) \to K^{\mathrm{Gro}}(R), \quad [p,q] \cdot [r,s] := [p \cdot r + q \cdot s, p \cdot s + q \cdot r].$$

The Grothendieck group $K^{\mathrm{Gro}}(R)$ - or more precisely $\big(K^{\mathrm{Gro}}(R), +, 0)\big)$ - supplemented by the operation $\cdot$ and the multiplicatively neutral element $[1,0]$ then becomes a ring. It is commutative, if R is.

Proof. The definition of $\cdot$ does not depend on representatives. To verify this assume $(p,q) \sim (p',q')$ and $(r,s) \sim (r',s')$. Then one can find $k, l \in M$ such that $p + q' + k = p' + q + k$ and $r + s' + l = r' + s + l$. Now compute

$$(p \cdot r + q \cdot s) + (p' \cdot s + q' \cdot r) + k \cdot (r + s) = (p' \cdot r + q' \cdot s) + (p \cdot s + q \cdot r) + k \cdot (r + s),$$
$$(p' \cdot r + q' \cdot s) + (p' \cdot s' + q' \cdot r') + (p' + q') \cdot l = (p' \cdot r' + q' \cdot s') + (p' \cdot s + q' \cdot r) + (p' + q') \cdot l,$$

and conclude

$$[p \cdot r + q \cdot s, p \cdot s + q \cdot r] = [p' \cdot r + q' \cdot s, p' \cdot s + q' \cdot r] = [p' \cdot r' + q' \cdot s', p' \cdot s' + q' \cdot r'].$$

So $[p,q] \cdot [r,s]$ is well-defined.

Let us now show that the operation $\cdot : K^{\mathrm{Gro}}(R) \times K^{\mathrm{Gro}}(R) \to K^{\mathrm{Gro}}(R)$ is associative and has $[1,0]$ as neutral element. The latter follows from $[1,0] \cdot [r,s] = [1 \cdot r, 1 \cdot s] = [r,s]$ and $[p,q] \cdot [1,0] = [p \cdot 1, q \cdot 1] = [p,q]$. To prove associativity let $[a,b], [p,q], [r,s] \in K^{\mathrm{Gro}}(R)$ and compute

$$\begin{aligned}
[a,b] \cdot ([p,q] \cdot [r,s]) &= [a,b] \cdot ([p \cdot r + q \cdot s, p \cdot s + q \cdot r]) = \\
&= [a \cdot (p \cdot r + q \cdot s) + b \cdot (p \cdot s + q \cdot r), a \cdot (p \cdot s + q \cdot r) + b \cdot (p \cdot r + q \cdot s)] = \\
&= [(a \cdot p + b \cdot q) \cdot r + (a \cdot q + b \cdot p) \cdot s, (a \cdot p + b \cdot q) \cdot s + (a \cdot q + b \cdot p) \cdot r] = \\
&= [a \cdot p + b \cdot q, a \cdot q + b \cdot p] \cdot [r,s] = ([a,b] \cdot [p,q]) \cdot [r,s].
\end{aligned}$$

The operation $\cdot$ distributes over addition from the left by

$$\begin{aligned}
[a,b] \cdot ([p,q] + [r,s]) &= [a,b] \cdot [p + r, q + s] = \\
&= [a \cdot (p + r) + b \cdot (q + s), b \cdot (p + r) + a \cdot (q + s)] = \\
&= [(a \cdot p + b \cdot q) + (a \cdot r + b \cdot s), (a \cdot q + b \cdot p) + (a \cdot s + b \cdot r)] = \\
&= [a \cdot p + b \cdot q, a \cdot q + b \cdot p] + [a \cdot r + b \cdot s, a \cdot s + b \cdot r] = \\
&= [a,b] \cdot [p,q] + [a,b] \cdot [r,s],
\end{aligned}$$

and from the right by

$$\begin{aligned}([a,b]\ +\ [p,q])\cdot[r,s] &= [a+p,b+q]\cdot[r,s] = \\ &= \big[(a+p)\cdot r+(b+q)\cdot s,(a+p)\cdot s+(b+q)\cdot r\big] = \\ &= \big[(a\cdot r+b\cdot s)+(p\cdot r+q\cdot s),(a\cdot s+b\cdot r)+(p\cdot s+q\cdot r)\big] = \\ &= [a\cdot r+b\cdot s,a\cdot s+b\cdot r]+[p\cdot r+q\cdot s,p\cdot s+q\cdot r] = \\ &= [a,b]\cdot[r,s]+[p,q]\cdot[r,s].\end{aligned}$$

Altogether this shows that $K^{\mathrm{Gro}}(R)$ is a ring. If the semiring R is commutative, then

$$[p,q]\cdot[r,s]=[p\cdot r+q\cdot s,p\cdot s+q\cdot r]=[r\cdot p+s\cdot q,r\cdot q+s\cdot p+]=[r,s]\cdot[p,q]$$

and multiplication in $K^{\mathrm{Gro}}(R)$ is commutative as well. This finishes the proof. □

The ring of integers $\mathbb{Z}$

2.2.22 Definition The *ring of integers* $\mathbb{Z}$ is defined as the Grothendieck ring $K^{\mathrm{Gro}}(\mathbb{N})$ associated to the semiring of natural numbers $\mathbb{N}$.

2.2.23 Theorem and Definition *The set $\mathbb{Z}$ together with addition $+$ and multiplication $\cdot$ as binary operations and the elements $0 := [0,0]$ and $1 := [1,0]$ as neutral elements is a commutative ring. Moreover, the canonical map $\mathbb{N} \hookrightarrow \mathbb{Z}$, $n \mapsto n := [n,0]$ is an embedding of abelian monoids, so $\mathbb{N}$ can be identified with its image $\big\{[n,0] \in \mathbb{Z} \mid n \in \mathbb{N}\big\}$ in $\mathbb{Z}$. Under this identification, $\mathbb{Z}$ is the disjoint union of $\mathbb{N}_{>0}$, $\{0\}$ and $-\mathbb{N}_{>0}$. Elements of $\mathbb{N}_{>0}$ are called* positive integers, *elements of $-\mathbb{N}_{>0}$* negative integers.

Proof. The first claim is a consequence of Proposition 2.2.21. The second claim follows from Proposition 2.2.18. The third claim is essentially a consequence of the trichotomy law for addition, Theorem 2.1.20. Namely, given $[m,n] \in \mathbb{Z}$ there are exactly three cases:

(1) There exists $k \in \mathbb{N}_{>0}$ such that $m = n + k$.

(2) $m = n$.

(3) There exists $l \in \mathbb{N}_{>0}$ such that $n = m + l$.

In the first case the equality $m + 0 = k + n$ holds true which in other words means that $[m,n] = [k,0] = k \in \mathbb{N}_{>0}$. If $m = n$, then $[m,n] = [0,0] = 0 \in \mathbb{N}$. In the third case one has $l + m = 0 + n$ which entails $[m,n] = [0,l] = -l \in -\mathbb{N}_{>0}$. The proof is finished. □

2.2.24 Remark From now on we will identify every natural number $n \in \mathbb{N}$ with its image in $\mathbb{Z}$. In particular we will denote the image of $n \in \mathbb{N}$ in $\mathbb{Z}$ by the same symbol for ease of notation. Further we will denote by $\mathbb{Z}_{\neq 0}$ the set of nonzero integers or in other words the union of $\mathbb{N}_{>0}$ and $-\mathbb{N}_{>0}$. The set $\mathbb{N}_{>0}$ of positive integers is sometimes denoted by $\mathbb{Z}_{>0}$, the set $-\mathbb{N}_{>0}$ of negative integers by $\mathbb{Z}_{<0}$.

2.2.25 Proposition *The product $p \cdot q$ of any non-zero integers p, q is non-zero.*

Proof. Assume first that p and q are elements of $\mathbb{N}_{>0}$. Then both are successors that means $p = s(k)$ and $q = s(l)$ for some $k, l \in \mathbb{N}$. Then

$$p \cdot q = s(k) \cdot s(l) = s(k) \cdot l + s(k) = s\big(s(k) \cdot l + k\big) \in \mathbb{N}_{>0} \ .$$

If $p \in -\mathbb{N}_{>0}$ and $q \in \mathbb{N}_{>0}$, then $-p \in \mathbb{N}_{>0}$ and

$$p \cdot q = -\big((-p) \cdot q\big) \in -\mathbb{N}_{>0} \ .$$

By commutativity of multiplication $p \cdot q = q \cdot p \in -\mathbb{N}_{>0}$ if $p \in \mathbb{N}_{>0}$ and $q \in -\mathbb{N}_{>0}$. Finally, if both p and q are negative, then

$$p \cdot q = (-p) \cdot (-q) \in \mathbb{N}_{>0} \ .$$

Quod erat demonstrandum. □

2.2.26 One can rephrase the preceding proposition by saying that the ring of integers $\mathbb{Z}$ does not have any nontrivial zero divisors. The notion of zero divisors makes sense in any ring, though, even if not commutative. In the following we give the precise definition.

2.2.27 Definition An element l of a ring R is called a *left zero divisor* if there exists an $s \in R \backslash \{0\}$ such that $ls = 0$. An element $r \in R$ is called a *right zero divisor* if there exists an $s \in R \backslash \{0\}$ such that $sr = 0$. If $a \in R$ is both a left and a right zero divisor it is called a *two-sided zero divisor*. In any nonzero ring the zero element is a zero divisor. It is called the *trivial zero divisor*. All other zero divisors are called *nontrivial*. A nonzero ring without any nontrivial zero divisors is called a *domain*. The elements of a ring which are neither left nor right zero divisors are said to be *regular* or are called *non-zero-divisors*.

In the case where R is a commutative ring, the left, right, and both-sided zero divisors are all the same, so one just speaks of *zero divisors* in this case. A nonzero commutative ring with no nontrivial zero divisors is called an *integral domain*.

2.2.28 Remark Using the language from this definition, Proposition 2.2.25 just tells that $\mathbb{Z}$ is an integral domain.

2.2.29 Proposition and Definition *Let $(G, \star, e)$ be a group, $G \to G$, $g \mapsto g^{-1}$ its inverse map, and $g \in G$ an element. Then the map*

$$\mathbb{Z} \to G, \quad p \mapsto \times^p g := \begin{cases} \bigtimes_{i=1}^{p} g & \text{if } p \geqslant 0, \\ \Big(\bigtimes_{i=1}^{-p} g \Big)^{-1} & \text{if } p < 0, \end{cases}$$

is a group homomorphism. If the group law on G is commutative and additively written, one writes $p \cdot g$ or briefly pg instead $\times^p g$, if the group law is multiplicatively written, then one denotes $\times^p g$ by g^p.

Proof. We already know by Proposition 2.1.59 that the map $\mathbb{N} \to G$, $n \mapsto \times^n g$ is a monoid homomorphism. So it only remains to show that $\times^{p+q} g = (\times^p g) \star (\times^q g)$ if p or q is negative. Assume that $p \geqslant 0$, $q < 0$. If $p + q \geqslant 0$, then by definition and Proposition 2.1.59

$$\times^{p+q} g = \times^{p+q} g \star (\times^{-q} g) \star (\times^q g) = (\times^p g) \star (\times^q g) \ .$$

If $p + q < 0$, then

$$\times^{p+q}g = \times^{p+q}g \star (\times^{-q}g) \star (\times^{q}g) = (\times^{p+q}g) \star (\times^{-(p+q)})g \star (\times^{p}g) \star (\times^{q}g) = (\times^{p}g) \star (\times^{q}g) \ .$$

Likewise one proves the case $p < 0$, $q \geqslant 0$. If $p, q < 0$, then

$$\begin{aligned} \times^{p+q}g &= \times^{p+q}g \star (\times^{-q}g) \star (\times^{q}g) = \times^{p+q}g \star (\times^{-q}g) \star (\times^{-p}g) \star (\times^{p}g) \star (\times^{q}g) \\ &= \times^{p+q}g \star \times^{-(p+q)}g \star (\times^{p}g) \star (\times^{q}g) = (\times^{p}g) \star (\times^{q}g) \ . \end{aligned}$$

This proves the claim. □

2.2.30 The ring of integers $\mathbb{Z}$ inherits from $\mathbb{N}$ also an order structure compatible with the algebraic operations.

2.2.31 Theorem and Definition *Given two integers p, q one calls p* less or equal than *q, in signs $p \leqslant q$, if $q - p \in \mathbb{N}$. The relation $\leqslant$ is a total order on $\mathbb{Z}$ which extends the one on $\mathbb{N}$ and satisfies the following monotony axioms:*

(M1) Monotony of addition
For all $p, q \in \mathbb{Z}$ and $a \in \mathbb{Z}$ the relation $p < q$ implies $p + a < q + a$.

(M2) Monotony of multiplication
For all $p, q \in \mathbb{Z}$ and $a \in \mathbb{N}_{>0}$ the relation $p < q$ implies $p \cdot a < q \cdot a$.

Proof. Note first that $p < q$ if and only if $q - p \in \mathbb{N}_{>0}$. Since $\mathbb{Z}$ is the disjoint union of $\mathbb{N}_{>0}$, $\{0\}$ and $-\mathbb{N}_{>0}$, this implies that the trichotomy law holds true for the relation $<$. This implies in particular that $\leqslant$ is total.

Reflexivity of $\leqslant$ is clear since $p - p = 0 \in \mathbb{N}$. Antisymmetry follows from the trichotomy law. If $p \leqslant q$ and $q \leqslant r$, then $q - p$ and $r - q$ are both elements of $\mathbb{N}$, hence $r - p = (r - q) + (q - p)$ is an element of $\mathbb{N}$ as well. So $\leqslant$ is transitive. Thus we have proved that $\leqslant$ is an order relation on $\mathbb{Z}$.

By Definition 2.1.35, the order relation on $\mathbb{Z}$ extends the one on $\mathbb{N}$.

Monotony of addition is trivial, since $q - p \in \mathbb{N}_{>0}$ implies $(q + a) - (p + a) \in \mathbb{N}_{>0}$. Similarly, if $a \in \mathbb{N}_{>0}$, then $q - p \in \mathbb{N}_{>0}$ implies $(q \cdot a) - (p \cdot a) = (q - p) \cdot a \in \mathbb{N}_{>0}$, so monotony of multiplication holds true as well. The proof is finished. □

Ordered commutative rings and integral domains

2.2.32 Definition A commutative ring $(R, +, \cdot, 0, 1)$ together with a total order $\leqslant$ on it is called an *ordered commutative ring* if the following weak monotony properties hold true:

(M1)' Weak monotony of addition
For all $r, s \in R$ and $a \in R$ the relation $r \leqslant s$ implies $r + a \leqslant s + a$.

(M2)' Weak monotony of multiplication
For all $r, sq \in R$ and $a \in R$ with $a \geqslant 0$ the relation $r \leqslant s$ implies $r \cdot a \leqslant s \cdot a$.

If R is an ordered commutative ring and does not have any non-trivial zero divisors, we call R - or better $(R, +, \cdot, 0, 1, \leqslant)$ - an ordered integral domain.

Given an ordered ring R, elements of $R_{>0} := \{x \in R \mid x > 0\}$ are called *positive*, elements of $R_{<0} := \{x \in R \mid x < 0\}$ *negative*. The set $R_{\geqslant 0} := \{x \in R \mid x \geqslant 0\}$ is the set of *nonnegative* elements of R, the set $R_{\leqslant 0} := \{x \in R \mid x \leqslant 0\}$ the set of its *nonpositive* elements.

2.2.33 Proposition *A commutative ring $(R, +, \cdot, 0, 1)$ on which a total order relation $\leqslant$ is given is an ordered commutative ring if the two monotony axioms* (M1) *and* (M2) *from Theorem 2.2.31 are satisfied with $\mathbb{Z}$ replaced by R and $\mathbb{N}_{>0}$ replaced by $R_{>0}$. If R is an integral domain, the converse holds true as well.*

Proof. To verify this assume first that $(R, +, \cdot, 0, 1, \leqslant)$ is a commutative ring and $\leqslant$ an order relation satisfying (M1) and (M2). Let $r, s, a \in R$ and assume $r \leqslant s$. If $r = s$, then $r + a = s + a$. If $r < s$ then $r + a < s + a$ by (M1). This proves (M1)'. Now assume in addition that $a \geqslant 0$. Then the inequality $r \leqslant s$ implies $r \cdot a = r \cdot a$ in case $a = 0$ or $p = q$ and, by (M2), $r \cdot a < s \cdot a$ if $a > 0$ and $p \neq q$. So (M2)' holds as well.

Now assume that $(R, +, \cdot, 0, 1, \leqslant)$ is an ordered ring. Assume that $r < s$ and let $a \in R$. Then $r + a \leqslant s + a$. □

2.2.34 Example By Theorem 2.2.31 and Proposition 2.2.33, the ring of integers is an example of an ordered integral domain. We will encounter several more examples, in particular the ordered fields $\mathbb{Q}$ and $\mathbb{R}$ of rational and real numbers, respectively.

2.2.35 Remark In $\mathbb{Z}$ and any integral domain, the monotony axioms (M1) and (M2) are equivalent to (M1)' and (M2)', respectively. When dealing with ordered integral domains we therefore will refer to both properties (M1) and (M1)' as monotony of addition, and to both (M2) and (M2)' as monotony of multiplication.

2.2.36 Proposition *The following holds true for integers p, q and r or, more generally, for elements p, q, and r of an ordered integral domain R:*

(i) *One has $1 > 0$ and $-1 < 0$.*

(ii) *If $0 < p$, then $-p < 0$. If $q < 0$, then $0 < -q$.*

(iii) *If $p < q$ and $r < 0$, then $rq < rp$.*

(iv) *If $p \neq 0$, then $p^2 > 0$.*

(v) *If $0 \leqslant p$ and $0 \leqslant q$, then the inequality $p < q$ is equivalent to $p^2 < q^2$.*

(vi) *If p is invertible that is if there is an element $p^{-1} \in R$ such that $p \cdot p^{-1} = 1$, then p is nonzero. Moreover, p is positive (respectively negative) if and only if p^{-1} is.*

(vii) *If both p and q are invertible and positive, then the inequality $p < q$ is equivalent to $\frac{1}{q} < \frac{1}{p}$.*

Proof. ad (ii). By monotony of addition, adding $-p$ to both sides of the inequality $0 < p$ entails $-p < 0$. Adding $-q$ to both sides of $q < 0$ gives $0 < -q$.

ad (iii). Since $-r > 0$, the inequality $p < q$ entails $-rp < -rq$. Adding $rp + rq$ to both sides proves the claimed inequality by monotony of addition.

ad (iv). For $p > 0$ the claim is clear by monotony of multiplication. So assume $p < 0$. Then $-p > 0$, hence $p^2 = (-1)^2 \cdot p^2 = (-p)^2 > 0$.

ad (i). Since in any integral domain $1 \neq 0$, one obtains $1 = 1^2 > 0$ by (iv). By (ii), $-1 < 0$ follows.

ad (v). By monotony of multiplication and transitivity, $p < q$ implies $p^2 \leqslant pq < q^2$ whereas $p \geqslant q$ entails $p^2 \geqslant pq \geqslant q^2$.

ad (vi). Since R is an integral domain, p is nonzero, because otherwise $1 = p \cdot p^{-1} = 0$. By trichotomy, p therefore is either positive or negative. If p is positive, then $p^{-1} \leqslant 0$ implies $1 \leqslant 0$ which can not be true. So $p^{-1} > 0$ if $p > 0$. Now recall that inverses of invertible elements are uniquely determined by (ii) in Proposition 2.2.7. Therefore, p^{-1} is invertible with inverse p, hence $p^{-1} > 0$ implies $p > 0$. The remainder of the claim now follows from trichotomy.

ad (vii). By (vi), both p^{-1} and q^{-1} are positive, hence $p < q$ implies by monotony of multiplication $q^{-1} = p^{-1}\,p\,q^{-1} < p^{-1}\,q\,q^{-1} = q^{-1}$. □

2.2.37 Definition Let $(R, +, \cdot, 0, 1, \leqslant)$ be an ordered commutative ring. Then the *abstract value function* on R is defined as follows:

$$|\cdot| : R \to R, \quad r \mapsto \begin{cases} r & \text{if } r > 0, \\ 0 & \text{if } r = 0, \\ -r & \text{if } r < 0\ . \end{cases}$$

One calls $|r|$ the *abstract value* of $r \in R$.

2.2.38 Proposition *Let $(R, +, \cdot, 0, 1, \leqslant)$ be an ordered commutative ring and consider its abstract value function $|\cdot| : R \to R$. Then the following holds true for all $r, s \in R$:*

(i) $|r| \geqslant 0$ *and* $|r| = 0$ *if and only if* $r = 0$.

(ii) *If* $s \geqslant 0$ *and* $-s \leqslant r \leqslant s$, *then* $|r| \leqslant s$. *Moreover,* $-|r| \leqslant r \leqslant |r|$.

(iii) Triangle Inequality: $|r + s| \leqslant |r| + |s|$.

(iv) Reverse Triangle Inequality: $\big||r| - |s|\big| \leqslant |r - s|$.

(v) $|r \cdot s| = |r| \cdot |s|$.

(vi) $|-r| = |r|$.

(vii) *If r is invertible, then* $\left|\frac{1}{r}\right| = \frac{1}{|r|}$.

Proof. ad (i). The first claim follows by definition, Proposition 2.2.36 (ii) and since $\mathbb{Z}$ is the disjoint union of $\mathbb{Z}_{>0}$, $\{0\}$ and $\mathbb{Z}_{<0}$.

ad (ii). The second claim follows for $r \geqslant 0$ by $-r \leqslant 0 \leqslant r = |r|$ and for $r < 0$ by $-|r| = r < 0 < -r = |r|$. Now assume that $-s \leqslant r \leqslant s$ for some $s \geqslant 0$. If $r \geqslant 0$, then $|r| = r \leqslant s$. If $r < 0$, then $|r| = -r \leqslant s$ by assumption and Proposition 2.2.36 (i) and (iii). So the first claim is proved as well.

ad (iii). One has $-|r| - |s| \leqslant r + s \leqslant |r| + |s|$ by (ii) and monotony of addition. Applying (ii) again gives the claim.

ad (iv). Douple application of the triangle inequality and monotony of addition entails $-|r-s| \leqslant |r| - |s| \leqslant |r - s|$, hence the reverse triangle inequality follows by (ii).

ad (v). If $r, s \geqslant 0$, then $r \cdot s \geqslant 0$ by monotony of multiplication, hence $|r \cdot s| = r \cdot s = |r| \cdot |s|$. If $r < 0$ and $s \geqslant 0$ or $r \geqslant 0$ and $s < 0$, then $r \cdot s \leqslant 0$, hence $|r \cdot s| = -(r \cdot s) = |r| \cdot |s|$. If both r and s are negative, then $r \cdot s > 0$, hence $|r \cdot s| = r \cdot s = (-1)^2 \cdot r \cdot s = (-r) \cdot (-s) = |r| \cdot |s|$.

ad (vi). This follows from the preceding relation and $|-1| = 1$.

ad (vii). The last equality is a consequence of $|r| \cdot \left|\frac{1}{r}\right| = \left|\frac{r}{r}\right| = |1| = 1$. □

2.3. Arithmetic in $\mathbb{Z}$

The fundamental theorem of arithmetic

2.3.1 Theorem (The Division Theorem) *Let a, b be integers with $b \neq 0$. Then there exist uniquely determined $q, r \in \mathbb{Z}$ such that $a = qb + r$ and $0 \leqslant r < |b|$.*

Proof. Consider the set $M = \{m \in \mathbb{N} \mid \exists q \in \mathbb{Z} : \ m = a - qb\}$. This set is non-empty, since by the archimedean property for $\mathbb{Z}$ there exists a natural number n such that $n|b| > -a$. If $b > 0$ put $q = -n$ otherwise let $q = n$. In either case $a - qb = a + n|b| > 0$, hence $a - qb \in M$. Since $\mathbb{N}$ is well-ordered, the set M has a least element which we denote by r. Let $q \in \mathbb{Z}$ such that $r = a - qb$. Assume that $r \geqslant |b|$. Then $s = a - (q + \operatorname{sgn}(b))b = r - \operatorname{sgn}(b)b = r - |b| \in \mathbb{N}$, where $\operatorname{sgn}(b)$ is defined to be equal to 1 if $b > 0$ and to be -1 else. Hence $s \in M$ and $s < r$ which contradicts the minimality of r. Therefore $r < |b|$.

Now assume that q', r' are a second pair of integers such that $a = q'b + r'$ and $0 \leqslant r' < |b|$. Then $\big|(q - q')b\big| = |r' - r| < |b|$ and $\big|(q - q')b\big| = |q - q'| \cdot |b|$. This can only be possible if $|q - q'| = 0$ which entails $q' = q$ and $r' = r$. The proof is finished. □

2.3.2 Definition If r, s are integers or more generally elements of an integral domain R, then r is said to *divide* s or that r is a *divisor* of s if there exists some element q of $\mathbb{Z}$ respectively R such that $q \cdot r = s$. One also says in this situation that s is a *multiple* of r and denotes it by writing $r \mid s$. If r does not divide s, one denotes this by $r \nmid s$.

2.3.3 Lemma *If r is a divisor of the non-zero integer s, then $|r| \leqslant |s|$.*

Proof. By definition there exists an integer q such that $q \cdot r = s$. Then q must be non-zero and $|q| \cdot |r| = |s|$. Hence $|q| \geqslant 1$ and $|r| \leqslant |s|$. □

2.3.4 Every natural number larger than 1 has at least two positive divisors, namely 1 and the number itself. The natural numbers $n > 1$ for which these two divisors are all the divisors of n have a particular name, they are called *prime numbers.*

2.3.5 Definition Let $r_1, \ldots, r_k$ with $k \in \mathbb{N}_{>0}$ be integers. A natural number d is called *greatest common divisor* of $r_1, \ldots, r_k$ if d is a divisor of all the r_j and if any natural number which is a divisor of all the r_j is a divisor of d. If 1 is a greatest common divisor of $r_1, \ldots, r_k$, then one calls $r_1, \ldots, r_k$ *coprime* or *relatively prime.*

2.3.6 Proposition and Definition *The greatest common divisor of integers $r_1, \ldots, r_k$ exists and is uniquely determined. It is denoted* $\gcd(r_1, \ldots, r_k)$.

Proof. First assume that $r_1 = \ldots = r_k = 0$. Then any natural number is a divisor of each of the r_j. In particular 0 is a divisor of all the r_j and is divided by all natural numbers. Since 0 is the only natural number with that property 0 is a greatest common divisor of $r_1, \ldots, r_k$ and it is the only one.

Now assume that at least one of the r_j is non-zero. Let us first prove uniqueness in this situation and assume that d and d' are greatest common divisors of $r_1, \ldots, r_k$. Since 0 is not a divisor of the non-zero elements of the r_j both d and d' have to be positive. By definition of greatest common divisors d divides d' and vice versa. Hence there are $n, m \in \mathbb{N}_{>0}$ such that $d = nd'$ and $d' = nd$. But then $d = nmd$ which implies $nm = 1$ since d is non-zero. This means that $n = m = 1$ and $d = d'$.

Next we prove existence. Let L be the set of all positive integers which can be written in the form $x_1 r_1 + \ldots + x_k r_k$ for integers $x_1, \ldots, x_k$. The set L is non-empty since it contains $r_1^2 + \ldots + r_k^2$. Let d be its smallest element. Clearly, if c divides all the r_j, then c divides d. So it remains to show that d divides all the r_j. By the the Division Theorem 2.3.1 there exist $a_j, b_j \in \mathbb{Z}$, $j = 1, \ldots, k$ with $0 \leqslant b_j < d$ such that $r_j = a_j \cdot d + b_j$. Assume that $b_i > 0$. Then

$$b_i = r_i - a_i \cdot d = (1 - a_i x_i) r_i - \sum_{j \neq i} (a_i x_j) r_j,$$

where the $x_j \in \mathbb{Z}$ are chosen so that $d = x_1 r_1 + \ldots + x_k r_k$. But this means that $b_i \in L$ which contradicts the minimality of d. Hence $b_j = 0$ for $j = 1, \ldots, k$, and d divides all r_j. $\square$

2.3.7 Example If p is a prime number and r an integer, then p is a divisor of r if and only if $\gcd(p, r) = p$. Otherwise $\gcd(p, r) = 1$. The greatest common divisor of 0 and an integer r is $|r|$, the greatest common divisor of 1 and an integer r is 1.

2.3.8 With the proof of Proposition 2.3.6 we have also shown the following result.

2.3.9 Lemma (Bézout's Lemma) *A natural number is the greatest common divisor of integers $r_1, \ldots, r_k$ if and only if it is the smallest positive natural number d for which there exist $x_1, \ldots, x_k \in \mathbb{Z}$ such that*

$$d = x_1 r_1 + \ldots + x_k r_k \ .$$

2.3.10 Lemma (Euclid's Lemma) *If a prime number p divides the product rs of two integers r and s, then p divides at least one of the integers r and s.*

Proof. If p divides r we are done. So assume that p does not divide r. Then $\gcd(p,r) = 1$, hence by Bezout's Lemma 2.3.9 there exist $x, y \in \mathbb{Z}$ such that $xp + yr = 1$. Multiplication by s gives $s = xsp + yrs$. The right side is divisible by p, hence p divides s. □

2.4. Rational numbers

2.4.1 Even though the integers form an abelian group with respect to addition, multiplicative inverses of integers $n \neq 1, -1$ do not exist in $\mathbb{Z}$. The argument is as follows. First observe that 0 is not multiplicatively invertible in $\mathbb{Z}$ and even not in any extension ring R of $\mathbb{Z}$ because if it were with inverse m, then $1 = 0 \cdot m = 0$ and $r = 1 \cdot r = 0 \cdot r = 0$ for all $r \in R$. But this contradicts that R is assumed to be an extension ring of $\mathbb{Z}$, which, to remind the reader, is a ring R in which $\mathbb{Z}$ is embedded by an injective ring homomorphism $\mathbb{Z} \hookrightarrow R$. To verify that also every integer $n \neq 0, 1, -1$ does not have a multiplicative inverse in $\mathbb{Z}$, assume that $m \in \mathbb{Z}$ is one, i.e. that $n \cdot m = m \cdot n = 1$. If $n > 0$, then $m > 0$ as well, since otherwise $n \cdot (-m) = -1 < 0$ and $n \cdot (-m) > 0$ by monotony of multiplication which is a contradiction. But then $m \geqslant 1$ and $n > 1$, hence $n \cdot m > 1$ which contradicts the assumption that m is a multiplicative inverse of n. If $n < 0$, then $-m$ is a multiplicative inverse of $-n$ which is strictly positive and which we already have ruled out to have a multiplicative inverse. So the elements of $\mathbb{Z}\backslash\{1, -1\}$ are all not invertible in $\mathbb{Z}$.

The ring (or better field as we will later see) of rational numbers $\mathbb{Q}$ will be defined as the minimal extension of $\mathbb{Z}$ in which all non-zero integers are multiplicatively invertible. The construction of $\mathbb{Q}$ is via *localization* by the set of non-zero integers meaning by forming abstract quotients, called fractions, of integers by non-zero ones. The process resembles the construction of the Grothendieck group, but it is not the same since we do not want to invert the zero element of $\mathbb{Z}$. The symbol $\mathbb{Q}$ for the field of rational numbers goes back to Giuseppe Peano who introduced it 1895 after *quoziente*, the Italian word for *quotient*.

Localization

2.4.2 Definition A subset S of a commutative ring R is called *multiplicative* if it contains 1 and if for all $r, s \in S$ the product rs is in S again.

Let R be a commutative ring, and $S \subset R$ a multiplicative subset. On the cartesian product $R \times S$ we introduce an equivalence relation as follows. Two pairs $(p,q), (r,s) \in R \times S$ are called equivalent, in signs $(p,q) \sim (r,s)$, if $pst = rqt$ for some $t \in S$. Obviously, $\sim$ is reflexive and symmetric. To verify transitivity, assume that $(a,b) \sim (p,q)$ and $(p,q) \sim (r,s)$. Choose $d, t \in S$ such that $aqd = pbd$ and $pst = rqt$. Then

$$as(qdt) = aqd(st) = pbd(st) = pst(bd) = rqt(bd) = rb(qdt) ,$$

hence $(a,b) \sim (r,s)$, so $\sim$ is transitive and an equivalence relation indeed. We denote the equivalence class of (p,q) by

$$\frac{p}{q}$$

and call it the *abstract quotient* or *fraction* of p by q. The set of fractions $\frac{p}{q}$ with $p \in R$, $q \in S$ is denoted by $S^{-1}R$ and called the *localization* of R by S.

2.4.3 Lemma *For every element t of a multiplicative subset S of a commutative ring R and every element $\frac{p}{q} \in S^{-1}R$ the fractions $\frac{p}{q}$ and $\frac{pt}{qt}$ coincide.*

Proof. This is clear since $p\,qt = pt\,q$. □

2.4.4 Proposition *Let R be a commutative ring and $S \subset R$ a multiplicative subset.*

(i) *The localization $S^{-1}R$ carries a ring structure with addition*

$$+ : S^{-1}R \times S^{-1}R \to S^{-1}R, \quad \left(\frac{p}{q}, \frac{r}{s}\right) \mapsto \frac{ps + rq}{qs} \ ,$$

multiplication

$$\cdot : S^{-1}R \times S^{-1}R \to S^{-1}R, \quad \left(\frac{p}{q}, \frac{r}{s}\right) \mapsto \frac{pq}{qs} \ ,$$

zero element $0 := \frac{0}{1}$ and multiplicative identity $1 := \frac{1}{1}$.

(ii) *For every $s \in S$, the fraction $\frac{1}{s}$ is the inverse of $\frac{s}{1}$ in $S^{-1}R$.*

(iii) *The map $R \to S^{-1}R$, $r \mapsto \frac{r}{1}$ is a ring homomorphism.*

(iv) *In case S does not contain any zero divisors that is if $sr \neq 0$ for all $s \in S$ and $r \in R_{\neq 0}$, then the canonical ring homomorphism $R \to S^{-1}R$ is injective. Moreover in this case, two fractions $\frac{p}{q}$ and $\frac{r}{s}$ are identical if and only if $ps = rq$.*

Proof. ad (i). First one needs to verify that addition and multiplication are well-defined. To this end let $\frac{p}{q} = \frac{p'}{q'}$ and $\frac{r}{s} = \frac{r'}{s'}$. This means there are $t, u \in S$ such that $pq't = p'qt$ and $rs'u = r'su$. Then

$$(ps + rq)\,q's'\,tu = pq't(ss'u) + rs'u(qq't) = p'qt(ss'u) + r'su(qq't) = (p's' + r'q')\,qs\,tu \ ,$$

so the sum of two fractions is well-defined. Next

$$pr\,q's'\,ut = (pq't)(rs'u) = (p'qt)(r'su) = p'r'\,qs\,ut \ ,$$

so the product of two fractions is well-defined as well.

To prove associativity of addition compute

$$\left(\frac{a}{b} + \frac{p}{q}\right) + \frac{r}{s} = \frac{aq + pb}{bq} + \frac{r}{s} = \frac{(aq + pb)s + rbq}{bqs} = \\ = \frac{aqs + (ps + rq)b}{bqs} = \frac{a}{b} + \frac{ps + rq}{qs} = \frac{a}{b} + \left(\frac{p}{q} + \frac{r}{s}\right) .$$

Since R is a commutative ring,

$$\frac{p}{q} + \frac{r}{s} = \frac{ps + rq}{qs} = \frac{rq + ps}{sq} = \frac{r}{s} + \frac{p}{q} \ ,$$

hence addition in $S^{-1}R$ is commutative. The fraction $\frac{0}{1}$ acts as neutral element by addition:

$$\frac{0}{1} + \frac{p}{q} = \frac{0 \cdot q + p \cdot 1}{1 \cdot q} = \frac{p}{q} \ .$$

And the additive inverse of $\frac{p}{q}$ is given by the fraction $\frac{-p}{q}$:

$$\frac{p}{q} + \frac{-p}{q} = \frac{pq + (-p)q}{q^2} = \frac{(p-p)q}{q^2} = \frac{0}{q^2} = \frac{0}{1} \ .$$

So we have shown that $S^{-1}R$ with addition is an abelian group. Let us consider multiplication now. Multiplication in $S^{-1}R$ is obviously associative since it is in R. More precisely,

$$\left(\frac{a}{b} \cdot \frac{p}{q}\right) \cdot \frac{r}{s} = \frac{ap}{bq} \cdot \frac{r}{s} = \frac{(ap)r}{(bq)s} = \frac{a(pr)}{b(qs)} = \frac{a}{b} \cdot \frac{pr}{qs} = \frac{a}{b} \cdot \left(\frac{p}{q} \cdot \frac{r}{s}\right) .$$

Similarly, multiplication in $S^{-1}R$ is commutative:

$$\frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs} = \frac{rp}{sq} = \frac{r}{s} \cdot \frac{p}{q} \ .$$

The element $\frac{1}{1}$ acts neutrally by multiplication, since

$$\frac{1}{1} \cdot \frac{p}{q} = \frac{1 \cdot p}{1 \cdot q} = \frac{p}{q} \ .$$

Note that $\frac{1}{1} = \frac{t}{t}$ for all $t \in S$ since $1 \cdot t = t \cdot 1$. Using this, one proves that multiplication distributes over addition:

$$\begin{aligned} \frac{a}{b} \cdot \left(\frac{p}{q} + \frac{r}{s}\right) &= \frac{a}{b} \cdot \frac{ps + rq}{qs} = \frac{a(ps+rq)}{bqs} = \frac{b}{b} \cdot \frac{a(ps+rq)}{bqs} = \\ &= \frac{aps + arq}{bqs} = \frac{ap\,bs + ar\,bq}{bq\,bs} = \frac{ap}{bq} + \frac{ar}{bs} = \frac{a}{b} \cdot \frac{p}{q} + \frac{a}{b} \cdot \frac{r}{s} \ . \end{aligned}$$

We have verified that $S^{-1}R$ is a commutative ring.

ad (*ii*). Since for every $s \in S$ the relation

$$\lambda(s) \cdot \frac{1}{s} = \frac{s}{1} \cdot \frac{1}{s} = \frac{s}{s} = 1$$

holds true, $\frac{1}{s}$ is the inverse of $\lambda(s)$.

ad (*iii*). The canonical mapping $\lambda : R \to S^{-1}R$, $r \mapsto \frac{r}{1}$ is a ring homomorphism, since

$$\frac{r}{1} + \frac{s}{1} = \frac{r+s}{1}, \qquad \frac{r}{1} \cdot \frac{s}{1} = \frac{rs}{1},$$

and since $\frac{0}{1}$ and $\frac{1}{1}$ are neutral with respect to addition and multiplication, respectively.

ad (*iv*). Now we show that λ is injective if S does not contain any zero divisors. Assume that $\lambda(r) = \lambda(s)$. Then $\lambda(r-s) = 0$, hence there exists a $t \in S$ such that $(r-s)t = 0$. But since S does not have any zero divisors $r = s$, so λ is injective.

Finally consider the equation $\frac{p}{q} = \frac{r}{s}$. It is equivalent to $pst = rqt$ for some $t \in S$, hence to $(ps - rq)t = 0$. If S does not have zero divisors, the latter is equivalent to $ps = rq$. This finishes the proof of the proposition. □

2.4.5 Remark In the case where the multiplicative subset S of the commutative ring R does not contain any zero divisors, one identifies R with its image in $S^{-1}R$ and denotes every element in $S^{-1}R$ of the form $\frac{r}{1}$ just by r.

2.4.6 Theorem *Let R be a commutative ring and $S \subset R$ a multiplicative subset. Then the localization $S^{-1}R$ fulfills the following universal property:*

(Loc) *Let A be a ring and $f : R \to A$ a ring homomorphism such that for every $s \in S$ the element $f(s)$ is invertible in A. Then there exists a unique ring homomorphism $f_{S^{-1}R} : S^{-1}R \to A$ which makes the diagram*

$$\begin{tikzcd} R \arrow[r, "f"] \arrow[d] & A \\ S^{-1}R \arrow[ur, "f_{S^{-1}R}"'] & \end{tikzcd}$$

commute.

Proof. Observe that for all $r, s \in R$ the equality

$$f(s)f(r) = f(rs) = f(sr) = f(r)f(s)$$

holds true. Hence $f(s)^{-1}f(r) = f(r)f(s)^{-1}$ if $s \in S$. If $p, r \in R$, $q, s \in S$ and $(p, q) \sim (r, s)$, choose some $t \in S$ such that $pst = rqt$. This entails $f(p)f(s)f(t) = f(r)f(q)f(t)$ and, by invertibility of $f(t)$ in A, $f(q)^{-1}f(p) = f(s)^{-1}f(r)$. Hence the map

$$f_{S^{-1}R} : S^{-1} \to A, \ \frac{r}{s} \mapsto f(s)^{-1}f(r)$$

is well-defined. Moreover, it is a ring homomorphism by the following calculations:

$$\begin{aligned}
f_{S^{-1}R}(0) = f_{S^{-1}R}\left(\frac{0}{1}\right) &= f(1)^{-1}f(0) = 0 \ , \\
f_{S^{-1}R}(1) = f_{S^{-1}R}\left(\frac{1}{1}\right) &= f(1)^{-1}f(1) = 1 \ , \\
f_{S^{-1}R}\left(\frac{r}{s}\right) + f_{S^{-1}R}\left(\frac{p}{q}\right) &= f(s)^{-1}f(r) + f(q)^{-1}f(p) = \\
&= f(sq)^{-1}\big(f(q)f(r) + f(s)f(p)\big) = f_{S^{-1}R}\left(\frac{qr+sp}{sq}\right) = f_{S^{-1}R}\left(\frac{r}{s} + \frac{p}{q}\right) \ , \\
f_{S^{-1}R}\left(\frac{r}{s}\right) \cdot f_{S^{-1}R}\left(\frac{p}{q}\right) &= f(s)^{-1}f(r)f(q)^{-1}f(p) = f(sq)^{-1}f(rp) = f_{S^{-1}R}\left(\frac{r}{s} \cdot \frac{p}{q}\right) \ .
\end{aligned}$$

This proves the existence claim. To verify uniqueness, let $\overline{f} : S^{-1}R \to A$ be another ring homomorphism such that the composition $R \to S^{-1}R \xrightarrow{\overline{f}} A$ coincides with f. Then compute for $r \in R$ and $s \in S$

$$\overline{f}\left(\frac{r}{s}\right) = \overline{f}(s)^{-1}\overline{f}(r) = f(s)^{-1}f(r) = f_{S^{-1}R}\left(\frac{r}{s}\right) \ .$$

So $\overline{f} = f_{S^{-1}R}$ and $f_{S^{-1}R}$ is uniquely determined. □

The field of rational numbers $\mathbb{Q}$

By Proposition 2.2.25, the set $S := \mathbb{Z}_{\neq 0}$ of non-zero integers is multiplicative. Hence the following definition makes sense.

2.4.7 Definition The *field of rational numbers* $\mathbb{Q}$ is defined as the localization $S^{-1}\mathbb{Z}$ of $\mathbb{Z}$ by the multiplicative set S of non-zero integers.

2.4.8 Theorem and Definition *The set $\mathbb{Q}$ together with addition $+$ and multiplication $\cdot$ as binary operations and the elements $0 := \frac{0}{1}$ and $1 := \frac{1}{1}$ as neutral elements is a* field *which means that the following axioms hold true:*

(Fld1) *$\mathbb{Q}$ together with addition $+$ and the element 0 is an abelian group.*

(Fld2) *$\mathbb{Q}$ together with multiplication $\cdot$ and the element 1 is an abelian monoid such that every element $\mathbb{Q}_{\neq 0} := \mathbb{Q}\backslash\{0\}$ has a multiplicative inverse.*

(Fld3) *Multiplication distributes from the left and the right over addition.*

(Fld4) *The neutral elements 0 and 1 are not equal.*

Moreover, the canonical map $\mathbb{Z} \hookrightarrow \mathbb{Q}$, $p \mapsto n := \frac{p}{1}$ is an injective ring homomorphism, so $\mathbb{Z}$ can be identified with its image $\left\{\frac{n}{1} \in \mathbb{Q} \mid n \in \mathbb{Z}\right\}$ in $\mathbb{Q}$.

Proof. By Proposition 2.4.4 we know that $\mathbb{Q}$ is a commutative ring and that $\mathbb{Z} \hookrightarrow \mathbb{Q}$, $p \to \frac{p}{1}$ an injective ring homomorphism. In particular this verifies Axioms (Fld1) and (Fld3) and that $\mathbb{Q}$ with multiplication $\cdot$ and the element 1 is an abelian monoid. Since $0 \neq 1$ in $\mathbb{Z}$ (because $\mathbb{N} \hookrightarrow \mathbb{Z}$ is injective, $1 = s(0)$ and 0 is not in the image of the successor map $s : \mathbb{N} \to \mathbb{N}$), Axiom (Fld4) holds true. It remains to show that every non-zero element of $\mathbb{Q}$ has a multiplicative inverse. So let $\frac{p}{q} \in \mathbb{Q}$ be non-zero. Then p and q are both in $\mathbb{Z}_{\neq 0}$. The element $\frac{q}{p}$ now is the multiplicative inverse of $\frac{p}{q}$. This finishes the proof. □

2.4.9 Definition A set $\mathbb{F}$ equipped with binary operations $+$ and $\cdot$ on $\mathbb{F}$ and two elements $0, 1 \in \mathbb{F}$ is called a *field* if Axioms (Fld1) to (Fld4) above are satisfied (after replacing $\mathbb{Q}$ by $\mathbb{F}$). If in addition $\leqslant$ is a total order relation on $\mathbb{F}$ such that the monotony axioms below are satisfied as well, then $\mathbb{F}$, or more precisley $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ is called an *ordered field*:

(M1) Monotony of addition
For all $a, b \in \mathbb{F}$ and $c \in \mathbb{F}$ the relation $a < b$ implies $a + c < b + c$.

(M2) Monotony of multiplication
For all $a, b \in \mathbb{F}$ and $c \in \mathbb{F}_{>0} := \{x \in \mathbb{F} \mid x > 0\}$ the relation $a < b$ implies $a \cdot c < b \cdot c$.

2.4.10 Definition A rational number $\frac{p}{q} \in \mathbb{Q}$ is called *less or equal* than a rational number $\frac{r}{s} \in \mathbb{Q}$, in signs $\frac{p}{q} \leqslant \frac{r}{s}$, if the integer $(rq - ps)qs$ lies in $\mathbb{N}$.

2.4.11 Theorem *The field $\mathbb{Q}$ together with the binary relation $\leqslant$ is an ordered field. Moreover, the canonical embedding $\mathbb{Z} \hookrightarrow \mathbb{Q}$, $p \mapsto \frac{p}{1}$ is order preserving which means that $p < q$ for $p, q \in \mathbb{Z}$ implies $\frac{p}{1} < \frac{q}{1}$.*

Proof. Obviously, $\leqslant$ is reflexive, since for $\frac{p}{q} \in \mathbb{Q}$ the integer $(pq - pq)q^2$ vanishes.

If $\frac{p}{q} \leqslant \frac{r}{s}$ and $\frac{r}{s} \leqslant \frac{p}{q}$, then both $(rq - ps)qs$ and its additive inverse $(ps - rq)qs$ are elements of $\mathbb{N}$, hence $(rq - ps)qs = 0$ which implies that $\frac{r}{s} - \frac{p}{q} = \frac{rq-ps}{qs} = \frac{(rq-ps)qs}{(qs)^2} = 0$. So $\leqslant$ is antisymmetric.

If $\frac{a}{b} \leqslant \frac{p}{q}$ and $\frac{p}{q} \leqslant \frac{r}{s}$, then $(pb - aq)bq \geqslant 0$ and $(rq - ps)qs \geqslant 0$. Hence $\big((rb - as)bs\big)qq = (rbq - aqs)bqs = (rq - ps)qsb^2 + (pb - aq)bqs^2 > 0$, since $b^2 > 0$ and $s^2 > 0$. Since $q^2 > 0$, $(rb - as)bs$ follows by monotony of multiplication in $\mathbb{Z}$. This proves that $\leqslant$ is transitive. So we have shown that $\leqslant$ is an order relation on $\mathbb{Q}$.

Since for two rational numbers $\frac{p}{q}$ and $\frac{r}{s}$ the integer $(rq-ps)qs$ is either positive, zero, or negative, the trichotomy law holds true, and $\leqslant$ is a total order.

Note that the relation $\frac{r}{s} > 0$ is equivalent to $rs > 0$. Hence, if $\frac{a}{b} < \frac{p}{q}$ and $\frac{r}{s} > 0$, then $\frac{ar}{bs} < \frac{pr}{qs}$ since $(prbs - arqs)bsqs = \big((pb - aq)bq\big)(rs)s^2 > 0$. Therefore, monotony of multiplication holds. Now let $\frac{r}{s} \in \mathbb{Q}$ be arbitrary and $\frac{a}{b} < \frac{p}{q}$ as before. Then

$$\frac{a}{b} + \frac{r}{s} = \frac{as + rb}{sb} \quad \text{and} \quad \frac{p}{q} + \frac{r}{s} = \frac{ps + rq}{sq} \ .$$

Now compute

$$\big((ps + rq)sb - (as + rb)(sq)\big)sbsq = \big((pb - aq)bq\big)s^4 > 0 \ .$$

Hence $\frac{a}{b} + \frac{r}{s} < \frac{p}{q} + \frac{r}{s}$, and monotony of addition is finally proved as well.

Finally assume $p < q$ for integers p, q. Then $(q \cdot 1 - p \cdot 1) \cdot 1^2 = p - q > 0$, hence $\frac{p}{1} \leqslant \frac{q}{1}$. But $p \neq q$ since the ring homomorphism $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is injective. So one even has $\frac{p}{1} < \frac{q}{1}$ as claimed. $\square$

2.4.12 Theorem *The field of rational numbers $\mathbb{Q}$ is an archimedean ordered field that is for every pair of rational numbers $\frac{p}{q}$, $\frac{r}{s}$ with $\frac{r}{s} > 0$ there exists a natural number $n \in \mathbb{N}$ such that $\frac{p}{q} < n\,\frac{r}{s}$.*

Proof. By Proposition 2.2.36 (vi) and since $\frac{r}{s} > 0$, the claim is equivalent to the existence of a natural number $n \in \mathbb{N}$ such that $\frac{ps}{qr} < n$. So it suffices to prove that for every rational $\frac{p}{q}$ there exists $n \in \mathbb{N}$ such that $\frac{p}{q} < n$. Let us show this. After possibly multiplying both p and q by -1 we can assume that $q \in \mathbb{N}_{>0}$. If $p \leqslant 0$, put $n = 1$ and observe $p \leqslant 0 < q = n \cdot q$. If $p > 0$, then $p \in \mathbb{N}$, so by the archimedean property for natural numbers 2.1.41 there exists $n \in \mathbb{N}$ such that $p < n \cdot q$. Now observe that $0 < \frac{1}{q}$ because $(1 \cdot 1 - 0 \cdot q) \cdot 1 \cdot q = q > 0$. By monotony of multiplication, we can now conclude from $p < n \cdot q$ that $\frac{p}{q} < n$. This proves the claim. $\square$

2.4.13 An important observation for the field of rational numbers is that not all equations of the form $x^2 + a = 0$ for given $a \in \mathbb{Q}$ have a solution $x \in \mathbb{Q}$.

2.4.14 Proposition *Let $n \in \mathbb{N}$ be a prime number. Then the equation $x^2 - n = 0$ does not have a rational solution.*

Proof. Assume that there is a rational number x such that $x^2 = n$. After possibly passing to $-x$ we can assume that there exist $p \in \mathbb{N}$ and $q \in \mathbb{N}_{>0}$ such that $x = \frac{p}{q}$. We can even assume that q is the least positive natural number for which there exist a $p \in \mathbb{N}$ such that $x = \frac{p}{q}$. Then $n \cdot q^2 = p^2$. Since n is prime, Euclid's Lemma 2.3.10 implies that n is a divisor of p. Write $p = n \cdot r$ for some $r \in \mathbb{N}$. Then $q^2 = n \cdot r^2$, which entails by Euclid's Lemma again that n is a divisor of q. Hence $q = n \cdot s$ for some $s \in \mathbb{N}_{>0}$. Observe that s has to be smaller than q. Now $x = \frac{p}{q} = \frac{r}{s}$ which contradicts the minimality of q. □

Ordered fields

2.4.15 In the following we introduce new concepts for ordered fields and will derive properties which hold in any ordered field, so in particular in $\mathbb{Q}$.

Let us recall some notation and assume that $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ is an ordered field. Then $\mathbb{F}$ (or better $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$) is in particular an ordered integral domain, so the sets $\mathbb{F}_{>0}$ of *positive* elements and $\mathbb{F}_{<0}$ of *negative* elements are defined as $\{x \in \mathbb{F} \mid 0 < x\}$ and $\{x \in \mathbb{F} \mid x < 0\}$, respectively, cf. Definition 2.2.32. The sets $\mathbb{F}_{\geqslant 0} := \{x \in \mathbb{F} \mid 0 \leqslant x\}$ and $\mathbb{F}_{\leqslant 0} := \{x \in \mathbb{F} \mid x \leqslant 0\}$ are the the sets of *non-negative* and *non-positive* elements of $\mathbb{F}$, respectively.

2.4.16 Proposition and Definition *Let* $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ *be an ordered field. Then the set* $\mathbb{F}^+ := \mathbb{F}_{\geqslant 0}$ *of non-negative elements is a* positive cone *that is it satisfies the following axioms:*

(PC1) *For all* $x, y \in \mathbb{F}^+$ *the sum* $x + y$ *lies in* $\mathbb{F}^+$.

(PC2) *For all* $x, y \in \mathbb{F}^+$ *the product* $x \cdot y$ *lies in* $\mathbb{F}^+$.

(PC3) *The square* x^2 *is an element of* $\mathbb{F}^+$ *for every* $x \in \mathbb{F}$.

(PC4) *The element* -1 *does not lie in* $\mathbb{F}^+$.

(PC5) *The field* $\mathbb{F}$ *is the union of* $\mathbb{F}^+$ *and* $\mathbb{F}^- := -\mathbb{F}^+$.

Proof. (PC1) follows by monotony of addition and transitivity: $x+y \geqslant x \geqslant 0$, (PC2) by monotony of multiplication: $x \cdot y \geqslant x \cdot 0 = 0$. (PC3) is a consequence of Proposition 2.2.36 (iv). Adding -1 to the inequality $0 < 1$ gives $-1 < 0$ which entails (PC4). Next observe that by Proposition 2.2.36 (ii) $\mathbb{F}^- = \mathbb{F}_{\leqslant 0}$. (PC5) then follows by the trichotomy law for $\leqslant$. □

2.4.17 Remark The proof of the proposition also shows $\mathbb{F}^- = \mathbb{F}_{\leqslant 0}$.

2.4.18 Corollary *In an ordered field* $\mathbb{F}$ *the equation* $x^2 = 1$ *does not have a solution.*

Proof. This is an immediate consequence of (PC3) and (PC4). □

2.4.19 Proposition *Let* $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ *be an ordered field. Then the map*

$$\mathbb{Z} \to \mathbb{F}, \quad p \mapsto p \cdot 1 := \begin{cases} \sum\limits_{i=1}^{p} 1 & \textit{if } p \in \mathbb{Z}_{>0}, \\ 0 & \textit{if } p = 0, \\ -\sum\limits_{i=1}^{-p} 1 & \textit{if } p \in \mathbb{Z}_{<0}, \end{cases}$$

is a strictly order preserving embedding of rings. It is the only ring homomorphism from $\mathbb{Z}$ *to* $\mathbb{F}$.

Proof. By Proposition 2.2.29, the map $\iota : \mathbb{Z} \to \mathbb{F}$, $p \mapsto p \cdot 1$ is a group homomorphism with respect to the additive group structures on $\mathbb{Z}$ and $\mathbb{F}$. Observe that by definition $\iota(1) = 1$. Next let us show by induction on n that $\iota(np) = \iota(n)\iota(p)$ for all $n \in \mathbb{N}$ and $p \in \mathbb{Z}$. For $n = 0, 1$ the claim is trivial. Assume that it holds for some $n > 0$. Then

$$\iota((n+1)p) = \iota(np + p) = \iota(n)\iota(p) + \iota(p) = (\iota(n) + 1)\iota(p) = \iota(n+1)\iota(p) \ ,$$

which shows that the claim holds for $n + 1$ as well. Moreover,

$$\iota((-n)p) = -\iota(np) = -\iota(n)\iota(p) = \iota(-n)\iota(p) \ .$$

Hence ι is a ring homomorphism. Now let $p, q \in \mathbb{Z}$ and assume $q < p$. Then $n = p - q \in \mathbb{N}_{>0}$ and $p \cdot 1 - q \cdot 1 = n \cdot 1 = \sum_{i=1}^{n} 1 > 0$, where the latter inequality follows by induction on n. Therefore the map $\mathbb{Z} \to \mathbb{F}$, $p \mapsto p \cdot 1$ is strictly order preserving and in particular injective. If $\mu : \mathbb{Z} \to \mathbb{F}$ is another ring homomorphism, then for $n \in \mathbb{N}$

$$\mu(n) = \mu\left(\sum_{i=1}^{n} 1\right) = \sum_{i=1}^{n} \mu(1) = \iota(n) \quad \text{and} \quad \mu(-n) = -\mu\left(\sum_{i=1}^{n} 1\right) = -\sum_{i=1}^{n} \mu(1) = \iota(-n)$$

since $\mu(1) = 1$. Hence ι and μ coincide. □

2.4.20 Since by the preceding result the set of integers is a subset of any ordered field $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$, one can ask the question whether $\mathbb{N}$ or $\mathbb{Z}$ are unbounded within that field. It will turn out that this is not always so.

2.4.21 Definition An ordered field $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ is said to be *archimedean ordered* if for every pair of elements $x, y \in \mathbb{F}$ with $x > 0$ there exists a natural number $n \in \mathbb{N}$ such that $y < n \cdot x$.

2.4.22 Example The field of rational numbers is archimedean ordered by Theorem 2.4.12.

2.4.23 The field of rational numbers is the smallest archimedean field and contained in any other archimedean ordered field by the following result.

2.4.24 Proposition *Every archimedean ordered field $\mathbb{F}$ contains $\mathbb{Q}$ as subfield in a canonical way. More precisely, there exists a unique strictly order preserving ring homomorphism $\mathbb{Q} \hookrightarrow \mathbb{F}$.*

Proof. By Proposition 2.4.19 there exists a uniquely determined strictly order preserving ring homomorphism $\mathbb{Z} \to \mathbb{F}$. By the universal property of localization, see Theorem 2.4.6, this ring homomorphism extends in a unique way to a ring homomorphism $\mathbb{Q} \to \mathbb{F}$. Let us show that this ring homomorphism, which we denote by ι, is order preserving. Assume that $\frac{p}{q}$ and $\frac{r}{s}$ are rational numbers and that $\frac{p}{q} < \frac{r}{s}$. After possibly passing to $\frac{-p}{-q}$ respectively $\frac{-r}{-s}$ we can assume that $q, s \in \mathbb{N}_{>0}$. Then $n := (rq - ps)qs \in \mathbb{N}_{>0}$ and $(\iota(r)\iota(q) - \iota(p)\iota(s))\iota(qs) = \iota(n) > 0$ by Proposition 2.4.19. Since $\iota(qs) = \iota(q)\iota(s) > 0$, $\iota(qs)^{-1}$ is positive by Proposition 2.2.36 (vi). Hence

$$\iota\left(\frac{r}{s}\right) - \iota\left(\frac{p}{q}\right) = \iota(r)\iota(s)^{-1} - \iota(p)\iota(q)^{-1} > 0$$

and ι is strictly order preserving. □

2.4.25 Remark By the proposition one can identify for every ordered field $\mathbb{F}$ the field of rationals with its image under the canonical embedding $\mathbb{Q} \hookrightarrow \mathbb{F}$. We will follow that convention in this work and will call the elements in the image of the embedding $\mathbb{Q} \hookrightarrow \mathbb{F}$ the *rational elements* of $\mathbb{F}$.

2.4.26 Proposition *For an ordered field* $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ *the following properties are equivalent:*

(i) *The field* $\mathbb{F}$ *is archimedean ordered.*

(ii) *For every* $y \in \mathbb{F}$ *there exists* $n \in \mathbb{N}$ *such that* $y < n$.

(iii) *For all* $x, y \in \mathbb{F}$ *with* $x < y$ *there exists a rational element* r *such that* $x < r < y$.

(iv) *For all* $x \in \mathbb{F}$ *with* $x \geqslant 0$ *there exists a unique* $n \in \mathbb{N}$ *such that* $n \leqslant x < n + 1$.

Proof. First observe that (ii) implies (i) since for $x, y \in \mathbb{F}$ with $x > 0$ (ii) entails the exsistence of a natural n with $\frac{y}{x} < n$. By positivity of n one obtains $y < n \cdot x$, hence (i). The converse implication is trivial, so (i) and (ii) are equivalent.

Assume that $\mathbb{F}$ is archimedean ordered. Let us show that then (iv) holds true. So let $x \geqslant 0$. The set of natural numbers $\leqslant x$ contains at least 0 and is finite by the archimedean property. Hence it has a maximal element. Denote that element by n. Then $x < n + 1$ by maximality of n. To verify uniqueness let m be another natural number with $m \leqslant x < m + 1$. Then $m \leqslant n$ by maximality of n. One concludes $n \leqslant m$ from $n \leqslant x < m + 1$. Hence $n = m$ and (iv) is proved. Note that (iv) implies (ii). So (i) and (iv) are equivalent.

Now assume $\mathbb{F}$ is archimedean and that x, y are elements of $\mathbb{F}$ with $x < y$. Since $y - x > 0$ there exists a $q \in \mathbb{N}_{>0}$ with $1 < q \cdot (y - x)$. Then choose $p \in \mathbb{N}$ such that $p \leqslant qx < p + 1$. One concludes $qx < p + 1 \leqslant qx + 1 < qy$, hence $x < \frac{p}{q} < y$. This proves (iii). If conversely (iii) holds true and $x > 0$ then there exist $p, q \in \mathbb{N}_{>0}$ with $x < \frac{p}{q} < x + 1$. Therefore $x \leqslant qx < p$ which entails (ii).□

2.4.27 Definition Let $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ be an ordered field. The *absolute value function* on $\mathbb{F}$ is the following map:

$$|\cdot| : \mathbb{F} \to \mathbb{F}, \quad x \mapsto \begin{cases} x & \text{if } x \in \mathbb{F}_{>0} = \mathbb{F}^{+} \backslash \{0\}, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x \in \mathbb{F}_{<0} = \mathbb{F}^{-} \backslash \{0\}\ . \end{cases}$$

One calls $|x|$ the *absolute value* of $x \in \mathbb{F}$.

2.4.28 Remark Since an ordered field is in particular an ordered integral domain, all properties of Proposition 2.2.36 and Proposition 2.2.38 hold true for the absolute value function of an ordered field.

The order topology

2.4.29 Let $(X, \leqslant)$ be a totally ordered set. *Intervals* in X are subsets $I \subset X$ which contain with every pair of elements $x < y$ also every element z between them which means that for all $x, y \in I$ with $x < y$ and all $z \in X$ with $x \leqslant z \leqslant y$ the relation $z \in I$ holds true. The following subsets are clearly intervals, where x, y denote arbitrary elements of X and $\pm\infty$ are two symbols assumed not to stand for elements of X:

$$\begin{aligned}
(x,y) := (x,y)_X &:= \{z \in X \mid x < z < y\} \ , \\
[x,y) := [x,y)_X &:= \{z \in X \mid x \leqslant z < y\} \ , \\
(x,y] := (x,y]_X &:= \{z \in X \mid x < z \leqslant y\} \ , \\
[x,y] := [x,y]_X &:= \{z \in X \mid x \leqslant z \leqslant y\} \ , \\
(x,\infty) := (x,\infty)_X &:= \{z \in X \mid x < z\} \ , \\
(-\infty,y) := (-\infty,y)_X &:= \{z \in X \mid z < y\} \ , \\
(-\infty,\infty) := (-\infty,\infty)_X &:= X \ , \\
[x,\infty) := [x,\infty)_X &:= \{z \in X \mid x \leqslant z\} \ , \\
(-\infty,y] := (-\infty,y]_X &:= \{z \in X \mid z \leqslant y\} \ .
\end{aligned}$$

One calls intervals of the form $(x,y)_X$, $(x,\infty)_X$ $(-\infty,y)$, or $(-\infty,\infty)_X$ *open intervals* in X, of the form $[x,y]_X$, $[x,\infty)_X$ or $(-\infty,y]_X$ *closed intervals*, and those of the form $[x,y)_X$ or $(x,y]_X$ *half-open intervals*.

2.4.30 Remark When the context makes it clear which underlying set X is meant, we will usually use the notation (x,y) instead of $(x,y)_X$, $[x,y)$ instead of $[x,y)_X$ and so on.

2.4.31 Definition Let $(X, \leqslant)$ be a totally ordered set with more than one elements. The *order topology* $\mathcal{T}_{(X,\leqslant)}$ on $(X, \leqslant)$ (or just X) is then defined as the set of all subsets $O \subset X$ such that for each $x \in O$ one of the following holds:

(i) there exist $a, b \in X$ with $x \in (a,b) \subset O$,

(ii) the element x is a minimum of X and there exists $b \in X$ such that $[x,b) \subset O$,

(iii) the element x is a maximum of X and there exists $a \in X$ such that $(a,x] \subset O$,

2.4.32 Remark Using the abbreviation $\overline{X} := X \cup \{\pm\infty\}$, the order topology on X can be equivalently defined as the set of all subsets $O \subset X$ such that for each $x \in O$ there exists $a, b \in \overline{X}$ such that $x \in (a,b) \subset O$.

2.4.33 Theorem and Definition *The set $\mathcal{T} = \mathcal{T}_{(X,\leqslant)}$ associated to a totally ordered set $(X, \leqslant)$ is a* topology *on X which means that the following axioms hold true:*

(Top0) *The sets X and $\varnothing$ are both elements of $\mathcal{T}$.*

(Top1) *The union of any collection of elements of $\mathcal{T}$ is again in $\mathcal{T}$ that means if $(U_i)_{i\in I}$ is a family of elements $U_i \in \mathcal{T}$, then $\bigcup_{i\in I} U_i \in \mathcal{T}$.*

(Top2) *The intersection of finitely many elements of $\mathcal{T}$ is again in $\mathcal{T}$ that is for every natural n and $U_1, \ldots, U_n \in \mathcal{T}$ the set $\bigcap_{i=1}^n U_i$ lies in $\mathcal{T}$.*

Proof. The emptyset is an element of $\mathcal{T}$ by definition. Let $x \in X$. If x is neither a minimum nor a maximum, then there exist $a, b \in X$ with $a < x < b$ or in other words $x \in (a,b)$. If x is a minimum of X, then there exists a $b > x$. Hence $x \in [x,b)$. Similarly, if x is a maximum, there exists an $a \in X$ such that $x \in (a,b]$. So $X \in \mathcal{T}$ and (Top0) is proved.

Next let $x \in \bigcup_{i \in I} U_i$, where each $U_i \in \mathcal{T}$, $i \in I$. Choose $j \in I$ such that $x \in U_j$. Then either $I = (a, b) \in U_j$ for some a, b with $a < x < b$, $I = [x, b) \in U_j$ for some $b > x$, or $I = (a, x] \in U_j$ for some $a < x$. In each case one has $I \subset \bigcup_{i \in I} U_i$ which proves (Top1).

Finally let $x \in \bigcap_{i=1}^n U_i$, where $U_1, \dots, U_n \in \mathcal{T}$. Assume first that x is neither a minimal nor a maximal element of X. Then one can find $a_1, \dots, a_n, b_1, \dots, b_n \in X$ such that $a_i < x < b_i$ and $(a_i, b_i) \subset U_i$ for $i = 1, \dots, n$. Let a be the maximum of the a_i and b the minimum of the b_i. Then $a < x < b$ and $(a, b) \subset \bigcap_{i=1}^n U_i$. Next assume x to be a minimum of X. Then there are $b_1, \dots, b_n \in X$ such that $[x, b_i) \subset U_i$ for $i = 1, \dots, n$. As before let b denote the minimum of the b_i. Then $[x, b) \subset \bigcap_{i=1}^n U_i$. The final case where x is a maximum of X works analogously. Choose $a_1, \dots, a_n \in X$ such that $(a_i, x] \subset U_i$ for $i = 1, \dots, n$ and let a be the maximum of the a_i. Then $(a, x] \subset \bigcap_{i=1}^n U_i$. So (Top2) holds true as well. □

2.4.34 Remarks (a) A set X together with a subset $\mathcal{T} \subset \mathcal{P}(X)$ such that the above axioms (Top0) to (Top2) hold true is called a *topological space*. We sometimes denote such a topological space as a pair $(X, \mathcal{T})$. Subsets of X which are elements of the topology $\mathcal{T}$ are called *open*, subsets of X whose complement in X is open are called *closed*. Note that both the sets of open and of closed subsets of a topological space X are ordered by set-theoretic inclusion. From now on we will use the notion of topological spaces even though the class of topological spaces will be studied in detail only later in Chapter II.1.

(b) For every set X the power set $\mathcal{P}(X)$ and the set $\{\emptyset, X\}$ are topologies on X called the *finest* and the *coarsest* topology on X, respectively. Note that this language makes fully sense since the set of topologies on a set X is also ordered by set-theoretic inclusion, and $\mathcal{P}(X)$ is the largest, $\{\emptyset, X\}$ the smallest topology on X with respect to this order relation.

2.4.35 Definition Let $(X, \mathcal{T})$ be a topological space and $A \subset X$ a subset. By the *interior* of A one understands the union of all open subsets contained in A, by the *closure* of A the intersection of all closed subsets containing A. The interior of A is denoted by the symbol $\mathring{A}$, its closure by $\bar{A}$. The complement $\bar{A} \setminus \mathring{A}$ is called the *boundary* of A and is denoted ∂A.

2.4.36 Lemma *Let $(X, \mathcal{T})$ be a topological space and $A \subset X$ a subset.*

(i) *The interior $\mathring{A}$ is the largest open subset of X contained in A.*

(ii) *The closure $\bar{A}$ is the smallest closed subset of X containing A.*

Proof. This follows immediately from the observation that the union of open sets of $(X, \mathcal{T})$ is open and that the intersection of closed sets of $(X, \mathcal{T})$ is closed. □

2.4.37 Definition Let $x \in X$ be a point of a topological space $(X, \mathcal{T})$. A subset $V \subset X$ is called a *neighborhood* of x if there exists an open set $U \in \mathcal{T}$ such that $x \in U \subset V$. The set of neighborhoods of a point $x \in X$ will be denoted $\mathcal{U}_x$. If $\mathbb{F}$ is an ordered field and $\mathcal{T}$ the order topology on $\mathbb{F}$, one calls the elements of $\mathcal{U}_0$ *zero neighborhoods*.

2.4.38 Lemma *Let $(X, \mathcal{T})$ be a topological space and $x \in X$ a point.*

(i) *Any superset of a neighborhood of x is a neighborhood of x.*

(ii) *The intersection of finitely many neighborhoods of x is a neighborhood of x.*

Proof. The claim is an immediate consequence of the definition of a neighborhood of x and axiom (Top2) for a topology. □

2.4.39 Definition

2.5. The real numbers

Complete ordered fields

2.5.1 Unless some ambiguity could arise we will from now on denote a an ordered field $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ simply by the symbol $\mathbb{F}$.

2.5.2 Definition An ordered field $\mathbb{F}$ is called *Dedekind complete* if every set bounded above has a least upper bound.

2.5.3 Lemma *An ordered field $\mathbb{F}$ is Dedekind complete if and only if every set bounded below has a greatest lower bound.*

Proof. The claim is an immediate consequence of the fact that multiplication by -1 is a strictly order reversing automorphism of the abelian group $(\mathbb{F}, +, 0)$. □

2.5.4 Proposition *A Dedekind complete ordered field $\mathbb{F}$ is archimedean.*

Proof. Assume that $\mathbb{F}$ is non-archimedean which in other words means that $\mathbb{N}$ is a bounded subset of $\mathbb{F}$. By Dedekind completeness there exists a least upper bound M of $\mathbb{N}$. Since $M - 1 < M$ the element $M - 1$ is not an upper bound of $\mathbb{N}$, hence there exists an $n \in \mathbb{N}$ such that $M - 1 < n$. But then $M < n + 1$ which is a contradiction. Therefore $\mathbb{F}$ is archimedean. □

2.5.5 Example The converse of the proposition does not hold true in general. For example the ordered field of rational numbers is archimedean but not Dedekind complete. To prove this consider the set $X = \{x \in \mathbb{Q} \mid x^2 \leqslant 2\}$. Since $|x|^2 = x^2 \leqslant 2 < 2^2$ for $x \in X$, the estimate $|x| < 2$ holds true for all $x \in X$ by Proposition 2.2.36 (v). Hence X is a bounded subset of $\mathbb{Q}$. But X does not have a least upper bound in $\mathbb{Q}$. To prove this we will lead the assumption that X has a least upper bound in $\mathbb{Q}$ to a contradiction. So assume that b is rational and a least upper bound of X. First note that $b > 1$ since $1 \in X$. If $b^2 < 2$ choose $n \in \mathbb{N}_{>0}$ so that $2b + 1 < n \cdot \left(2 - b^2\right)$. Then $\left(b + \frac{1}{n}\right)^2 \leqslant b^2 + \frac{2b+1}{n} < 2$ which contradicts the assumption that b is an upper bound of X. Hence $b^2 \geqslant 2$. Assume that $b^2 > 2$ and put $c = \frac{1}{2}\left(b + \frac{2}{b}\right)$. Then $c^2 - 2 = \frac{1}{4}\left(b - \frac{2}{b}\right)^2 \geqslant 0$, hence c is an upper bound of X. But $b - c = \frac{1}{2b}\left(b^2 - 2\right) > 0$ which contradicts that b is the least upper bound of X. So $b^2 = 2$. But that contradicts Proposition 2.4.14, hence $\mathbb{Q}$ is not Dedekind complete.

2.5.6 Definition By a *Dedekind cut* or shortly a *cut* in an archimedean ordered field $\mathbb{F}$ one understand a pair (A, B) of subsets $A, B \subset \mathbb{F}$ such that the following holds true

(DC1) The sets A, B form a partition of $\mathbb{F}$ that means $\mathbb{F} = A \cup B$.

(DC2) The set A does not have a greatest element.

(DC3) For all elements $a \in A$ and $b \in B$ one has $a < b$.

A cut (A, B) of $\mathbb{F}$ is called a *gap* if B does not have a least element.

2.5.7 Example Let $\mathbb{F}$ be archimedean ordered, $x \in \mathbb{F}$ and put $A = \{y \in \mathbb{F} \mid y < x\}$ and $B = \{z \in \mathbb{F} \mid z \geqslant x\}$. Then (A, B) is a Dedekind cut which is not a gap.

2.5.8 Proposition *Let (A, B) be a Dedekind cut in the archimedean ordered field $\mathbb{F}$. Then the following holds true:*

(i) *If $a, x \in \mathbb{F}$ satisfy $a \in A$ and $x \leqslant a$, then $x \in A$.*

(ii) *If $b, y \in \mathbb{F}$ satisfy $b \in B$ and $b \leqslant y$, then $y \in B$.*

(iii) *The lower part A is always open in $\mathbb{F}$ and satisfies $\mathring{\overline{A}} = A$.*

(iv) *The upper part B is always closed in $\mathbb{F}$ and satisfies $\overline{\mathring{B}} = B$.*

(v) *If (A, B) is a gap, then A and B are both open and closed.*

Proof. Properties (i) and (ii) are immediate consequences of (DC1) and (DC3).

To verify that the lower part A is open let $x \in A$. Since A does not have a greatest element there exists a $a \in A$ with $x < a$. Since $\mathbb{F}$ is archimedean there exists $n \in \mathbb{N}$ such that $a - x < n$. Hence $a - n < x < a$. Since every element $y \in \mathbb{F}$ with $y < a$ is an element of A by (i), the interval $(a - n, a)$ is contained in A and contains x. So A is open in the order topology. Its complement B therefore is closed.

The relation $A \subset \overline{A}$ entails $A = \mathring{A} \subset \mathring{\overline{A}}$. To verify (iv) it therefore remains to prove that $\mathring{\overline{A}} \subset A$. Assume that $x \in \mathring{\overline{A}}$. Then there exist $a, b \in \mathbb{F}$ such that $x \in (a, b) \subset \mathring{\overline{A}}$. But this means that every open neighborhood of b contains a point of $\overline{A}$ hence meets A. Therefore $b \in \overline{A}$. Choose $c \in \mathbb{F}$ such that $x < c < b$. Then $(c, b+1)$ is an open neighborhood of b, so contains some element $d \in A$. Since $x < d$, the relation $x \in A$ follows by (i) and (iv) is proved.

Claim (iv) follows from the equality

$$\overline{\mathring{B}} = \mathbb{F} \setminus \mathring{\overline{A}} = \mathbb{F} \setminus A = B \ .$$

To prove (v) is suffices to show that B is open. Let y be an element of B. Since by assumption B does not have a least element there exists $b \in B$ with $b < y$. Choose $n \in \mathbb{N}$ such that $y - b < n$. Then $y \in (b, b+n) \subset B$, so B is open and the last claim is verified. □

2.5.9 Proposition *For an archimedean ordered field $(\mathbb{F}, +, \cdot, 0, 1, \leqslant)$ the following properties are equivalent:*

(i) *The ordered field $\mathbb{F}$ is Dedekind complete.*

(ii) *No Dedekind cut in $\mathbb{F}$ is a gap.*

(iii) *The* nested interval property *holds for $\mathbb{F}$ which means that for all sequences $\big([a_n, b_n]\big)_{n\in\mathbb{N}}$ of intervals fulfilling $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$ for all $n \in \mathbb{N}$ the intersection $\bigcap_{n\in\mathbb{N}} [a_n, b_n]$ is non-empty.*

(iv) *The archimedean ordered field* $\mathbb{F}$ *is* Cauchy complete *that is every Cauchy sequence in* $\mathbb{F}$ *converges.*

Proof.

Definition and uniqueness of real number fields

2.5.10 Definition By a *real number field* one understands a complete archimedean ordered field.

2.5.11 As we will show in this section all real number fields coincide up to isomorphism. Moreover, there exists a real number field. The proof will be broken up in a a uniqueness part and an existence part. For the existence we need the notion of a morphism of (ordered) fields which will be introduced next. Existence will be shown in two ways, first in the way real numbers were constructed by R. Dedekind , then by constructung the complete hull of $\mathbb{Q}$ via equivalence classes of Cauchy sequences.

2.5.12 Definition A *field homomorphism* or shortly a *homorphism* between two fields $\mathbb{K}$ and $\mathbb{F}$ is a ring homomorphism $f : \mathbb{K} \to \mathbb{F}$. If both fields are ordered, an order preserving field homomorphism $f : \mathbb{K} \to \mathbb{F}$ is called a *morphism of ordered fields.*

2.5.13 Proposition *Fields with field homomorphisms as morphisms form a category denoted* Fld. *Ordered fields together with order preserving field homomorphism as morphisms are a subcategory* OFld. *The real number fields form a full subcategory of* OFld.

Proof. Obviously, the identity map $\mathrm{id}_{\mathbb{K}}$ of a field $\mathbb{K}$ is a morphism. If $\mathbb{K}$ is ordered, the identity map is also order preserving. Moreover, the composition of field morphisms is a morphism as well. If both morphisms are order preserving the composition is so as well. The claim follows. □

2.5.14 Proposition *Every field homomorphism* $f : \mathbb{K} \to \mathbb{F}$ *is injective.*

Proof. Let $x, y \in \mathbb{K}$ and assume that $f(x) = f(y)$. Then $f(x - y) = 0$. If $x \neq y$, the difference $x - y$ has an inverse z. But that implies

$$1 = f(1) = f\big(z(x - y)\big) = f(z)f(x - y) = 0$$

which is impossible. So $x = y$ and f is injective. □

2.5.15 Theorem *Let* $\mathbb{K}$ *and* $\mathbb{F}$ *be real number fields. Then there exists an order preserving field homomorphism* $f : \mathbb{K} \to \mathbb{F}$. *Every such morphism is an isomorphism.*

Proof. By the preceding proposition we only need to show that f is surjective. □

Real numbers à la Dedekind

Cauchy completion of the field of rational numbers

2.5.16 Let $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ denote the set of all Cauchy sequences in $\mathbb{Q}$. Addition and multiplication can be extended pointwise from $\mathbb{Q}$ to $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ as follows:

$$\begin{aligned} + : \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \times \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \to \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}, &\quad \Big((x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}\Big) \mapsto (x_n + y_n)_{n\in\mathbb{N}} \\ \cdot : \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \times \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \to \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}, &\quad \Big((x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}\Big) \mapsto (x_n \cdot y_n)_{n\in\mathbb{N}} \end{aligned}$$

Addition and multiplication on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ are both associative and commutative since these properties hold componentwise. Furthermore, multiplication distributes from the left and from the right over addition, again since this property holds componentwise. Next define an embedding $\mathbb{Q} \hookrightarrow \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ by $r \mapsto (r)_{n\in\mathbb{N}}$, where $(r)_{n\in\mathbb{N}}$ denotes the constant sequence with each component being equal to r. Obviously, $\mathbb{Q} \hookrightarrow \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ preserves the operations of addition and multiplication. Moreover, the sequence $(0)_{n\in\mathbb{N}}$ serves as neutral element with respect to addition in $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$, and $(1)_{n\in\mathbb{N}}$ as neutral element with respect to multiplication. Hence $\big(\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}, +, \cdot, (0)_{n\in\mathbb{N}}, (1)_{n\in\mathbb{N}}\big)$ is a commutative ring and $\mathbb{Q} \hookrightarrow \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ an injective ring hommorphism. Now we define an equivalence relation $\sim$ on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ as follows. Call two Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ *equivalent*, in signs $(x_n)_{n\in\mathbb{N}} \sim (y_n)_{n\in\mathbb{N}}$, if the sequence $(x_n - y_n)_{n\in\mathbb{N}}$ is a null sequence. Denote the equivalence class of a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}$ by $[(x_n)_{n\in\mathbb{N}}]$, the quotient space $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \big/ \sim$ by $\mathbb{R}$, and let $q : \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}} \to \mathbb{R}$ be the quotient map which assigns to every element of $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ its equivalence class. We will now show that $\sim$ is a congruence relation on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$. This means that for Cauchy sequences $(x_n)_{n\in\mathbb{N}} \sim (x'_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}} \sim (y'_n)_{n\in\mathbb{N}}$ the two added sequences $(x_n + y_n)_{n\in\mathbb{N}}$ and $(x'_n + y'_n)_{n\in\mathbb{N}}$ and the two multiplied sequences $(x_n \cdot y_n)_{n\in\mathbb{N}}$ and $(x'_n \cdot y')_{n\in\mathbb{N}}$ are equivalent.

2.5.17 Lemma *The relation $\sim$ is a congruence relation on the commutative ring $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$.*

Proof. Asssume that $(x_n)_{n\in\mathbb{N}} \sim (x'_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}} \sim (y'_n)_{n\in\mathbb{N}}$, where all sequences are Cauchy sequences in $\mathbb{Q}$. Then $(x_n - x'_n)_{n\in\mathbb{N}}$ and $(y_n - y'_n)_{n\in\mathbb{N}}$ are both null sequences, hence there exist for every $\varepsilon > 0$ natural numbers $N_1(\varepsilon)$ and $N_2(\varepsilon)$ such that $|x_n - x'_n| < \varepsilon$ for $n \geqslant N_1(\varepsilon)$ and $|y_n - y'_n| < \varepsilon$ for $n \geqslant N_2(\varepsilon)$. Fix $\varepsilon > 0$ and put $N := \max\{N_1(\frac{\varepsilon}{2}), N_2(\frac{\varepsilon}{2})\}$. Then one obtains for all $n \geqslant N$

$$\big|(x_n + y_n) - (x'_n + y'_n)\big| = \big|(x_n - x'_n) + (y_n - y'_n)\big| \leqslant |x_n - x'_n| + |y_n - y'_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \ .$$

Next observe that $(x'_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are bounded since both sequences are Cauchy. Choose $M_1, M_2 > 0$ such that $|x'_n| \leqslant M_1$ and $|y_n| \leqslant M_2$ for all $n \in \mathbb{N}$. Now put $N := \max\{N_1(\frac{\varepsilon}{2M_2}), N_2(\frac{\varepsilon}{2M_1})\}$. Then one gets for all $n \geqslant N$

$$\begin{aligned} \big|(x_n \cdot y_n) - (x'_n \cdot y'_n)\big| &= \big|(x_n - x'_n) \cdot y_n + x'_n \cdot (y_n - y'_n)\big| \leqslant \\ &\leqslant |x_n - x'_n|\,|y_n| + |y_n - y'_n|\,|x'_n| < \frac{\varepsilon}{2M_2}\,M_2 + \frac{\varepsilon}{2M_1}\,M_1 = \varepsilon \ . \end{aligned}$$

Hence $(x_n + y_n)_{n\in\mathbb{N}} \sim (x'_n + y'_n)_{n\in\mathbb{N}}$ and $(x_n \cdot y_n)_{n\in\mathbb{N}} \sim (x'_n \cdot y'_n)_{n\in\mathbb{N}}$. □

As a consequence of the lemma, addition and multiplication on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ descend to unique operations $+$ and $\cdot$ on the quotient space $\mathbb{R}$ such that for all rational Cauchy sequences $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$:

$$q\big((x_n)_{n\in\mathbb{N}} + (y_n)_{n\in\mathbb{N}}\big) = q\big((x_n)_{n\in\mathbb{N}}\big) + q\big((y_n)_{n\in\mathbb{N}}\big) \quad \text{and} \tag{2.5.1}$$

$$q\big((x_n)_{n\in\mathbb{N}} \cdot (y_n)_{n\in\mathbb{N}}\big) = q\big((x_n)_{n\in\mathbb{N}}\big) \cdot q\big((y_n)_{n\in\mathbb{N}}\big) \tag{2.5.2}$$

The resulting binary operations $+$ and $\cdot$ on $\mathbb{R}$ are both associative and commutative since they are on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$. Moreover, multiplication on $\mathbb{R}$ distributes over addition, again since $+$ and $\cdot$ have that property on $\mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$. The mapping $i : \mathbb{Q} \hookrightarrow \mathbb{R}$, $r \mapsto [(r)_{n\in\mathbb{N}}]$ is injective and preserves addition and multiplication by Equations 2.5.2 and 2.5.1 and since $\mathbb{Q} \to \mathbb{Q}^{\mathbb{N}}_{\text{Cauchy}}$ is a ring homomorphism. Injectivity follows from the fact that for rational r, s the sequence $(r-s)_{n\in\mathbb{N}}$ is constant, thus a null sequence if and only if $r = s$. From now on, we will denote the image of a rational number r under the embedding i by r as well. In particular the elements $[(0)_{n\in\mathbb{N}}]$ and $[(1)_{n\in\mathbb{N}}]$ will be denoted by 0 and 1, respectively. Last let us define the set $\mathbb{R}^+$ as the set of all equivalence classes of Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ such that $x_n \in \mathbb{Q}^+ = \{r \in \mathbb{Q} \mid r \geqslant 0\}$ for all $n \in \mathbb{N}$.

2.5.18 Theorem and Definition *The set $\mathbb{R}$ together with addition $+$, multiplication $\cdot$, and the elements $0, 1$ becomes a field. Moreover, the set $\mathbb{R}^+$ is a positive cone on $\mathbb{R}$. Thus one obtains a total order $\leqslant$ on $\mathbb{R}$ respecting the monotony laws by defining $x \leqslant y$ for $x, y \in \mathbb{R}$ if $y - x \in \mathbb{R}^+$. Altogether, $(\mathbb{R}, +, \cdot, 0, 1, \leqslant)$ is a complete ordered field called the* field of real numbers. *Finally, $i : \mathbb{Q} \hookrightarrow \mathbb{R}$ an embedding of fields.*

Proof.

2.6. The complex numbers

Part I.

Fundamentals of Algebra

I.1. Group theory

1.1. The category of groups

1.1.1 Definition A set G together with a binary operation $\cdot : G \times G \to G$ and an element $e \in G$ is called a *group* if it satisfies the following axioms:

(Grp1) The operation $\cdot$ is *associative* that means

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \quad \text{for all } a, b, c \in G \ .$$

(Grp2) The element e is *neutral* with respect to the operation $\cdot$ which means that

$$e \cdot a = a \cdot e = g \quad \text{for all } a \in G \ .$$

(Grp3) For each $a \in G$ there exists an *inverse*, i.e. an element $b \in G$ such that

$$a \cdot b = b \cdot a = e \ .$$

If in addition the following axiom of commutativity holds true, the group G is called *abelian*:

(Grp4) The operation $\cdot$ is *commutative* that means

$$a \cdot b = b \cdot a \quad \text{for all } a, b \in G \ .$$

If $\cdot : G \times G \to G$ is a binary operation on G such that (Grp1) is fulfilled, the pair $(G, \cdot)$ is called a *semigroup*. In case G together with the binary operation $\cdot$ and an element $e \in G$ satisfies axioms (Grp1) and (Grp2), one calls $(G, \cdot, e)$ a *monoid*, if it satisfies axioms (Grp1), (Grp2), and (Grp4), the monoid $(G, \cdot, e)$ is called *abelian*.

1.1.2 Examples (a) The set $\mathbb{N}$ of natural numbers together with addition and 0 is an abelian monoid by Theorem 2.1.14.

(b) The set $\mathbb{Z}$ of integers together with addition as group law and 0 is an abelian group by Theorem 2.2.23.

(c) The set $\mathbb{Z}$ of integers together with multiplication as binary operation and 1 is an abelian monoid by Theorem 2.2.23.

I.2. Rings and modules

2.1. The category of rings

Definitions and first examples

Even though we shall mostly work with commutative rings in this book, we will introduce the general notion of rings which are allowed to be non-commutative.

2.1.1 Definition A *ring* is a set R together with an addition map $+ : R \times R \to R$, a multiplication map $\cdot : R \times R \to R$, and elements $0, 1 \in R$ that satisfy the following conditions:

(Ring1) R together with addition as binary operation and 0 as neutral element is an abelian group.

(Ring2) R together with multiplication as binary operation and 1 as neutral element is a monoid

(Ring3) Multiplication *distributes* from the left and the right over addition that is

$$r \cdot (s + t) = r \cdot s + r \cdot t \quad \text{for all } r, s, t \in R \text{, and}$$
$$(r + s) \cdot t = r \cdot t + s \cdot t \quad \text{for all } r, s, t \in R.$$

If, in addition, the following axiom holds, the ring R is called *commutative*:

(Ring4) Multiplication is *commutative* that is $r \cdot s = s \cdot r$ for all $r, s \in R$.

We shall typically write rs for $r \cdot s$. Sometimes, when we want to particularly denote the structure maps and structure elements of a ring we write a ring as a quintuple $(R, +, \cdot, 0, 1)$.

If R is a ring, an *invertible element* or a *unit* is an element $r \in R$, such that there exists an element $s \in R$, called (*multiplicative*) *inverse of* r, which satisfies

$$r \cdot s = 1 \quad \text{and} \quad s \cdot r = 1 \ .$$

The set of units of a ring R will be denoted by $R^\times$.

Given a ring R, a *subring* is a subset $S \subset R$ that contains the zero and identity elements, is closed under addition and multiplication and is closed under forming additive inverses. In other words, $S \subset R$ is a *subring*, if $0, 1 \in S$ and if for all $r, s \in S$ the elements $r + s$, $-r$ and rs are in S as well.

Following (Bourbaki, 1989, p. 98), a *pseudo-ring* (or in other words *non-unital-ring*) is a set R together with binary operations $+$ and $\cdot$ and an element 0 such that the Axioms (Ring1) and (Ring3) are satisfied and such that $\cdot$ is associative that is (Grp1) holds true. A pseudo-ring R is called *commutative* if Axiom (Ring4) is satisfied as well. A subset S of a pseudo-ring R is called a *sub-pseudo-ring*, if it contains the zero element, is closed under addition and multiplication, and is closed under forming additive inverses. Sometimes we write a pseudo-ring as a quadrupel $(R, +, \cdot, 0)$ to include the structure maps and its zero element.

If R is a pseudo-ring, the *center* of R is defined as the set of all $r \in R$ commuting with all ring elements that is as the set

$$\mathrm{Z}(R) := \left\{ r \in R \mid rs = sr \text{ for all } s \in R \right\} .$$

The following result is essentially a repetition of Proposition 2.2.15.

2.1.2 Proposition *Let R be a pseudo-ring. Then*

(i) $0 \cdot r = r \cdot 0 = 0$ *for all* $r \in R$.

(ii) $(-r)s = r(-s) = -(rs)$ *for all* $r, s \in R$.

(iii) *If R possesses a multiplicative identity* 1*, then* $(-1)^2 = 1$.

(iv) *A multiplicative identity in R is uniquely determined.*

(v) *Assume that R possesses an identity element. Then the inverse for an invertible $r \in R$ is uniquely determined. If it exists, the inverse of r is denoted by r^{-1}.*

Proof. ad (i). First compute using associativity, distributivity and that 0 is a zero element:

$$0 \cdot r = (0 + 0) \cdot r = (0 \cdot r) + (0 \cdot r) .$$

Adding $-(0 \cdot r)$ on both sides gives $0 = 0 \cdot r$. By an analogous argument we obtain $0 = r \cdot 0$.

ad (ii). By (i) we obtain

$$0 = 0 \cdot s = (r + (-r)) \cdot s = rs + (-r)s ,$$

which entails $(-r)s = -(rs)$. Similarly, one shows $r(-s) = -(rs)$.

ad (iii). If 1 is a multiplicative identity, then $0 = 0 \cdot (-1) = (1 + (-1)) \cdot (-1) = (-1) + (-1)^2$ which entails the claim by adding 1 on both sides.

ad (iv). Assume that 1 and $1'$ are two identity elements in R. Then

$$1 = 1 \cdot 1' = 1' .$$

ad (v). Let R be a ring with identity 1 and $s, s' \in R$ be two inverses of r. Then

$$s = s \cdot 1 = s \cdot (r \cdot s') = (s \cdot r) \cdot s' = 1 \cdot s' = s' . \qquad \square$$

2.1.3 Examples (a) The *zero ring* or *trivial ring* is the ring with underlying set $\{0\}$. Its identity element coincides with 0, and it is obviously commutative.

(b) The simplest non-trivial example of a ring is the commutative ring $\mathbb{Z}_2 = \{0, 1\}$. Note that as a consequence of the ring axioms and the fact that $0 \neq 1$ one has $-1 = 1$ in $\mathbb{Z}_2$.

(c) The most important example of a commutative ring is the ring of integers $\mathbb{Z}$.

(d) The sets $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ of rational, real, and complex numbers, respectively, form all commutative rings.

(e) The set $\mathbb{H}$ of quaternions is a ring which is not commutative.

2.1.4 Example The center $\mathrm{Z}(R)$ of a pseudo-ring R is a sub-pseudo-ring of R by Proposition 2.1.2. It is commutative by definition. Moreover, if R possesses an identity element, then $\mathrm{Z}(R)$ is even a subring.

2.1.5 Examples The following are examples of function rings.

(a) Let X be a set and R a pseudo-ring. The set R^X of functions $f : X \to R$ is a pseudo-ring. Hereby, addition and multiplication of functions $f, g : X \to R$ are defined pointwise: $(f + g)(x) := f(x) + g(x)$ and $(f \cdot g)(x) := f(x) \cdot g(x)$ for $x \in X$. Obviously, R^X with addition as binary operation then becomes an abelian group, where the zero function $0_X : X \to R$, $x \mapsto 0$ serves as neutral element, and the additive inverse $-f$ of $f \in R^X$ is given by $(-f)(x) := -f(x)$ for $x \in R$. Associativity and commutativity of addition in R^X hold true because they hold pointwise over each $x \in X$. Likewise, multiplication in R^X is associative. The distributivity law holds in R^X also, because it holds pointwise when evaluating at $x \in X$. So R^X becomes a pseudo-ring. If R is even a ring, the function $1_X : X \to R$, $x \mapsto 1$ serves as an identity element, so R^X then is a ring as well.

(b) A sub-pseudo-ring of R^X (independently of wether R is a ring or pseudo-ring) is given by the subset

$$R^{(X)} := \left\{ f \in R^X \;\middle|\; f(x) \neq 0 \text{ for at most finitely many } x \in X \right\} .$$

This follows immediately from the observation that the sum and the product of two elements $f, g \in R^{(X)}$ lie again in $R^{(X)}$, that $0_X \in R^{(X)}$, and that $R^{(X)}$ contains with an element f also its negative $-f$. Unless X is finite and R a ring, the pseudo-ring $R^{(X)}$ is not unital or in other words not a ring.

(c) If X is a topological space and $R = \mathbb{R}$, the subspace

$$\mathcal{C}(X) := \{f \in \mathbb{R}^X \mid f \text{ is continuous}\}$$

is a subring of $\mathbb{R}^X$, since the constant functions 0_X and 1_X are continuous, and since the sum and product of two real-valued functions on X are again continuous.

(d) If M is a smooth manifold, the subspace

$$\mathcal{C}^\infty(M) := \{f \in \mathcal{C}(M) \mid f \text{ is smooth}\}$$

is a subring of $\mathcal{C}(M)$, since the constant function 1_X is smooth, and since the sum and product of two real-valued smooth functions on M is again smooth.

2.1.6 Example Let R be a commutative ring. One defines $R[x]$, the *ring of polynomials in one variable over* R, as follows. As a set, $R[x]$ coincides with $R^{(\mathbb{N})}$. For an element $a \in R[x]$ denote by a_n for every $n \in \mathbb{N}$ its n-th component, that means let $a = (a_n)_{n\in\mathbb{N}}$. Using this agreement, the sum and product of two elements $a, b \in R[x]$ are defined by

$$\begin{aligned}(a+b)_n &:= (a_n + b_n) \text{ for all } n \in \mathbb{N}, \text{ and} \\ (a \cdot b)_n &:= \sum_{k+l=n} a_k b_l \text{ for all } n \in \mathbb{N} .\end{aligned}$$

By Example 2.1.5 (b) , $(R[x], +, 0_{\mathbb{N}})$ is an abelian group with zero element $0_{\mathbb{N}} : \mathbb{N} \to R$, $n \mapsto 0$. Let us show that $\cdot$ is an associative and commutative operation on $R[x]$. To this end let $a, b, c \in R[x]$ and compute for $n \in \mathbb{N}$ using associativity of multiplication in R

$$\begin{aligned}((a \cdot b) \cdot c)_n &= \sum_{l+k=n} \sum_{i+j=l} (a_i b_j) c_k = \sum_{i+j+k=n} (a_i b_j) c_k = \\ &= \sum_{i+j+k=n} a_i (b_j c_k) = \sum_{i+l=n} \sum_{j+k=l} a_i (b_j c_k) = (a \cdot (b \cdot c))_n .\end{aligned}$$

Next verify that by commutativity of R

$$(a \cdot b)_n = \sum_{k+l=n} a_k b_l = \sum_{k+l=n} b_l a_k = (b \cdot a)_n .$$

Denote by $1_{\mathbb{N}}$ the element of $R[x]$ defined by $(1_{\mathbb{N}})_0 = 1$ and $(1_{\mathbb{N}})_n = 0$ for $n \in \mathbb{N}_{>0}$. One checks immediately that $a \cdot 1_{\mathbb{N}} = 1_{\mathbb{N}} \cdot a = a$ for all $a \in R[x]$. Hence $(R[x], +, \cdot, 0_{\mathbb{N}}, 1_{\mathbb{N}})$ is a commutative ring as claimed. Let us now denote by x the element of $R[x]$ uniquely defined by the property that its n-th component is 1 for $n = 1$ and 0 otherwise. Moreover, for $r \in R$ and $a \in R[x]$ denote by ra the element with components $(ra)_n = ra_n$, $n \in \mathbb{N}$. With this agreement and the observation that the k-th component of x^n is 1 for $k = n$ and 0 otherwise one checks that every $a \in R[x]$ can be written in the form

$$a = \sum_{n\in\mathbb{N}} a_n x^n .$$

Note that the sum on the right hand side is finite indeed, since only finitely many a_n are allowed to be nonzero. The representation of a in the form $\sum_{n\in\mathbb{N}} a_n x^n$ is even unique. This follows from the observation that the k-th component of a sum $\sum_{n\in\mathbb{N}} b_n x^n$, where only finitely many $b_n \in R$ are nonzero, is the sum of the k-th components of the elements $b_n x^n$ and that the k-th component of $b_n x^n$ is b_k if $n = k$ and 0 otherwise. We will later see how one can embed R into $R[x]$ and how polynomial rings $R[x_1, \ldots, x_n]$ in several variables $x_1, \ldots, x_n$ are defined.

The class of rings forms a category. Its morphisms are called ring homomorphisms. Let us formulate this in more detail.

2.1.7 Definition A *morphism of pseudo-rings* is a map $f : R \to S$ between pseudo-rings $(R, +_R, \cdot_R, 0_R)$ and $(S, +_S, \cdot_S, 0_S)$ that respects addition and multiplication. That is

(Ring5) $f(x +_R y) = f(y) +_S f(y)$ for all $x, y \in R$.

(Ring6) $f(x \cdot_R y) = f(x) \cdot_S f(y)$ for all $x, y \in R$.

If R and S are rings, a morphism of pseudo-rings $f : R \to S$ which preserves the identity elements is called a *ring homomorphism*. More precisely, $f : R \to S$ is a *ring homomorphism* if it satisfies Axioms (Ring5) and (Ring6) and in addition the axiom

(Ring8) $f(1_R) = 1_S$, where 1_R and 1_S are the respective identity elements.

2.1.8 Proposition *Let $f : R \to S$ be a morphism of pseudo-rings. Then*

(Ring7) $f(0_R) = 0_S$*, where 0_R and 0_S are the respective zero elements.*

Proof. By additivity of f one has $f(0_R) = f(0_R +_R 0_R) = f(0_R) +_S f(0_R)$. From this one concludes $f(0_R) = 0_S$ since $(S, +_S, 0_S)$ is an abelian group, so has the cancellation property. □

2.1.9 Theorem and Definition (a) *The identity map* id_R *on a ring R is a ring homomorphism. If R is a pseudo-ring,* id_R *is a morphism of pseudo-rings.*

(b) *Let R, S and T be three rings and $f : R \to S$ and $S : Y \to T$ ring homomrphisms. Then $g \circ f$ is a ring homomorphism. If R, S, T are pseudo-rings and f, g morphisms of pseudo-rings, then so is $g \circ f$.*

(c) *Rings as objects together with ring homomorphisms as morphisms form a category. It is called the* category of rings *and will be denoted by* Ring*. Analogously, one obtains the category* PRing *of pseudo-rings and morphisms of pseudo-rings. The category* Ring *can be understood in a canonical way as a subcategory of* PRing.

Proof. The identity map on a (pseudo-) ring obviously preserves all structure maps and neutral elements. If f and g are morphisms of pseudo-rings, then, for all $x, y \in R$

$$\begin{aligned} (g \circ f)(x +_R y) &= g\big(f(x +_R y)\big) = g\big(f(x) +_S f(y)\big) = g\big(f(x)\big) +_T g\big(f(y)\big) = \\ &= (g \circ f)(x) +_T (g \circ f)(y) \ , \\ g \circ f(x \cdot_R y) &= g\big(f(x \cdot_R y)\big) = g\big(f(x) \cdot_S f(y)\big) = g\big(f(x)\big) \cdot_T g\big(f(y)\big) = \\ &= (g \circ f)(x) \cdot_T (g \circ f)(y) \ , \end{aligned}$$

hence $g \circ f$ is a morphism of pseudo-rings. If R, S, and T are even rings, and f, g ring homomorphisms, one checks

$$(g \circ f)(1_R) = g\big(f(1_R)\big) = g(1_S) = 1_T \ ,$$

so $g \circ f$ is a ring homomorphism. It is now clear that pseudo-rings together with morphisms of pseudo-rings and rings together ring homomorphisms form categories PRing and Ring, respectively. The canonical embedding of Ring into PRing is obtained by mapping a ring $(R, +, \cdot, 0, 1)$ to the pseudo-ring $(R, +, \cdot, 0)$ or in other words by forgetting the multiplicative identity. Note that this is an embedding indeed since by Proposition 2.1.2 (iv) identity elements in a ring are uniquely determined. □

Unital algebras over a commutative ring

2.1.10 The philosophy of Grothendieck, as expounded in his EGA **?**, is that one should always do things in a relative context. This means that instead of working with objects, one should work with *morphisms* of objects. Motivated by this, we introduce:

2.1.11 Definition Given a commutative ring R, a *unital R-algebra* is a ring A together with a morphism of rings (the *structure morphism*) $R \to Z(A) \subset A$. In other words, the structure morphism $R \to A$ has image in the center of the ring A. A unital R-algebra A is called *commutative* if A is a commutative ring.

A morphism between R-algebras is a ring homomorphism that is required to commute with the structure morphisms. So if A is an R-algebra, then A is not only a ring, but there is a way to multiply elements of A by elements of R. Namely, to multiply $a \in A$ with $x \in R$, take the image of x in A, and multiply that by a.

For instance, any ring is an algebra over any subring.

We can think of an A-algebra as an arrow $A \to R$, and a morphism from $A \to R$ to $A \to S$ as a commutative diagram

$$\begin{tikzcd} R \arrow[rr] & & S \\ & A \arrow[ul] \arrow[ur] & \end{tikzcd}$$

This is a special case of the *undercategory* construction.

If B is an A-algebra and C a B-algebra, then C is an A-algebra in a natural way. Namely, by assumption we are given morphisms of rings $A \to B$ and $B \to C$, so composing them gives the structure morphism $A \to C$ of C as an A-algebra.

2.1.12 Example Every ring is a $\mathbb{Z}$-algebra in a natural and unique way. There is a unique map (of rings) $\mathbb{Z} \to R$ for any ring R because a ring-homomorphism is required to preserve the identity. In fact, $\mathbb{Z}$ is the *initial object* in the category of rings: this is a restatement of the preceding discussion.

2.1.13 Example If R is a ring, the polynomial ring $R[x]$ is an R-algebra in a natural manner. Each element of R is naturally viewed as a "constant polynomial."

2.1.14 Example The field of complex numbers $\mathbb{C}$ is an $\mathbb{R}$-algebra.

2.1.15 Example For any ring R, we can consider the polynomial ring $R[x_1, \dots, x_n]$ which consists of the polynomials in n variables with coefficients in R. This can be defined inductively as $(R[x_1, \dots, x_{n-1}])[x_n]$, where the procedure of adjoining a single variable comes from the previous **??**.

We shall see a more general form of this procedure in Example 2.1.16.

Here is an example that generalizes the case of the polynomial ring.

2.1.16 Example If R is a ring and G a commutative monoid,[1] then the set $R[G]$ of formal finite sums $\sum r_i g_i$ with $r_i \in R, g_i \in G$ is a commutative ring, called the **monoid ring** or **group ring** when G is a group. Alternatively, we can think of elements of $R[G]$ as infinite sums $\sum_{g \in G} r_g g$ with R-coefficients, such that almost all the r_g are zero. We can define the multiplication law such that

$$\left(\sum r_g g\right)\left(\sum s_g g\right) = \sum_h \left(\sum_{gg'=h} r_g s_{g'}\right) h.$$

This process is called *convolution.* We can think of the multiplication law as extended the group multiplication law (because the product of the ring-elements corresponding to g, g' is the ring element corresponding to $gg' \in G$).

The case of $G = \mathbb{N}$ is the polynomial ring. In some cases, we can extend this notion to formal infinite sums, as in the case of the formal power series ring; see definition 1.3.5 below.

2.1.17 Remark The ring $\mathbb{Z}$ is an *initial object* in the category of rings. That is, for any ring R, there is a *unique* morphism of rings $\mathbb{Z} \to R$. We discussed this briefly earlier; show more generally that A is the initial object in the category of A-algebras for any ring A.

2.1.18 Remark The ring where $0 = 1$ (the **zero ring**) is a *final object* in the category of rings. That is, every ring admits a unique map to the zero ring.

2.1.19 Remark Let $\mathcal{C}$ be a category and $F : \mathcal{C} \to \mathbf{Sets}$ a covariant functor. Recall that F is said to be **corepresentable** if F is naturally isomorphic to $X \to \hom_{\mathcal{C}}(U, X)$ for some object $U \in \mathcal{C}$. For instance, the functor sending everything to a one-point set is corepresentable if and only if $\mathcal{C}$ admits an initial object.

Prove that the functor $\mathsf{Ring} \to \mathbf{Sets}$ assigning to each ring its underlying set is representable. (Hint: use a suitable polynomial ring.)

The category of rings is both complete and cocomplete. To show this in full will take more work, but we can here describe what certain cases (including all limits) look like. As we saw in remark 2.1.19, the forgetful functor $\mathsf{Ring} \to \mathbf{Sets}$ is corepresentable. Thus, if we want to look for limits in the category of rings, here is the approach we should follow: we should take the limit first of the underlying sets, and then place a ring structure on it in some natural way.

2.1.20 Example (Products) The **product** of two rings R_1, R_2 is the set-theoretic product $R_1 \times R_2$ with the multiplication law $(r_1, r_2)(s_1, s_2) = (r_1 s_1, r_2 s_2)$. It is easy to see that this is a product in the category of rings. More generally, we can easily define the product of any collection of rings.

To describe the coproduct is more difficult: this will be given by the *tensor product* to be developed in the sequel.

2.1.21 Example (Equalizers) Let $f, g : R \rightrightarrows S$ be two ring-homomorphisms. Then we can construct the **equalizer** of f, g as the subring of R consisting of elements $x \in R$ such that $f(x) = g(x)$. This is clearly a subring, and one sees quickly that it is the equalizer in the category of rings.

[1] That is, there is a commutative multiplication on G with an identity element, but not necessarily with inverses.

As a result, we find:

2.1.22 Proposition *The category* Ring *is complete.*

As we said, we will not yet show that Ring is cocomplete. But we can describe filtered colimits. In fact, filtered colimits will be constructed just as in the set-theoretic fashion. That is, the forgetful functor Ring $\to$ **Sets** commutes with *filtered* colimits (though not with general colimits).

2.1.23 Example (Filtered colimits) Let I be a filtering category, $F : I \to \mathsf{Ring}$ a functor. We can construct $\varinjlim_I F$ as follows. An object is an element (x, i) for $i \in I$ and $x \in F(i)$, modulo equivalence; we say that (x, i) and (y, j) are equivalent if there is a $k \in I$ with maps $i \to k, j \to k$ sending x, y to the same thing in the ring $F(k)$.

To multiply (x, i) and (y, j), we find some $k \in I$ receiving maps from i, j, and replace x, y with elements of $F(k)$. Then we multiply those two in $F(k)$. One easily sees that this is a well-defined multiplication law that induces a ring structure, and that what we have described is in fact the filtered colimit.

Zerodivisors

Let R be a commutative ring.

2.1.24 Definition If $r \in R$, then r is called a **zero divisor** if there is $s \in R, s \neq 0$ with $sr = 0$. Otherwise r is called a **non-zero-divisor.**

As an example, we prove a basic result on the zero divisors in a polynomial ring.

2.1.25 Proposition *Let $A = R[x]$. Let $f = a_n x^n + \cdots + a_0 \in A$. If there is a non-zero polynomial $g \in A$ such that $fg = 0$, then there exists $r \in R \smallsetminus \{0\}$ such that $f \cdot r = 0$.*

So all the coefficients are zero divisors.

Proof. Choose g to be of minimal degree, with leading coefficient bx^d. We may assume that $d > 0$. Then $f \cdot b \neq 0$, lest we contradict minimality of g. We must have $a_i g \neq 0$ for some i. To see this, assume that $a_i \cdot g = 0$, then $a_i b = 0$ for all i and then $fb = 0$. Now pick j to be the largest integer such that $a_j g \neq 0$. Then $0 = fg = (a_0 + a_1 x + \cdots a_j x^j)g$, and looking at the leading coefficient, we get $a_j b = 0$. So $\deg(a_j g) < d$. But then $f \cdot (a_j g) = 0$, contradicting minimality of g. ∎

2.1.26 Remark The product of two non-zero-divisors is a non-zero-divisor, and the product of two zero divisors is a zero divisor. It is, however, not necessarily true that the *sum* of two zero divisors is a zero divisor.

2.2. Further examples

We now illustrate a few important examples of commutative rings. The section is in large measure an advertisement for why one might care about commutative algebra; nonetheless, the reader is encouraged at least to skim this section.

Rings of holomorphic functions

The following subsec may be omitted without impairing understanding.

There is a fruitful analogy in number theory between the rings $\mathbb{Z}$ and $\mathbb{C}[t]$, the latter being the polynomial ring over $\mathbb{C}$ in one variable (**??**). Why are they analogous? Both of these rings have a theory of unique factorization: that is, factorization into primes or irreducible polynomials. (In the latter, the irreducible polynomials have degree one.) Indeed we know:

1. Any nonzero integer factors as a product of primes (possibly times -1).
2. Any nonzero polynomial factors as a product of an element of $\mathbb{C}^* = \mathbb{C} - \{0\}$ and polynomials of the form $t - a, a \in \mathbb{C}$.

There is another way of thinking of $\mathbb{C}[t]$ in terms of complex analysis. This is equal to the ring of holomorphic functions on $\mathbb{C}$ which are meromorphic at infinity. Alternatively, consider the Riemann sphere $\mathbb{C} \cup \{\infty\}$; then the ring $\mathbb{C}[t]$ consists of meromorphic functions on the sphere whose poles (if any) are at ∞. This description admits generalizations. Let X be a Riemann surface. (Example: take the complex numbers modulo a lattice, i.e. an elliptic curve.) Suppose that $x \in X$. Define R_x to be the ring of meromorphic functions on X which are allowed poles only at x (so are everywhere else holomorphic).

2.2.1 Example Fix the notations of the previous discussion. Fix $y \neq x \in X$. Let R_x be the ring of meromorphic functions on the Riemann surface X which are holomorphic on $X - \{x\}$, as before. Then the collection of functions that vanish at y forms an *ideal* in R_x.

There are lots of other ideals. For instance, fix two points $y_0, y_1 \neq x$; we look at the ideal of R_x that vanish at both y_0, y_1.

For any Riemann surface X, the conclusion of Dedekind's theorem (??) applies. In other words, the ring R_x as defined in the example admits unique factorization of ideals. We shall call such rings **Dedekind domains** in the future.

2.2.2 Example Keep the preceding notation.

Let $f \in R_x$, nonzero. By definition, f may have a pole at x, but no poles elsewhere. f vanishes at finitely many points $y_1, \ldots, y_m$. When X was the Riemann sphere, knowing the zeros of f told us something about f. Indeed, in this case f is just a polynomial, and we have a nice factorization of f into functions in R_x that vanish only at one point. In general Riemann surfaces, this is not generally possible. This failure turns out to be very interesting.

Let $X = \mathbb{C}/\Lambda$ be an elliptic curve (for $\Lambda \subset \mathbb{C}^2$ a lattice), and suppose $x = 0$. Suppose we are given $y_1, y_2, \ldots, y_m \in X$ that are nonzero; we ask whether there exists a function $f \in R_x$ having simple zeros at $y_1, \ldots, y_m$ and nowhere else. The answer is interesting, and turns out to recover the group structure on the lattice.

2.2.3 Proposition *A function $f \in R_x$ with simple zeros only at the $\{y_i\}$ exists if and only if $y_1 + y_2 + \cdots + y_n = 0$ (modulo Λ).*

So this problem of finding a function with specified zeros is equivalent to checking that the specific zeros add up to zero with the group structure.

In any case, there might not be such a nice function, but we have at least an ideal I of functions that have zeros (not necessarily simple) at $y_1, \dots, y_n$. This ideal has unique factorization into the ideals of functions vanishing at y_1, functions vanishing at y_2, so on.

Ideals and varieties

We saw in the previous subsec that ideals can be thought of as the vanishing of functions. This, like divisibility, is another interpretation, which is particularly interesting in algebraic geometry.

Recall the ring $\mathbb{C}[t]$ of complex polynomials discussed in the last subsec. More generally, if R is a ring, we saw in **??** that the set $R[t]$ of polynomials with coefficients in R is a ring. This is a construction that can be iterated to get a polynomial ring in several variables over R.

2.2.4 Example Consider the polynomial ring $\mathbb{C}[x_1, \dots, x_n]$. Recall that before we thought of the ring $\mathbb{C}[t]$ as a ring of meromorphic functions. Similarly each element of the polynomial ring $\mathbb{C}[x_1, \dots, x_n]$ gives a function $\mathbb{C}^n \to \mathbb{C}$; we can think of the polynomial ring as sitting inside the ring of all functions $\mathbb{C}^n \to \mathbb{C}$.

A question you might ask: What are the ideals in this ring? One way to get an ideal is to pick a point $x = (x_1, \dots, x_n) \in \mathbb{C}^n$; consider the collection of all functions $f \in \mathbb{C}[x_1, \dots, x_n]$ which vanish on x; by the usual argument, this is an ideal.

There are, of course, other ideals. More generally, if $Y \subset \mathbb{C}^n$, consider the collection of polynomial functions $f : \mathbb{C}^n \to \mathbb{C}$ such that $f \equiv 0$ on Y. This is easily seen to be an ideal in the polynomial ring. We thus have a way of taking a subset of $\mathbb{C}^n$ and producing an ideal. Let I_Y be the ideal corresponding to Y.

This construction is not injective. One can have $Y \neq Y'$ but $I_Y = I_{Y'}$. For instance, if Y is dense in $\mathbb{C}^n$, then $I_Y = (0)$, because the only way a continuous function on $\mathbb{C}^n$ can vanish on Y is for it to be zero.

There is a much closer connection in the other direction. You might ask whether all ideals can arise in this way. The quick answer is no—not even when $n = 1$. The ideal $(x^2) \subset \mathbb{C}[x]$ cannot be obtained in this way. It is easy to see that the only way we could get this as I_Y is for $Y = \{0\}$, but I_Y in this case is just (x), not (x^2). What's going wrong in this example is that (x^2) is not a *radical* ideal.

2.2.5 Definition An ideal $I \subset R$ is **radical** if whenever $x^2 \in I$, then $x \in I$.

The ideals I_Y in the polynomial ring are all radical. This is obvious. You might now ask whether this is the only obstruction. We now state a theorem that we will prove later.

2.2.6 Theorem (Hilbert's Nullstellensatz) *If $I \subset \mathbb{C}[x_1, \ldots, x_n]$ is a radical ideal, then $I = I_Y$ for some $Y \subset \mathbb{C}^n$. In fact, the canonical choice of Y is the set of points where all the functions in Y vanish.*[2]

This will be one of the highlights of the present course. But before we can get to it, there is much to do.

2.2.7 Remark Assuming the Nullstellensatz, show that any *maximal* ideal in the polynomial ring $\mathbb{C}[x_1, \ldots, x_n]$ is of the form $(x_1 - a_1, \ldots, x_n - a_n)$ for $a_1, \ldots, a_n \in \mathbb{C}$. An ideal of a ring is called **maximal** if the only ideal that contains it is the whole ring (and it itself is not the whole ring).

As a corollary, deduce that if $I \subset \mathbb{C}[x_1, \ldots, x_n]$ is a proper ideal (an ideal is called **proper** if it is not equal to the entire ring), then there exists $(x_1, \ldots, x_n) \in \mathbb{C}^n$ such that every polynomial in I vanishes on the point $(x_1, \ldots, x_n)$. This is called the **weak Nullstellensatz.**

2.3. Ideals

2.3.1 An *ideal* in a ring is analogous to a normal subgroup of a group. As we shall see, one may quotient by ideals just as one quotients by normal subgroups. The idea is that one wishes to have a suitable equivalence relation on a ring R such that the relevant maps (addition and multiplication) factor through this equivalence relation. It is easy to check that any such relation arises via an ideal, more precisely a two-sided ideal. Note that in the case where the ring is not assumed to be commutative

2.3.2 Definition Let R be a ring. An *ideal* in R is a subset $I \subset R$ that satisfies the following.

(I1) $(I, +, 0)$ is a subgroup of $(R, +, 0)$ that is $0 \in I$ and $x + y \in I$ for all $x, y \in I$.

(I2) If $x \in I$ and $y, z \in R$, then $xy \in I$ and $yz \in I$.

2.3.3 There is a simple way of obtaining ideals, which we now describe. Given elements $x_1, \ldots, x_n \in R$, we denote by $(x_1, \ldots, x_n) \subset R$ the subset of linear combinations $\sum_{i=1}^n r_i x_i s_i$, where $r_1, s_1, \ldots, r_n, s_n \in R$. This is clearly an ideal, and in fact the smallest one containing all x_i. It is called the ideal *generated* by $x_1, \ldots, x_n$. A *principal ideal* (x) is one generated by a single $x \in R$.

2.3.4 Example Ideals generalize the notion of divisibility. Note that in $\mathbb{Z}$, the set of elements divisible by $n \in \mathbb{Z}$ forms the ideal $I = n\mathbb{Z} = (n)$. We shall see that every ideal in $\mathbb{Z}$ is of this form: $\mathbb{Z}$ is a *principal ideal domain.*

Indeed, one can think of an ideal as axiomatizing the notions that "divisibility" ought to satisfy. Clearly, if two elements are divisible by something, then their sum and product should also be divisible by it. More generally, if an element is divisible by something, then the product of that

[2] Such a subset is called an algebraic variety.

element with anything else should also be divisible. In general, we will extend (in the chapter on Dedekind domains) much of the ordinary arithmetic with $\mathbb{Z}$ to arithmetic with *ideals* (e.g. unique factorization).

2.3.5 Example We saw in examples 2.1.5 that if X is a set and R a ring, then the set R^X of functions $X \to R$ is naturally a ring. If $Y \subset X$ is a subset, then the subset of functions vanishing on Y is an ideal.

2.3.6 Remark Show that the ideal $(2, 1+\sqrt{-5}) \subset \mathbb{Z}[\sqrt{-5}]$ is not principal.

Operations on ideals

There are a number of simple operations that one may do with ideals, which we now describe.

2.3.7 Definition The sum $I + J$ of two ideals $I, J \subset R$ is defined as the set of sums

$$\{x + y : x \in I, y \in J\}.$$

2.3.8 Definition The product IJ of two ideals $I, J \subset R$ is defined as the smallest ideal containing the products xy for all $x \in I, y \in J$. This is just the set

$$\left\{\sum x_i y_i : x_i \in I, y_i \in J\right\}.$$

We leave the basic verification of properties as an exercise:

2.3.9 Remark Given ideals $I, J \subset R$, verify the following.

1. $I + J$ is the smallest ideal containing I and J.
2. IJ is contained in I and J.
3. $I \cap J$ is an ideal.

2.3.10 Example In $\mathbb{Z}$, we have the following for any m, n.

1. $(m) + (n) = (\gcd\{m, n\})$,
2. $(m)(n) = (mn)$,
3. $(m) \cap (n) = (\operatorname{lcm}\{m, n\})$.

2.3.11 Proposition *For ideals $I, J, K \subset R$, we have the following.*

1. *Distributivity: $I(J + K) = IJ + IK$.*

Quotient rings

We next describe a procedure for producing new rings from old ones. If R is a ring and $I \subset R$ an ideal, then the quotient group R/I is a ring in its own right. If $a+I, b+I$ are two cosets, then the multiplication is $(a+I)(b+I) = ab+I$. It is easy to check that this does not depend on the coset representatives a, b. In other words, as mentioned earlier, the arithmetic operations on R factor *through the equivalence relation defined by I.*

As one easily checks, this becomes to a multiplication

$$R/I \times R/I \to R/I$$

which is commutative and associative, and whose identity element is $1+I$. In particular, R/I is a ring, under multiplication $(a+I)(b+I) = ab+I$.

2.3.12 Definition R/I is called the **quotient ring** by the ideal I.

The process is analogous to quotienting a group by a normal subgroup: again, the point is that the equivalence relation induced on the algebraic structure—either the group or the ring—by the subgroup (or ideal)—is compatible with the algebraic structure, which thus descends to the quotient.

The reduction map $\phi\colon R \to R/I$ is a ring-homomorphism with a universal property*. Namely, for any ring B, there is a map*

$$\hom(R/I, B) \to \hom(R, B)$$

on the hom-sets by composing with the ring-homomorphism ϕ; this map is injective and the image consists of all homomorphisms $R \to B$ which vanish on I. Stated alternatively, to map out of R/I (into some ring B) is the same thing as mapping out of R while killing the ideal $I \subset R$.

This is best thought out for oneself, but here is the detailed justification. The reason is that any map $R/I \to B$ pulls back to a map $R \to R/I \to B$ which annihilates I since $R \to R/I$ annihilates I. Conversely, if we have a map

$$f : R \to B$$

killing I, then we can define $R/I \to B$ by sending $a+I$ to $f(a)$; this is uniquely defined since f annihilates I.

2.3.13 Remark If R is a commutative ring, an element $e \in R$ is said to be **idempotent** if $e^2 = e$. Define a covariant functor **Rings** $\to$ **Sets** sending a ring to its idempotents. Prove that it is corepresentable. (Answer: the corepresenting object is $\mathbb{Z}[X]/(X - X^2)$.)

2.3.14 Remark Show that the functor assigning to each ring the set of elements annihilated by 2 is corepresentable.

2.3.15 Remark If $I \subset J \subset R$, then J/I is an ideal of R/I, and there is a canonical isomorphism

$$(R/I)/(J/I) \simeq R/J.$$

2. $I \cap (J + K) = I \cap J + I \cap K$ *if* $I \supset J$ *or* $I \supset K$.
3. *If* $I + J = R$, $I \cap J = IJ$.

Proof. 1 and 2 are clear. For 3, note that $(I + J)(I \cap J) = I(I \cap J) + J(I \cap J) \subset IJ$. Since $IJ \subset I \cap J$, the result follows. □

2.3.16 Remark There is a *contravariant* functor **Rings** → **Sets** that sends each ring to its set of ideals. Given a map $f : R \to S$ and an ideal $I \subset S$, we define an ideal $f^{-1}(I) \subset R$; this defines the functoriality. This functor is not representable, as it does not send the initial object in **Rings** to the one-element set. We will later use a *subfunctor* of this functor, the Spec construction, when we replace ideals with "prime" ideals.

2.4. Introduction

In this chapter we will introduce the notions of a ring and that of a module over a ring. The focus of the present book will be on commutative rings, though, and the spaces represented by them. Most of the chapter will be definitions.

We begin with a few historical remarks on the origin of commutative ring theory. Fermat's last theorem states that the equation

$$x^n + y^n = z^n$$

has no nontrivial solutions in the integers, for $n \geqslant 3$. We could try to prove this by factoring the expression on the left hand side. We can write

$$(x + y)(x + \zeta y)(x + \zeta^2 y) \dots (x + \zeta^{n-1} y) = z^n,$$

where ζ is a primitive nth root of unity. Unfortunately, the factors lie in $\mathbb{Z}[\zeta]$, not the integers $\mathbb{Z}$. Though $\mathbb{Z}[\zeta]$ is still a *ring* where we have notions of primes and factorization, just as in $\mathbb{Z}$, we will see that prime factorization is not always unique in $\mathbb{Z}[\zeta]$. (If it were always unique, then we could at least one important case of Fermat's last theorem rather easily; see the introductory chapter of **?** for an argument.)

For instance, consider the ring $\mathbb{Z}[\sqrt{-5}]$ of complex numbers of the form $a + b\sqrt{-5}$, where $a, b \in \mathbb{Z}$. Then we have the two factorizations

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

Both of these are factorizations of 6 into irreducible factors, but they are fundamentally different.

In part, commutative algebra grew out of the need to understand this failure of unique factorization more generally. We shall have more to say on factorization in the future, but here we just focus on the formalism. The basic definition for studying this problem is that of a *ring*, which we now introduce.

2.5. Modules over a ring

We will now establish some basic terminology about modules. Throughout this section, R denotes always a ring.

Definitions

2.5.1 Definition A *left R-module* M is an abelian group $(M, +)$ together with a map $\cdot : R \times M \to M$, which is usually called the *scalar multiplication* and written $(x, m) \mapsto xm$, such that

(Mod1) Scalar multiplication is associative, i.e. $(xy)m = x(ym)$ for all $x, y \in R$ and $m \in M$.

(Mod2) The unit $1 \in R$ acts as identity that means $1 \cdot m = m$ for all $m \in M$.

(Mod3) There are distributive laws on both sides:
$(x + y)m = xm + ym$ and $x(m + n) = xm + xn$ for all $x, y \in R$ and $m, n \in M$.

A *right R-module* N is an abelian group $(N, +)$ together with a map $\cdot : N \times R \to N$, which is usually called *scalar multiplication* as well and written $(n, y) \mapsto ny$, such that

(Mod 1)$^\circ$ Scalar multiplication is associative, i.e. $n(xy) = (nx)y$ for all $x, y \in R$ and $n \in N$.

(Mod 2)$^\circ$ The unit $1 \in R$ acts as identity that means $n \cdot 1 = n$ for all $n \in N$.

(Mod 3)$^\circ$ There are distributive laws on both sides:
$n(x + y) = nx + ny$ and $(m + n)y = my + ny$ for all $x, y \in R$ and $m, n \in N$.

By an *R-module* we always understand a left R-module if not explicitely mentioned differently.

2.5.2 Remark Another definition of a left R module can be given as follows. If M is an abelian group, $\mathrm{End}(M)$ is the set of homomorphisms $f : M \to M$. This can be made into a (noncommutative) ring. Addition is defined pointwise, and multiplication is by composition. The identity element is the identity function id_M. If R is a ring and $R \to \mathrm{End}(M)$ a homomorphism, then M is made into a left R-module, and vice versa.

2.5.3 Examples (a) If R is a ring, then R is a left R-module by multiplication on the left, and a right R-module by multiplication on the right.

(b) A $\mathbb{Z}$-module is the same thing as an abelian group.

2.5.4 Definition If M is a left (respectively right) R-module, a non-empty subset $N \subset M$ is a *submodule* if it is an additive subgroup (meaning closed under addition and inversion) and is closed under multiplication by elements of R, i.e. $aN \subset N$ (respectively $Na \subset N$) for $a \in R$. A submodule is a left (respectively right) R-module in its own right. If $N \subset M$ is a submodule, there is a commutative diagram:

$$\begin{array}{ccccccc} R \times M_0 & \longrightarrow & M_0 & \quad \text{respectively} \quad & N \times R & \longrightarrow & N \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ R \times M & \longrightarrow & M & & M \times R & \longrightarrow & M \ , \end{array}$$

depending on whether M is a left or right R-module. Here the horizontal maps are multiplication by scalars.

2.5.5 Examples (a) Let R be a commutative ring; then an ideal in R is the same thing as a submodule of R.

(b) If R is a commutative ring, an R-algebra is an R-module in an obvious way. More generally, if R is a commutative ring and A is an R-algebra, any A-module becomes an R-module by pulling back the multiplication map via $R \to A$.

Dual to submodules is the notion of a quotient module, which we define next.

2.5.6 Definition Suppose M is an R-module and N a submodule. Then the abelian group $M/N = \{m + N \in \mathcal{P}(M) \mid m \in M\}$ (of cosets) is an R-module, called the *quotient module* of M by by N. Multiplication is as follows. If one has a coset $m + N \in M/N$, one multiplies this by $a \in R$ to get the coset $ax + M_0$. This does not depend on the coset representative.

The categorical structure on modules

So far, we have talked about modules, but we have not discussed morphisms between modules, and have yet to make the class of modules over a given ring into a category. This we do next.

Let us thus introduce a few more basic notions.

2.5.7 Definition Let R be a ring. Suppose M, N are R-modules. A map $f : M \to N$ is a **module-homomorphism** if it preserves all the relevant structures. Namely, it must be a homomorphism of abelian groups, $f(x + y) = f(x) + f(y)$, and second it must preserve multiplication:

$$f(ax) = af(x)$$

for $a \in R, x \in M$.

A simple way of getting plenty of module-homomorphisms is simply to consider multiplication by a fixed element of the ring.

2.5.8 Example If R is a commutative ring, M an R-module, and $a \in R$, then multiplication by a is a module-homomorphism $M \xrightarrow{a} M$ for any R-module M. Such homomorphisms are called *homotheties*. When one considers modules over noncommutative rings, this is no longer true.

If $M \xrightarrow{f} N$ and $N \xrightarrow{g} P$ are module-homomorphisms, their composite $M \xrightarrow{g \circ f} P$ clearly is too. Thus, for any commutative ring R, the class of R-modules and module-homomorphisms forms a **category**.

2.5.9 Remark The initial object in this category is the zero module, and this is also the final object.

In general, a category where the initial object and final object are the same (that is, isomorphic) is called a *pointed category*. The common object is called the *zero object*. In a pointed category

$\mathcal{C}$, there is a morphism $X \to Y$ for any two objects $X, Y \in \mathcal{C}$: if $*$ is the zero object, then we can take $X \to * \to Y$. This is well-defined and is called the *zero morphism.* One can easily show that the composition (on the left or the right) of a zero morphism is a zero morphism (between a possibly different set of objects).

In the case of the category of modules, the zero object is clearly the zero module, and the zero morphism $M \to N$ sends $m \mapsto 0$ for each $m \in M$.

2.5.10 Definition Let $f : M \to N$ be a module homomorphism. In this case, the *kernel* $\operatorname{Ker} f$ of f is the set of elements $m \in M$ with $f(m) = 0$. This is a submodule of M, as is easy to see.

The *image* $\operatorname{Im} f$ of f (the set-theoretic image, i.e. the collection of all $f(x)$, $x \in M$) is also a submodule of N.

The *cokernel* of f is defined by $N/\operatorname{Im}(f)$.

2.5.11 Remark The universal property of the kernel is as follows. Let $M \xrightarrow{f} N$ be a morphism with kernel $K \subset M$. Let $T \to M$ be a map. Then $T \to M$ factors through the kernel $K \to M$ if and only if its composition with f (a morphism $T \to N$) is zero. That is, an arrow $T \to K$ exists in the diagram (where the dotted arrow indicates we are looking for a map that need not exist)

$$\begin{tikzcd} & T \arrow[dl, dashed] \arrow[d] & \\ K \arrow[r] & M \arrow[r, "f"] & N \end{tikzcd}$$

if and only if the composite $T \to N$ is zero. In particular, if we think of the hom-sets as abelian groups (i.e. $\mathbb{Z}$-modules)

$$\hom_R(T, K) = \ker\left(\hom_R(T, M) \to \hom_R(T, N)\right).$$

In other words, one may think of the kernel as follows. If $X \xrightarrow{f} Y$ is a morphism, then the kernel $\ker(f)$ is the equalizer of f and the zero morphism $X \xrightarrow{0} Y$.

2.5.12 Remark What is the universal property of the cokernel?

2.5.13 Remark On the category of modules, the functor assigning to each module M its underlying set is corepresentable (cf. remark 2.1.19). What is the corepresenting object?

We shall now introduce the notions of *direct sum* and *direct product*. Let I be a set, and suppose that for each $i \in I$, we are given an R-module M_i.

2.5.14 Definition The **direct product** $\prod M_i$ is set-theoretically the cartesian product. It is given the structure of an R-module by addition and multiplication pointwise on each factor.

2.5.15 Definition The **direct sum** $\bigoplus_I M_i$ is the set of elements in the direct product such that all but finitely many entries are zero. The direct sum is a submodule of the direct product.

2.5.16 Example The direct product is a product in the category of modules, and the direct sum is a coproduct. This is easy to verify: given maps $f_i : M \to M_i$, then we get get a unique map $f : M \to \prod M_i$ by taking the product in the category of sets. The case of a coproduct is dual: given maps $g_i : M_i \to N$, then we get a map $\bigoplus M_i \to N$ by taking the *sum* g of the g_i: on a family $(m_i) \in \bigoplus M_i$, we take $g(m_i) = \sum_I g_i(m_i)$; this is well-defined as almost all the m_i are zero.

example 2.5.16 shows that the category of modules over a fixed commutative ring has products and coproducts. In fact, the category of modules is both complete and cocomplete (see definition 1.5.68 for the definition). To see this, it suffices to show that (by **??** and its dual) that this category admits equalizers and coequalizers.

The equalizer of two maps

$$M \overset{f,g}{\rightrightarrows} N$$

is easily checked to be the submodule of M consisting of $m \in M$ such that $f(m) = g(m)$, or, in other words, the kernel of $f - g$. The coequalizer of these two maps is the quotient module of N by the submodule $\{f(m) - g(m), m \in M\}$, or, in other words, the cokernel of $f - g$.

Thus:

2.5.17 Proposition *If R is a ring, the category of R-modules is complete and cocomplete.*

2.5.18 Example Note that limits in the category of R-modules are calculated in the same way as they are for sets, but colimits are not. That is, the functor from R-modules to **Sets**, the forgetful functor, preserves limits but not colimits. Indeed, we will see that the forgetful functor is a right adjoint (proposition 2.8.3), which implies it preserves limits (by proposition 1.7.10).

Exactness

Finally, we introduce the notion of *exactness.*

2.5.19 Definition Let $f : M \to N$ be a morphism of R-modules. Suppose $g : N \to P$ is another morphism of R-modules. The pair of maps is a **complex** if $g \circ f = 0 : M \to N \to P$. This is equivalent to the condition that $\mathrm{Im}(f) \subset \mathrm{Ker}(g)$.

This complex is *exact* (or *exact at N*) if $\mathrm{Im}(f) = \mathrm{Ker}(g)$. In other words, anything that is killed when mapped to P actually comes from something in M.

We shall often write pairs of maps as sequences

$$A \xrightarrow{f} B \xrightarrow{g} C$$

and say that the sequence is exact if the pair of maps is, as in Definition 2.5.19. A longer (possibly infinite) sequence of modules

$$A_0 \to A_1 \to A_2 \to \dots$$

will be called a **complex** if each set of three consecutive terms is a complex, and **exact** if it is exact at each step.

2.5.20 Example The sequence $0 \to A \xrightarrow{f} B$ is exact if and only if the map f is injective. Similarly, $A \xrightarrow{f} B \to 0$ is exact if and only if f is surjective. Thus, $0 \to A \xrightarrow{f} B \to 0$ is exact if and only if f is an isomorphism.

One typically sees this definition applied to sequences of the form

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0,$$

which, if exact, is called a **short exact sequence**. Exactness here means that f is injective, g is surjective, and f maps onto the kernel of g. So M'' can be thought of as the quotient M/M'.

2.5.21 Example Conversely, if M is a module and $M' \subset M$ a submodule, then there is a short exact sequence

$$0 \to M' \to M \to M/M' \to 0.$$

So every short exact sequence is of this form.

Suppose F is a functor from the category of R-modules to the category of S-modules, where R, S are rings. Then:

2.5.22 Definition 1. F is called **additive** if F preserves direct sums.

2. F is called **exact** if F is additive and preserves exact sequences.
3. F is called **left exact** if F is additive and preserves exact sequences of the form $0 \to M' \to M \to M''$. Equivalently, F preserves kernels.
4. F is **right exact** if F is additive and F preserves exact sequences of the form $M' \to M \to M'' \to 0$, i.e. F preserves cokernels.

The reader should note that much of homological algebra can be developed using the more general setting of an *abelian category,* which axiomatizes much of the standard properties of the category of modules over a ring. Such a generalization turns out to be necessary when many natural categories, such as the category of chain complexes or the category of sheaves on a topological space, are not naturally categories of modules. We do not go into this here, cf. **?**.

A functor F is exact if and only if it is both left and right exact. This actually requires proof, though it is not hard. Namely, right-exactness implies that F preserves cokernels. Left-exactness implies that F preserves kernels. F thus preserves images, as the image of a morphism is the kernel of its cokernel. So if

$$A \to B \to C$$

is a short exact sequence, then the kernel of the second map is equal to the image of the first; we have just seen that this is preserved under F.

From this, one can check that left-exactness is equivalent to requiring that F preserve finite limits (as an additive functor, F automatically preserves products, and we have just seen that F is left-exact iff it preserves kernels). Similarly, right-exactness is equivalent to requiring that F preserve finite colimits. So, in *any* category with finite limits and colimits, we can talk about right or left exactness of a functor, but the notion is used most often for categories with an additive structure (e.g. categories of modules over a ring).

2.5.23 Remark Suppose whenever $0 \to A' \to A \to A'' \to 0$ is short exact, then $FA' \to FA \to FA'' \to 0$ is exact. Prove that F is right-exact. So we get a slightly weaker criterion for right-exactness.

Do the same for left-exact functors.

Split exact sequences

Let $f : A \to B$ be a map of sets which is injective. Then there is a map $g : A \to B$ such that the composite $g \circ f : A \xrightarrow{f} B \xrightarrow{g} A$ is the identity. Namely, we define g to be the inverse of f on $f(A)$ and arbitrarily on $B - f(A)$. Conversely, if $f : A \to B$ admits an element $g : B \to A$ such that $g \circ f = 1_A$, then f is injective. This is easy to see, as any $a \in A$ can be "recovered" from $f(a)$ (by applying g).

In general, however, this observation does not generalize to arbitrary categories.

2.5.24 Definition Let $\mathcal{C}$ be a category. A morphism $A \xrightarrow{f} B$ is called a **split injection** if there is $g : B \to A$ with $g \circ f = 1_A$.

2.5.25 Remark (General nonsense) Suppose $f : A \to B$ is a split injection. Show that f is a categorical monomorphism. (Idea: the map $\hom(C, A) \to \hom(C, B)$ becomes a split injection of sets thanks to g.)

add: what is a categorical monomorphism? Maybe omit the exercise

In the category of sets, we have seen above that *any* monomorphism is a split injection. This is not true in other categories, in general.

2.5.26 Remark Consider the morphism $\mathbb{Z} \to \mathbb{Z}$ given by multiplication by 2. Show that this is not a split injection: no left inverse g can exist.

We are most interested in the case of modules over a ring.

2.5.27 Proposition *A morphism $f : A \to B$ in the category of R-modules is a split injection if and only if:*

1. *f is injective.*
2. *$f(A)$ is a direct summand in B.*

The second condition means that there is a submodule $B' \subset B$ such that $B = B' \oplus f(A)$ (internal direct sum). In other words, $B = B' + f(A)$ and $B' \cap f(A) = \{0\}$.

Proof. Suppose the two conditions hold, and we have a module B' which is a complement to $f(A)$. Then we define a left inverse

$$B \xrightarrow{g} A$$

by letting $g|_{f(A)} = f^{-1}$ (note that f becomes an *isomorphism* $A \to f(A)$) and $g|_{B'} = 0$. It is easy to see that this is indeed a left inverse, though in general not a right inverse, as g is likely to be non-injective.

Conversely, suppose $f : A \to B$ admits a left inverse $g : B \to A$. The usual argument (as for sets) shows that f is injective. The essentially new observation is that $f(A)$ is a direct summand in B. To define the complement, we take $\ker(g) \subset B$. It is easy to see (as $g \circ f = 1_A$) that $\ker(g) \cap f(A) = \{0\}$. Moreover, $\ker(g) + f(A)$ fills B: given $b \in B$, it is easy to check that

$$b - f(g(b)) \in \ker(g).$$

Thus we find that the two conditions are satisfied. ∎

add: further explanation, exactness of filtered colimits

The five lemma

The five lemma will be a useful tool for us in proving that maps are isomorphisms. Often this argument is used in inductive proofs. Namely, we will see that often "long exact sequences" (extending infinitely in one or both directions) arise from short exact sequences in a natural way. In such events, the five lemma will allow us to prove that certain morphisms are isomorphisms by induction on the dimension.

2.5.28 Theorem *Suppose given a commutative diagram*

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[r] \arrow[d] & D \arrow[r] \arrow[d] & E \arrow[d] \\ A' \arrow[r] & B' \arrow[r] & C' \arrow[r] & D' \arrow[r] & E' \end{tikzcd}$$

such that the rows are exact and the four vertical maps $A \to A', B \to B', D \to D', E \to E'$ are isomorphisms. Then $C \to C'$ is an isomorphism.

This is the type of proof that goes by the name of "diagram-chasing," and is best thought out visually for oneself, even though we give a complete proof.

Proof. We have the diagram

$$\begin{tikzcd} A \arrow[r, "k"] \arrow[d, "a"] & B \arrow[r, "l"] \arrow[d, "b"] & C \arrow[r, "m"] \arrow[d, "g"] & D \arrow[r, "n"] \arrow[d, "d"] & E \arrow[d, "e"] \\ F \arrow[r, "p"'] & G \arrow[r, "q"'] & H \arrow[r, "r"'] & I \arrow[r, "s"'] & J \end{tikzcd}$$

∎

where the rows are exact at B, C, D, G, H, I and the squares commute. In addition, suppose that a, b, d, e are isomorphisms. We will show that g is an isomorphism.

We show that g is surjective:

Suppose that $h \in H$. Since d is surjective, there exists an element $d \in D$ such that $r(h) = d(d) \in I$. By the commutativity of the rightmost square, $s(r(h)) = e(n(d))$. The exactness at I means that $\operatorname{Im} r = \ker s$, so hence $e(n(d)) = s(r(h)) = 0$. Because e is injective, $n(d) = 0$. Then $d \in \operatorname{Ker}(n) = \operatorname{Im}(m)$ by exactness at D. Therefore, there is some $c \in C$ such that $m(c) = d$. Now, $d(m(c)) = d(d) = r(h)$ and by the commutativity of squares, $d(m(c)) = r(g(c))$, so therefore $r(g(c)) = r(h)$. Since r is a homomorphism, $r(g(c) - h) = 0$. Hence $g(c) - h \in \ker r = \operatorname{Im} q$ by exactness at H.

Therefore, there exists $g \in G$ such that $q(g) = g(c) - h$. b is surjective, so there is some $b \in B$ such that $b(b) = g$ and hence $q(b(b)) = g(c) - h$. By the commutativity of squares, $q(b(b)) = g(l(b)) = g(c) - h$. Hence $h = g(c) - g(l(b)) = g(c - l(b))$, and therefore g is surjective.

So far, we've used that b and g are surjective, e is injective, and exactness at D, H, I.

We show that g is injective:

Suppose that $c \in C$ and $g(c) = 0$. Then $r(g(c)) = 0$, and by the commutativity of squares, $d(m(c)) = 0$. Since d is injective, $m(c) = 0$, so $c \in \ker m = \operatorname{Im} l$ by exactness at C. Therefore, there is $b \in B$ such that $l(b) = c$. Then $g(l(b)) = g(c) = 0$, and by the commutativity of squares, $q(b(b)) = 0$. Therefore, $b(b) \in \ker q$, and by exactness at G, $b(b) \in \ker q = \operatorname{Im} p$.

There is now $f \in F$ such that $p(f) = b(b)$. Since a is surjective, this means that there is $a \in A$ such that $f = a(a)$, so then $b(b) = p(a(a))$. By commutativity of squares, $b(b) = p(a(a)) = b(k(a))$, and hence $b(k(a) - b) = 0$. Since b is injective, we have $k(a) - b = 0$, so $k(a) = b$. Hence $b \in \operatorname{Im} k = \ker l$ by commutativity of squares, so $l(b) = 0$. However, we defined b to satisfy $l(b) = c$, so therefore $c = 0$ and hence g is injective.

Here, we used that a is surjective, b, d are injective, and exactness at B, C, G.

Putting the two statements together, we see that g is both surjective and injective, so g is an isomorphism. We only used that b, d are isomorphisms and that a is surjective, e is injective, so we can slightly weaken the hypotheses; injectivity of a and surjectivity of e were unnecessary.

2.6. Ideals in commutative rings

The notion of an *ideal* has already been defined. Now we will introduce additional terminology related to the theory of ideals.

Prime and maximal ideals

Recall that the notion of an ideal generalizes that of divisibility. In elementary number theory, though, one finds that questions of divisibility basically reduce to questions about primes. The notion of a "prime ideal" is intended to generalize the familiar idea of a prime number.

2.6.1 Definition An ideal $I \subset R$ is said to be *prime* if

(PI1) $1 \notin I$ (by convention, 1 is not a prime number).

(PI2) If $xy \in I$, either $x \in I$ or $y \in I$.

2.6.2 Example If $R = \mathbb{Z}$ and $p \in R$, then $(p) \subset \mathbb{Z}$ is a prime ideal if and only if p or $-p$ is a prime number in $\mathbb{N}$ or if p is zero.

2.6.3 Example If R is any commutative ring, there are two obvious ideals. These obvious ones are the zero ideal (0) consisting only of the zero element, and the unit element (1) consisting of all of R.

2.6.4 Definition An ideal $I \subset R$ is called *maximal* if

(MI1) $1 \notin I$.

(MI2) Any larger ideal contains 1 (i.e., is all of R).

2.6.5 Remark So a maximal ideal is a maximal element in the partially ordered set of proper ideals. Recall that an ideal is called *proper* if it does not contain 1.

2.6.6 Remark Find the maximal ideals in $\mathbb{C}[t]$.

2.6.7 Proposition *A maximal ideal is prime.*

Proof. First, a maximal ideal does not contain 1.

Let $I \subset R$ be a maximal ideal. We need to show that if $xy \in I$, then one of x, y is in I. If $x \notin I$, then $(I, x) = I + (x)$ (the ideal generated by I and x) strictly contains I, so by maximality contains 1. In particular, $1 \in I + (x)$, so we can write

$$1 = a + xb$$

where $a \in I$, $b \in R$. Multiply both sides by y:

$$y = ay + bxy.$$

Both terms on the right here are in I ($a \in I$ and $xy \in I$), so we find that $y \in I$. □

Given a ring R, what can we say about the collection of ideals in R? There are two obvious ideals in R, namely (0) and (1). These are the same if and only if $0 = 1$, i.e. R is the zero ring. So for any nonzero commutative ring, we have at least two distinct ideals.

Next, we show that maximal ideals always *do* exist, except in the case of the zero ring.

2.6.8 Proposition *Let R be a commutative ring. Let $I \subset R$ be a proper ideal. Then I is contained in a maximal ideal.*

Proof. This requires the axiom of choice in the form of Zorn's lemma. Let P be the collection of all ideals $J \subset R$ such that $I \subset J$ and $J \neq R$. Then P is a poset with respect to inclusion. P is nonempty because it contains I. Note that given a (nonempty) linearly ordered collection of ideals $J_\alpha \in P$, the union $\bigcup J_\alpha \subset R$ is an ideal: this is easily seen in view of the linear ordering (if $x, y \in \bigcup J_\alpha$, then both x, y belong to some J_γ, so $x + y \in J_\gamma$; multiplicative closure is even easier). The union is not all of R because it does not contain 1.

This implies that P has a maximal element by Zorn's lemma. This maximal element may be called $\mathfrak{M}$; it's a proper element containing I. I claim that $\mathfrak{M}$ is a maximal ideal, because if it were contained in a larger ideal, that would be in P (which cannot happen by maximality) unless it were all of R. □

2.6.9 Corollary *Let R be a nonzero commutative ring. Then R has a maximal ideal.*

Proof. Apply the lemma to the zero ideal. □

2.6.10 Corollary *Let R be a nonzero commutative ring. Then $x \in R$ is invertible if and only if it belongs to no maximal ideal $\mathfrak{m} \subset R$.*

Proof. Indeed, x is invertible if and only if $(x) = 1$. That is, if and only if (x) is not a proper ideal; now proposition 2.6.8 finishes the argument. □

Fields and integral domains

Recall:

2.6.11 Definition A commutative ring R is called a **field** if $1 \neq 0$ and for every $x \in R - \{0\}$ there exists an **inverse** $x^{-1} \in R$ such that $xx^{-1} = 1$.

This condition has an obvious interpretation in terms of ideals.

2.6.12 Proposition *A commutative ring with $1 \neq 0$ is a field iff it has only the two ideals $(1), (0)$.*

Alternatively, a ring is a field if and only if (0) is a maximal ideal.

Proof. Assume R is a field. Suppose $I \subset R$. If $I \neq (0)$, then there is a nonzero $x \in I$. Then there is an inverse x^{-1}. We have $x^{-1}x = 1 \in I$, so $I = (1)$. In a field, there is thus no room for ideals other than (0) and (1).

To prove the converse, assume every ideal of R is (0) or (1). Then for each $x \in R$, $(x) = (0)$ or (1). If $x \neq 0$, the first cannot happen, so that means that the ideal generated by x is the unit ideal. So 1 is a multiple of x, implying that x has a multiplicative inverse. □

So fields also have an uninteresting ideal structure.

2.6.13 Corollary *If R is a ring and $I \subset R$ is an ideal, then I is maximal if and only if R/I is a field.*

Proof. The basic point here is that there is a bijection between the ideals of R/I and ideals of R containing I.

Denote by $\phi : R \to R/I$ the reduction map. There is a construction mapping ideals of R/I to ideals of R. This sends an ideal in R/I to its inverse image. This is easily seen to map to ideals of R containing I. The map from ideals of R/I to ideals of R containing I is a bijection, as one checks easily.

It follows that R/I is a field precisely if R/I has precisely two ideals, i.e. precisely if there are precisely two ideals in R containing I. These ideals must be (1) and I, so this holds if and only if I is maximal. $\square$

There is a similar characterization of prime ideals.

2.6.14 Definition A commutative ring R is an **integral domain** if for all $x, y \in R$, $x \neq 0$ and $y \neq 0$ imply $xy \neq 0$.

2.6.15 Proposition *An ideal $I \subset R$ is prime iff R/I is a domain.*

2.6.16 Remark Prove proposition 2.6.15.

Any field is an integral domain. This is because in a field, nonzero elements are invertible, and the product of two invertible elements is invertible. This statement translates in ring theory to the statement that a maximal ideal is prime.

Finally, we include an example that describes what *some* of the prime ideals in a polynomial ring look like.

2.6.17 Example Let R be a ring and P a prime ideal. We claim that $PR[x] \subset R[x]$ is a prime ideal.

Consider the map $\tilde{\phi} : R[x] \to (R/P)[x]$ with $\tilde{\phi}(a_0 + \cdots + a_n x^n) = (a_0 + P) + \cdots + (a_n + P)x^n$. This is clearly a homomorphism because $\phi : R \to R/P$ is, and its kernel consists of those polynomials $a_0 + \cdots + a_n x^n$ with $a_0, \ldots, a_n \in P$, which is precisely $P[x]$. Thus $R[x]/P[x] \simeq (R/P)[x]$, which is an integral domain because R/P is an integral domain. Thus $P[x]$ is a prime ideal.

However, if P is a maximal ideal, then $P[x]$ is never a maximal ideal because the ideal $P[x] + (x)$ (the polynomials with constant term in P) always strictly contains $P[x]$ (because if $x \in P[x]$ then $1 \in P$, which is impossible). Note that $P[x] + (x)$ is the kernel of the composition of $\tilde{\phi}$ with evaluation at 0, i.e $(\mathrm{ev}_0 \circ \tilde{\phi}) : R[x] \to R/P$, and this map is a surjection and R/P is a field, so that $P[x] + (x)$ is the maximal ideal in $R[x]$ containing $P[x]$.

2.6.18 Remark Let R be a domain. Consider the set of formal quotients $a/b, a, b \in R$ with $b \neq 0$. Define addition and multiplication using usual rules. Show that the resulting object $K(R)$ is a ring, and in fact a *field*. The natural map $R \to K(R)$, $r \to r/1$, has a universal property. If $R \hookrightarrow L$ is an injection of R into a field L, then there is a unique morphism $K(R) \to L$ of fields extending $R \to L$. This construction will be generalized when we consider *localization.* This construction is called the **quotient field.**

Note that a non-injective map $R \to L$ will *not* factor through the quotient field!

2.6.19 Remark Let R be a commutative ring. Then the **Jacobson radical** of R is the intersection $\bigcap \mathfrak{m}$ of all maximal ideals $\mathfrak{m} \subset R$. Prove that an element x is in the Jacobson radical if and only if $1 - yx$ is invertible for all $y \in R$.

Prime avoidance

The following fact will come in handy occasionally. We will, for instance, use it much later to show that an ideal consisting of zero divisors on a module M is contained in associated prime.

2.6.20 Theorem (Prime avoidance) *Let $I_1, \dots, I_n \subset R$ be ideals. Let $A \subset R$ be a subset which is closed under addition and multiplication. Assume that at least $n - 2$ of the ideals are prime. If $A \subset I_1 \cup \dots \cup I_n$, then $A \subset I_j$ for some j.*

The result is frequently used in the following specific case: if an ideal I is contained in a finite union $\bigcup \mathfrak{p}_i$ of primes, then $I \subset \mathfrak{p}_i$ for some i.

Proof. Induct on n. If $n = 1$, the result is trivial. The case $n = 2$ is an easy argument: if $a_1 \in A \smallsetminus I_1$ and $a_2 \in A \smallsetminus I_2$, then $a_1 + a_2 \in A \smallsetminus (I_1 \cup I_2)$.

Now assume $n \geqslant 3$. We may assume that for each j, $A \not\subset I_1 \cup \dots \cup \hat{I_j} \cup \cdots I_n$.[3] Fix an element $a_j \in A \smallsetminus (I_1 \cup \dots \cup \hat{I_j} \cup \cdots I_n)$. Then this a_j must be contained in I_j since $A \subset \bigcup I_j$. Since $n \geqslant 3$, one of the I_j must be prime. We may assume that I_1 is prime. Define $x = a_1 + a_2 a_3 \cdots a_n$, which is an element of A. Let's show that x avoids *all* of the I_j. If $x \in I_1$, then $a_2 a_3 \cdots a_n \in I_1$, which contradicts the fact that $a_i \notin I_j$ for $i \neq j$ and that I_1 is prime. If $x \in I_j$ for $j \geqslant 2$. Then $a_1 \in I_j$, which contradicts $a_i \notin I_j$ for $i \neq j$. ∎

The Chinese remainder theorem

Let m, n be relatively prime integers. Suppose $a, b \in \mathbb{Z}$; then one can show that the two congruences $x \equiv a \mod m$ and $x \equiv b \mod n$ can be solved simultaneously in $x \in \mathbb{Z}$. The solution is unique, moreover, modulo mn. The Chinese remainder theorem generalizes this fact:

2.6.21 Theorem (Chinese remainder theorem) *Let $I_1, \dots I_n$ be ideals in a ring R which satisfy $I_i + I_j = R$ for $i \neq j$. Then we have $I_1 \cap \dots \cap I_n = I_1 \dots I_n$ and the morphism of rings*

$$R \to \bigoplus R/I_i$$

is an epimorphism with kernel $I_1 \cap \dots \cap I_n$.

Proof. First, note that for any two ideals I_1 and I_2, we have $I_1 I_2 \subset I_1 \cap I_2$ and $(I_1 + I_2)(I_1 \cap I_2) \subset I_1 I_2$ (because any element of $I_1 + I_2$ multiplied by any element of $I_1 \cap I_2$ will clearly be a sum of products of elements from both I_1 and I_2). Thus, if I_1 and I_2 are coprime, i.e. $I_1 + I_2 = (1) = R$, then $(1)(I_1 \cap I_2) = (I_1 \cap I_2) \subset I_1 I_2 \subset I_1 \cap I_2$, so that $I_1 \cap I_2 = I_1 I_2$. This establishes the result for $n = 2$.

[3] The hat means omit I_j.

If the ideals $I_1, \ldots, I_n$ are pairwise coprime and the result holds for $n-1$, then

$$\bigcap_{i=1}^{n-1} I_i = \prod_{i=1}^{n-1} I_i.$$

Because $I_n + I_i = (1)$ for each $1 \leqslant i \leqslant n-1$, there must be $x_i \in I_n$ and $y_i \in I_i$ such that $x_i + y_i = 1$. Thus, $z_n = \prod_{i=1}^{n-1} y_i = \prod_{i=1}^{n-1}(1 - x_i) \in \prod_{i=1}^{n-1} I_i$, and clearly $z_n + I_n = 1 + I_n$ since each $x_i \in I_n$. Thus $I_n + \prod_{i=1}^{n-1} I_i = I_n + \bigcap_{i=1}^{n-1} I_i = (1)$, and we can now apply the $n = 2$ case to conclude that $\bigcap_{i=1}^{n} I_i = \prod_{i=1}^{n} I_i$.

Note that for any i, we can construct a z_i with $z_i \in I_j$ for $j \neq i$ and $z_i + I_i = 1 + I_i$ via the same procedure.

Define $\phi : R \to \bigoplus R/I_i$ by $\phi(a) = (a + I_1, \ldots, a + I_n)$. The kernel of ϕ is $\bigcap_{i=1}^{n} I_i$, because $a + I_i = 0 + I_i$ iff $a \in I_i$, so that $\phi(a) = (0 + I_1, \ldots, 0 + I_n)$ iff $a \in I_i$ for all i, that is, $a \in \bigcap_{i=1}^{n} I_i$. Combined with our previous result, the kernel of ϕ is $\prod_{i=1}^{n} I_i$.

Finally, recall that we constructed $z_i \in R$ such that $z_i + I_i = 1 + I_i$, and $z + I_j = 0 + I_j$ for all $j \neq i$, so that $\phi(z_i) = (0 + I_1, \ldots, 1 + I_i, \ldots, 0 + I_n)$. Thus, $\phi(a_1 z_1 + \cdots + a_n z_n) = (a_1 + I_1, \ldots, a_n + I_n)$ for all $a_i \in R$, so that ϕ is onto. By the first isomorphism theorem, we have that $R/I_1 \cdots I_n \simeq \bigoplus_{i=1}^{n} R/I_i$.

2.7. Some special classes of domains

Principal ideal domains

2.7.1 Definition A ring R is a **principal ideal domain** or **PID** if $R \neq 0$, R is not a field, R is a domain, and every ideal of R is principal.

These have the next simplest theory of ideals. Each ideal is very simple—it's principal—though there might be a lot of ideals.

2.7.2 Example $\mathbb{Z}$ is a PID. The only nontrivial fact to check here is that:

2.7.3 Proposition *Any nonzero ideal $I \subset \mathbb{Z}$ is principal.*

Proof. If $I = (0)$, then this is obvious. Else there is $n \in I - \{0\}$; we can assume $n > 0$. Choose $n \in I$ as small as possible and positive. Then I claim that the ideal I is generated by (n). Indeed, we have $(n) \subset I$ obviously. If $m \in I$ is another integer, then divide m by n, to find $m = nb + r$ for $r \in [0, n)$. We find that $r \in I$ and $0 \leqslant r < n$, so $r = 0$, and m is divisible by n. And $I \subset (n)$.

So $I = (n)$. ∎

A module M is said to be *finitely generated* if there exist elements $x_1, \ldots, x_n \in M$ such that any element of M is a linear combination (with coefficients in R) of the x_i. (We shall define this more formally below.) One reason that PIDs are so convenient is:

2.7.4 Theorem (Structure theorem) *If M is a finitely generated module over a principal ideal domain R, then M is isomorphic to a direct sum*

$$M \simeq \bigoplus_{i=1}^{n} R/a_i,$$

for various $a_i \in R$ (possibly zero).

add: at some point, the proof should be added. This is important!

Unique factorization domains

The integers $\mathbb{Z}$ are especially nice because of the fundamental theorem of arithmetic, which states that every integer has a unique factorization into primes. This is not true for every integral domain.

2.7.5 Definition An element of a domain R is **irreducible** if it cannot be written as the product of two non-unit elements of R.

2.7.6 Example Consider the integral domain $\mathbb{Z}[\sqrt{-5}]$. We saw earlier that

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}),$$

which means that 6 was written as the product of two non-unit elements in different ways. $\mathbb{Z}[\sqrt{-5}]$ does not have unique factorization.

2.7.7 Definition A domain R is a **unique factorization domain** or **UFD** if every non-unit $x \in R$ satisfies

1. x can be written as a product $x = p_1 p_2 \cdots p_n$ of irreducible elements $p_i \in R$
2. if $x = q_1 q_2 \cdots q_m$ where $q_i \in R$ are irreducible then the p_i and q_i are the same up to order and multiplication by units.

2.7.8 Example $\mathbb{Z}$ is a UFD, while $\mathbb{Z}[\sqrt{-5}]$ is not. In fact, many of our favorite domains have unique factorization. We will prove that all PIDs are UFDs. In particular, in remark 2.7.13 and remark 2.7.14, we saw that $\mathbb{Z}[i]$ and $F[t]$ are PIDs, so they also have unique factorization.

2.7.9 Theorem *Every principal ideal domain is a unique factorization domain.*

Proof. Suppose that R is a principal ideal domain and x is an element of R. We first demonstrate that x can be factored into irreducibles. If x is a unit or an irreducible, then we are done. Therefore, we can assume that x is reducible, which means that $x = x_1 x_2$ for non-units $x_1, x_2 \in R$. If there are irreducible, then we are again done, so we assume that they are reducible and repeat this process. We need to show that this process terminates.

Suppose that this process continued infinitely. Then we have an infinite ascending chain of ideals, where all of the inclusions are proper: $(x) \subset (x_1) \subset (x_{11}) \subset \cdots \subset R$. We will show that this is impossible because any infinite ascending chain of ideals $I_1 \subset I_2 \subset \cdots \subset R$ of a principal ideal domain eventually becomes stationary, i.e. for some n, $I_k = I_n$ for $k \geqslant n$. Indeed, let $I = \bigcup_{i=1}^{\infty} I_i$. This is an ideal, so it is principally generated as $I = (a)$ for some a. Since $a \in I$, we must have $a \in I_N$ for some N, which means that the chain stabilizes after I_N.

It remains to prove that this factorization of x is unique. We induct on the number of irreducible factors n of x. If $n = 0$, then x is a unit, which has unique factorization up to units. Now, suppose that $x = p_1 \cdots p_n = q_1 \cdots q_m$ for some $m \geqslant n$. Since p_1 divides x, it must divide the product $q_1 \cdots q_m$ and by irreducibility, one of the factors q_i. Reorder the q_i so that p_1 divides q_1. However, q_1 is irreducible, so this means that p_1 and q_1 are the same up to multiplication by a unit u. Canceling p_1 from each of the two factorizations, we see that $p_2 \cdots p_n = uq_2 \cdots q_m = q_2' \cdots q_m$. By induction, this shows that the factorization of x is unique up to order and multiplication by units. ☐

Euclidean domains

A euclidean domain is a special type of principal ideal domain. In practice, it will often happen that one has an explicit proof that a given domain is euclidean, while it might not be so trivial to prove that it is a UFD without the general implication below.

2.7.10 Definition An integral domain R is a **euclidean domain** if there is a function $|\cdot| : R \to Z_{\geqslant 0}$ (called the norm) such that the following hold.

1. $|a| = 0$ iff $a = 0$.
2. For any nonzero $a, b \in R$ there exist $q, r \in R$ such that $b = aq + r$ and $|r| < |a|$.

In other words, the norm is compatible with division with remainder.

2.7.11 Theorem *A euclidean domain is a principal ideal domain.*

Proof. Let R be an euclidean domain, $I \subset R$ and ideal, and b be the nonzero element of smallest norm in I. Suppose $a \in I$. Then we can write $a = qb + r$ with $0 \leqslant r < |b|$, but since b has minimal nonzero absolute value, $r = 0$ and $b|a$. Thus $I = (b)$ is principal. ☐

As we will see, this implies that any euclidean domain admits *unique factorization.*

2.7.12 Proposition *Let F be a field. Then the polynomial ring $F[t]$ is a euclidean domain. In particular, it is a PID.*

Proof. We define *add:* ☐

2.7.13 Remark Prove that $\mathbb{Z}[i]$ is principal. (Define the norm as $N(a + ib) = a^2 + b^2$.)

2.7.14 Remark Prove that the polynomial ring $F[t]$ for F a field is principal.

It is *not* true that a PID is necessarily euclidean. Nevertheless, it was shown in **?** that the converse is "almost" true. Namely, **?** defines the notion of an **almost euclidean domain.** A domain R is almost euclidean if there is a function $d : R \to \mathbb{Z}_{\geqslant 0}$ such that

1. $d(a) = 0$ iff $a = 0$.
2. $d(ab) \geqslant d(a)$ if $b \neq 0$.
3. If $a, b \in R - \{0\}$, then either $b \mid a$ or there is $r \in (a, b)$ with $d(r) < d(b)$.

It is easy to see by the same argument that an almost euclidean domain is a PID. (Indeed, let R be an almost euclidean domain, and $I \subset R$ a nonzero ideal. Then choose $x \in I - \{0\}$ such that $d(x)$ is minimal among elements in I. Then if $y \in I - \{0\}$, either $x \mid y$ or $(x, y) \subset I$ contains an element with smaller d. The latter cannot happen, so the former does.) However, in fact:

2.7.15 Proposition (?) *A domain is a PID if and only if it is almost euclidean.*

Proof. Indeed, let R be a PID. Then R is a UFD (theorem 2.7.9), so for any $x \in R$, there is a factorization into prime elements, unique up to units. If x factors into n elements, we define $d(x) = n$; we set $d(0) = 0$. The first two conditions for an almost euclidean domain are then evident.

Let $x = p_1 \dots p_m$ and $y = q_1 \dots q_n$ be two elements of R, factored into irreducibles. Suppose $x \nmid y$. Choose a generator b of the (principal) ideal (x, y); then obviously $y \mid b$ so $d(y) \leqslant d(b)$. But if $d(y) = d(b)$, then the number of factors of y and b is the same, so $y \mid b$ would imply that y and b are associates. This is a contradiction, and implies that $d(y) < d(b)$.

2.7.16 Remark We have thus seen that a euclidean domain is a PID, and a PID is a UFD. Both converses, however, fail. By Gauß's lemma (**??**), the polynomial ring $\mathbb{Z}[X]$ has unique factorization, though the ideal $(2, X)$ is not principal.

In **?**, it is shown that the ring $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ is a PID but not euclidean (i.e. there is *no* euclidean norm on it).

According to **?**, sec. 8.3, proposition 2.7.15 actually goes back to Hasse (and these norms are sometimes called "Dedekind-Hasse norms").

2.8. Basic properties of modules

Free modules

We now describe a simple way of constructing modules over a ring, and an important class of modules.

2.8.1 Definition A module M is *free* if it is isomorphic to $R^{(S)} = \bigoplus_S R$ for some index set S. The cardinality of S is called the *rank* of the free module.

2.8.2 Example R is the simplest example of a free module.

Free modules have a *universal property*. Namely, recall that if M is an R-module, then to give a homomorphism

$$R \to M$$

is equivalent to giving an element $m \in M$ (the image of 1). By the universal product of the direct sum (which is the coproduct in the category of modules), it follows that to give a map

$$\bigoplus_I \to M$$

is the same as giving a map of *sets* $I \to M$. In particular:

2.8.3 Proposition *The functor $S \mapsto \bigoplus_S R$ from* Ens *to R-modules is the left adjoint to the forgetful functor U from R-modules to* Ens.

The claim now is that the notion of "rank" is well-defined for a free module. To see this, we will have to use the notion of a *maximal ideal* (definition 2.6.4) and corollary 2.6.13. Indeed, suppose $\bigoplus_I R$ and $\bigoplus_J R$ are isomorphic; we must show that I and J have the same cardinality. Choose a maximal ideal $\mathfrak{m} \subset R$. Then, by applying the functor $M \to M/\mathfrak{m}M$, we find that the $R/\mathfrak{m}$-*vector spaces*

$$\bigoplus_I R/\mathfrak{m}, \quad \bigoplus_J R/\mathfrak{m}$$

are isomorphic. By linear algebra, I and J have the same cardinality.

Free modules have a bunch of nice properties. The first is that it is very easy to map out of a free module.

2.8.4 Example Let I be an indexing set, and M an R-module. Then to give a morphism

$$\bigoplus_I R \to M$$

is equivalent to picking an element of M for each $i \in I$. Indeed, given such a collection of elements $\{m_i\}$, we send the generator of $\bigoplus_I R$ with a 1 in the ith spot and zero elsewhere to m_i.

2.8.5 Example In a domain, every principal ideal (other than zero) is a free module of rank one.

Another way of saying this is that the free module $\bigoplus_I R$ represents the functor on modules sending M to the *set* M^I. We have already seen a special case of this for I a one-element set (remark 2.5.13).

The next claim is that free modules form a reasonably large class of the category of R-modules.

2.8.6 Proposition *Given an R-module M, there is a free module F and a surjection*

$$F \twoheadrightarrow M.$$

Proof. We let F to be the free R-module on the elements e_m, one for each $m \in M$. We define the map

$$F \to M$$

by describing the image of each of the generators e_m: we just send each e_m to $m \in M$. It is clear that this map is surjective. ∎

We close by making a few remarks on matrices. Let M be a free module of rank n, and fix an isomorphism $M \simeq R^n$. Then we can do linear algebra with M, even though we are working over a ring and not necessarily a field, at least to some extent. For instance, we can talk about n-by-n matrices over the ring R, and then each of them induces a transformation, i.e. a module-homomorphism, $M \to M$; it is easy to see that every module-homomorphism between free modules is of this form. Moreover, multiplication of matrices corresponds to composition of homomorphisms, as usual.

2.8.7 Example Let us consider the question of when the transformation induced by an n-by-n matrix is invertible. The answer is similar to the familiar one from linear algebra in the case of a field. Namely, the condition is that the determinant be invertible.

Suppose that an $n \times n$ matrix A over a ring R is invertible. This means that there exists A^{-1} so that $AA^{-1} = I$, so hence $1 = \det I = \det(AA^{-1}) = (\det A)(\det A^{-1})$, and therefore, $\det A$ must be a unit in R.

Suppose instead that an $n \times n$ matrix A over a ring R has an invertible determinant. Then, using Cramer's rule, we can actually construct the inverse of A.

We next show that if R is a commutative ring, the category of modules over R contains enough information to reconstruct R. This is a small part of the story of *Morita equivalence,* which we shall not enter into here.

2.8.8 Example Suppose R is a commutative ring, and let $\mathcal{C}$ be the category of R-modules. The claim is that $\mathcal{C}$, as an *abstract* category, determines R. Indeed, the claim is that R is canonically the ring of endomorphisms of the identity functor $1_{\mathcal{C}}$.

Such an *endomorphism* is given by a natural transformation $\phi : 1_{\mathcal{C}} \to 1_{\mathcal{C}}$. In other words, one requires for each R-module M, a homomorphism of R-modules $\phi_M : M \to M$ such that if $f : M \to N$ is any homomorphism of modules, then there is a commutative square

$$\begin{tikzcd} M \arrow[r, "\phi_M"] \arrow[d, "f"] & M \arrow[d] \\ N \arrow[r, "\phi_N"] & N. \end{tikzcd}$$

Here is a simple way of obtaining such endomorphisms. Given $r \in R$, we consider the map $r : M \to m$ which just multiplies each element by r. This is a homomorphism, and it is clear that it is natural in the above sense. There is thus a map $R \to \mathrm{End}(1_{\mathcal{C}})$ (note that multiplication corresponds to composition of natural transformations). This map is clearly injective; different $r, s \in R$ lead to different natural transformations (e.g. on the R-module R).

The claim is that *any* natural transformation of $1_{\mathcal{C}}$ is obtained in this way. Namely, let $\phi : 1_{\mathcal{C}} \to 1_{\mathcal{C}}$ be such a natural transformation. On the R-module R, ϕ must be multiplication by some element $r \in R$ (because $\hom_R(R, R)$ is given by such homotheties). Consequently, one sees by drawing commutative diagrams that $\phi : R^{\oplus S} \to R^{\oplus S}$ is of this form for any set S. So ϕ is multiplication by r on any free R-module. Since any module M is a quotient of a free module F, we can draw a diagram

$$\begin{tikzcd} F \arrow[r, "\phi_F"] \arrow[d] & F \arrow[d] \\ M \arrow[r, "\phi_M"] & M. \end{tikzcd}$$

Since the vertical arrows are surjective, we find that ϕ_F must be given by multiplication by r too.

Finitely generated modules

The notion of a "finitely generated" module is analogous to that of a finite-dimensional vector space.

2.8.9 Definition An R-module M is **finitely generated** if there exists a surjection $R^n \to M$ for some n. In other words, it has a finite number of elements whose "span" contains M.

The basic properties of finitely generated modules follow from the fact that they are stable under extensions and quotients.

2.8.10 Proposition *Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence. If M', M'' are finitely generated, so is M.*

Proof. Suppose $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is exact. Then g is surjective, f is injective, and $\ker(g) = \mathrm{im}(f)$. Now suppose M' is finitely generated, say by $\{a_1, \dots, a_s\}$, and M'' is finitely generated, say by $\{b_1, \dots, b_t\}$. Because g is surjective, each $g^{-1}(b_i)$ is non-empty. Thus, we can fix some $c_i \in g^{-1}(b_i)$ for each i.

For any $m \in M$, we have $g(m) = r_1 b_1 + \cdots + r_t b_t$ for some $r_i \in R$ because $g(m) \in M''$ and M'' is generated by the b_i. Thus $g(m) = r_1 g(c_i) + \cdots r_t g(c_t) = g(r_1 c_1 + \cdots + r_t c_t)$, and because g is a homomorphism we have $m - (r_1 c_1 + \cdots + r_t c_t) \in \ker(g) = \mathrm{im}(f)$. But M' is generated by the a_i, so the submodule $\mathrm{im}(f) \subset M$ is finitely generated by the $d_i = f(a_i)$.

Thus, any $m \in M$ has $m - (r_1 c_1 + \cdots + r_t c_t) = r_{t+1} d_1 + \cdots + r_{t+s} d_s$ for some $r_1, \dots, r_{t+s}$, thus M is finitely generated by $c_1, \dots, c_t, d_1, \dots, d_s$.

The converse is false. It is possible for finitely generated modules to have submodules which are *not* finitely generated. As we shall see in chapter III.2, this does not happen over *noetherian* rings.

2.8.11 Example Consider the ring $R = \mathbb{C}[X_1, X_2, \dots,]$ and the ideal $(X_1, X_2, \dots)$. This ideal is a submodule of the finitely generated R-module R, but it is not finitely generated.

2.8.12 Remark Show that a quotient of a finitely generated module is finitely generated.

2.8.13 Remark Consider a *split* exact sequence $0 \to M' \to M \to M'' \to 0$. In this case, show that if M is finitely generated, so is M'.

Finitely presented modules

Over messy rings, the notion of a finitely presented module is often a good substitute for that of a finitely generated one. In fact, we are going to see (**??**) that there is a general method of reducing questions about finitely presented modules over arbitrary rings to finitely generated modules over finitely generated $\mathbb{Z}$-algebras.

Throughout, fix a ring R.

2.8.14 Definition An R-module M is **finitely presented** if there is an exact sequence

$$R^m \to R^n \to M \to 0.$$

The point of this definition is that M is the quotient of a free module R^n by the "relations" given by the images of the vectors in R^m. Since R^m is finitely generated, M can be represented via finitely many generators *and* finitely many relations.

The reader should compare this with the definition of a **finitely generated** module; there we only require an exact sequence

$$R^n \to M \to 0.$$

As usual, we establish the usual properties of finitely presented modules.

We start by showing that if a finitely presented module M is generated by finitely many elements, the "module of relations" among these generators is finitely generated itself. The condition of finite presentation only states that there is *one* such set of generators such that the module of generators is finitely generated.

2.8.15 Proposition *Suppose M is finitely presented. Then if $R^m \twoheadrightarrow M$ is a surjection, the kernel is finitely generated.*

Proof. Let K be the kernel of $R^m \twoheadrightarrow M$. Consider an exact sequence

$$F' \to F \to M \to 0$$

where F', F are finitely generated and free, which we can do as M is finitely presented. Draw a commutative and exact diagram

$$\begin{tikzcd}
& F' \arrow[r] & F \arrow[r] \arrow[d, dashed] & M \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K \arrow[r] & R^m \arrow[r] & M \arrow[r] & 0
\end{tikzcd}$$

The dotted arrow $F \to R^m$ exists as F is projective. There is induced a map $F' \to K$. We get a commutative and exact diagram

$$\begin{tikzcd}
& F' \arrow[r] \arrow[d, "f"] & F \arrow[r] \arrow[d, "g"] & M \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K \arrow[r] & R^m \arrow[r] & M \arrow[r] & 0
\end{tikzcd},$$

to which we can apply the snake lemma. There is an exact sequence

$$0 \to \mathrm{coker}(f) \to \mathrm{coker}(g) \to 0,$$

which gives an isomorphism $\mathrm{coker}(f) \simeq \mathrm{coker}(g)$. However, $\mathrm{coker}(g)$ is finitely generated, as a quotient of R^m. Thus $\mathrm{coker}(f)$ is too. Since we have an exact sequence

$$0 \to \mathrm{Im}(f) \to K \to \mathrm{coker}(f) \to 0,$$

and $\mathrm{Im}(f)$ is finitely generated (as the image of a finitely generated object, F'), we find by proposition 2.8.10 that $\mathrm{coker}(f)$ is finitely generated. ☐

2.8.16 Proposition *Given an exact sequence*

$$0 \to M' \to M \to M'' \to 0,$$

if M', M'' are finitely presented, so is M.

In general, it is not true that if M is finitely presented, then M' and M'' are. For instance, it is possible that a submodule of the free, finitely generated module R (i.e. an ideal), might fail to be finitely generated. We shall see in chapter III.2 that this does not happen over a *noetherian* ring.

Proof. Indeed, suppose we have exact sequences

$$F_1' \to F_0' \to M' \to 0$$

and

$$F_1'' \to F_0'' \to M'' \to 0$$

where the F's are finitely generated and free. We need to get a similar sequence for M. Let us stack these into a diagram

$$\begin{tikzcd}
 & F_1' \arrow[d] & & F_1'' \arrow[d] & \\
 & F_0' \arrow[d] & & F_0'' \arrow[d] & \\
0 \arrow[r] & M' \arrow[r] & M \arrow[r] & M'' \arrow[r] & 0
\end{tikzcd}$$

However, now, using general facts about projective modules (**??**), we can splice these presentations into a resolution

$$F_1' \oplus F_1'' \to F_0' \oplus F_0'' \to M \to 0,$$

which proves the assertion. ☐

2.8.17 Corollary *The (finite) direct sum of finitely presented modules is finitely presented.*

Proof. Immediate from proposition 2.8.16 ☐

Modules of finite length

A much stronger condition on modules that of finite generation is that of *finite length*. Here, basically any operation one does will eventually terminate.

Let R be a commutative ring, M an R-module.

2.8.18 Definition M is **simple** if $M \neq 0$ and M has no nontrivial submodules.

2.8.19 Remark A torsion-free abelian group is never a simple $\mathbb{Z}$-module.

2.8.20 Proposition *M is simple if and only if it is isomorphic to $R/\mathfrak{m}$ for $\mathfrak{m} \subset R$ a maximal ideal.*

Proof. Let M be simple. Then M must contain a cyclic submodule Rx generated by some $x \in M - \{0\}$. So it must contain a submodule isomorphic to R/I for some ideal I, and simplicity implies that $M \simeq R/I$ for some I. If I is not maximal, say properly contained in J, then we will get a nontrivial submodule J/I of $R/I \simeq M$. Conversely, it is easy to see that $R/\mathfrak{m}$ is simple for $\mathfrak{m}$ maximal. □

2.8.21 Remark (Schur's lemma) Let $f : M \to N$ be a module-homomorphism, where M, N are both simple. Then either $f = 0$ or f is an isomorphism.

2.8.22 Definition M is of **finite length** if there is a finite filtration $0 \subset M^0 \subset \cdots \subset M^n = M$ where each M^i/M^{i-1} is simple.

2.8.23 Remark Modules of finite length are closed under extensions (that is, if $0 \to M' \to M \to M'' \to 0$ is an exact sequence, then if M', M'' are of finite length, so is M).

In the next result (which will not be used in this chapter), we shall use the notions of a *noetherian* and an *artinian* module. These notions will be developed at length in **??**, and we refer the reader there for more explanation. A module is *noetherian* if every ascending chain $M_1 \subset M_2 \subset \dots$ of submodules stabilizes, and it is *artinian* if every descending chain stabilizes.

2.8.24 Proposition *M is finite length iff M is both noetherian and artinian.*

Proof. Any simple module is obviously both noetherian and artinian: there are two submodules. So if M is finite length, then the finite filtration with simple quotients implies that M is noetherian and artinian, since these two properties are stable under extensions (proposition 2.1.7 and proposition 2.4.5 of chapter III.2).

Suppose $M \neq 0$ is noetherian and artinian. Let $M_1 \subset M$ be a minimal nonzero submodule, which exists as M is artinian. This is necessarily simple. Then we have a filtration

$$0 = M_0 \subset M_1.$$

If $M_1 = M$, then the filtration goes up to M, and we have that M is of finite length. If not, find a submodule M_2 that contains M_1 and is minimal among submodules containing M_1; then the quotient M_2/M_1 is simple. We have the filtration

$$0 = M_0 \subset M_1 \subset M_2,$$

which we can keep continuing until at some point we reach M. Note that since M is noetherian, we cannot continue this strictly ascending chain forever. □

2.8.25 Remark In particular, any submodule or quotient module of a finite length module is of finite length. Note that the analog is not true for finitely generated modules unless the ring in question is noetherian.

Our next goal is to show that the length of a filtration of a module with simple quotients is well-defined. For this, we need:

2.8.26 Lemma *Let $0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$ be a filtration of M with simple quotients. Let $N \subset M$. Then the filtration $0 = M_0 \cap N \subset M_1 \cap N \subset \cdots \subset N$ has simple or zero quotients.*

Proof. Indeed, for each i, $(N \cap M_i)/(N \cap M_{i-1})$ is a submodule of M_i/M_{i-1}, so is either zero or simple. □

2.8.27 Theorem (Jordan-Hölder) *Let M be a module of finite length. In this case, any two filtrations on M with simple quotients have the same length.*

2.8.28 Definition This number is called the **length** of M and is denoted $\ell(M)$.

Proof of theorem 2.8.27. Let us introduce a temporary definition: $l(M)$ is the length of the *minimal* filtration on M. We will show that any filtration of M (with simple quotients) is of length $l(M)$. This is the proposition in another form.

The proof of this claim is by induction on $l(M)$. Suppose we have a filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$$

with simple quotients. We would like to show that $n = l(M)$. By definition of $l(M)$, there is another filtration

$$0 = N_0 \subset \cdots \subset N_{l(M)} = M.$$

If $l(M) = 0, 1$, then M is zero or simple, which will necessarily imply that $n = 0, 1$ respectively. So we can assume $l(M) \geqslant 2$. We can also assume that the result is known for strictly smaller submodules of M.

There are two cases:

1. $M_{n-1} = N_{l(M)-1}$. Then $M_{n-1} = N_{l(M)-1}$ has l at most $l(M) - 1$. Thus by the inductive hypothesis any two filtrations on M_{n-1} have the same length, so $n-1 = l(M)-1$, implying what we want.

2. We have $M_{n-1} \cap N_{l(M)-1} \subsetneq M_{n-1}, N_{l(M)-1}$. Call this intersection K.

   Now we have two filtrations of these modules $M_{n-1}, N_{l(M)-1}$ whose quotients are simple. We can replace them such that the next term before them is K. To do this, consider the filtrations

   $$0 = M_0 \cap K \subset M_1 \subset K \subset \ldots M_{n-1} \cap K = K \subset M_{n-1}$$

   and

   $$0 = N_0 \cap K \subset M_1 \subset K \subset \ldots N_{l(M)-1} \cap K = K \subset N_{l(M)-1}.$$

   These filtrations have simple or zero quotients by lemma 2.8.26, and since $M_{n-1}/K = M_{n-1}/M_{n-1} \cap N_{l(M)-1} = M/M_{n-1}$ is simple, and similarly for $N_{l(M)-1}/K$. We can throw out redundancies to eliminate the zero terms. So we get two new filtrations of M_{n-1} and $N_{l(M)-1}$ whose second-to-last term is K.

By the inductive hypothesis any two filtrations on either of these proper submodules $M_{n-1}, N_{l(M)-1}$ have the same length. Thus the lengths of the two new filtrations are $n-1$ and $l(M)-1$, respectively. So we find that $n-1 = l(K)+1$ and $l(M)-1 = l(K)+1$ by the inductive hypothesis. This implies what we want. ∎

2.8.29 Remark Prove that the successive quotients M_i/M_{i-1} are also determined (up to permutation).

I.3. Fields and extensions

3.1. Introduction

In this chapter, we shall discuss the theory of fields. Recall that a field is an integral domain for which all non-zero elements are invertible; equivalently, the only two ideals of a field are (0) and (1) since any nonzero element is a unit. Consequently fields will be the simplest cases of much of the theory developed later.

The theory of field extensions has a different feel from standard commutative algebra since, for instance, any morphism of fields is injective. Nonetheless, it turns out that questions involving rings can often be reduced to questions about fields. For instance, any integral domain can be embedded in a field (its quotient field), and any local ring (that is, a ring with a unique maximal ideal; we have not defined this term yet) has associated to it its residue field (that is, its quotient by the maximal ideal). A knowledge of field extensions will thus be useful.

3.2. Fields

Recall once again:

3.2.1 Definition A *field* is an integral domain where every non-zero element is invertible. Alternatively, it is a set $\mathbb{K}$, endowed with binary operations of addition $+$ and multiplication $\cdot$ and two elements 0 and 1, such that the usual axioms of a field are satisfied:

(Fld1) $\mathbb{K}$ together with addition $+$ and the element 0 is an abelian group.

(Fld2) $\mathbb{K}$ together with multiplication $\cdot$ and the element 1 is an abelian monoid such that every non-zero element of $\mathbb{K}$ is invertible.

(Fld3) Multiplication distributes from the left and the right over addition that is

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{and} \quad (a + b) \cdot c = a \cdot c + a \cdot b \qquad \text{for all } a, b, c \in \mathbb{K} .$$

(Fld4) The neutral elements 0 and 1 are not equal.

A *subfield* of a field is a subset closed under these operations and containing 0 and 1. Equivalently, it is a subring that is itself a field.

For a field $\mathbb{K}$, we write $\mathbb{K}^*$ for the subset $\mathbb{K}\setminus\{0\}$. This generalizes the usual notation **??**, where R^* refers to the group of invertible elements in a ring R.

Examples

To get started, let us begin by providing several examples of fields. The reader should recall (corollary 2.6.13) that if R is a ring and $I \subset R$ an ideal, then R/I is a field precisely when I is maximal.

3.2.2 Example One of the most familiar examples of a field is the rational numbers $\mathbb{Q}$.

3.2.3 Example If p is a prime number, then $\mathbb{Z}/(p)$ is a field, denoted $\mathbb{F}_p$. Indeed, (p) is a maximal ideal in $\mathbb{Z}$. Thus, fields may be finite: $\mathbb{F}_p$ contains p elements.

3.2.4 Example (Quotients of the polynomial ring) In a principal ideal domain, every prime ideal is principal. Now, by 2.7.12, if k is a field, then the polynomial ring $k[x]$ is a PID. It follows that if $P \in k[x]$ is an irreducible polynomial (that is, a nonconstant polynomial that does not admit a factorization into terms of smaller degrees), then $k[x]/(P)$ is a field. It contains a copy of k in a natural way.

This is a very general way of constructing fields. For instance, the complex numbers $\mathbb{C}$ can be constructed as $\mathbb{R}[x]/(x^2 + 1)$.

3.2.5 Remark What is $\mathbb{C}[x]/(x^2 + 1)$?

3.2.6 Example (Quotient fields) Recall from remark 2.6.18 that, given an integral domain A, there is an imbedding $A \hookrightarrow K(A)$ into a field $K(A)$ formally constructed as quotients $a/b, a, b \in A$ (and $b \neq 0$) modulo an evident equivalence relation. This is called the **quotient field.** The quotient field has the following universal property: given an injection $\phi : A \hookrightarrow K$ for a field K, there is a unique map $\psi : K(A) \to K$ making the diagram commutative (i.e. a map of A-algebras). Indeed, it is clear how to define such a map: we set

$$\psi(a/b) = \phi(a)/\phi(b),$$

where injectivity of ϕ assures that $\phi(b) \neq 0$ if $b \neq 0$.

If the map is not injective, then such a factorization may not exist. Consider the imbedding $\mathbb{Z} \to \mathbb{Q}$ into its quotient field, and consider the map $\mathbb{Z} \to \mathbb{F}_p$: this last map goes from $\mathbb{Z}$ into a field, but it does not factor through $\mathbb{Q}$ (as p is invertible in $\mathbb{Q}$ and zero in $\mathbb{F}_p$!).

3.2.7 Example (Rational function field) If k is a field, then we can consider the field $k(x)$ of **rational functions** over k. This is the quotient field of the polynomial ring $k[x]$; in other words, it is the set of quotients F/G for $F, G \in k[x]$ with the obvious equivalence relation.

Here is a fancier example of a field.

3.2.8 Example Let X be a Riemann surface.[1] Let $\mathbb{C}(X)$ denote the set of meromorphic functions on X; clearly $\mathbb{C}(X)$ is a ring under multiplication and addition of functions. It turns out that in fact $\mathbb{C}(X)$ is a field; this is because if a nonzero function $f(z)$ is meromorphic, so is $1/f(z)$. For example, let S^2 be the Riemann sphere; then we know from complex analysis that the ring of meromorphic functions $\mathbb{C}(S^2)$ is the field of rational functions $\mathbb{C}(z)$.

[1]Readers not familiar with Riemann surfaces may ignore this example.

One reason fields are so nice from the point of view of most other chapters in this book is that the theory of k-modules (i.e. vector spaces), for k a field, is very simple. Namely:

3.2.9 Proposition *If k is a field, then every k-module is free.*

Proof. Indeed, by linear algebra we know that a k-module (i.e. vector space) V has a *basis* $\mathcal{B} \subset V$, which defines an isomorphism from the free vector space on $\mathcal{B}$ to V. □

3.2.10 Corollary *Every exact sequence of modules over a field splits.*

Proof. This follows from **??** and proposition 3.2.9, as every vector space is projective. □

This is another reason why much of the theory in future chapters will not say very much about fields, since modules behave in such a simple manner. Note that corollary 3.2.10 is a statement about the *category* of k-modules (for k a field), because the notion of exactness is inherently arrow-theoretic (i.e. makes use of purely categorical notions, and can in fact be phrased within a so-called *abelian category*).

Henceforth, since the study of modules over a field is linear algebra, and since the ideal theory of fields is not very interesting, we shall study what this chapter is really about: *extensions* of fields.

The characteristic of a field

In the category of rings, there is an *initial object* $\mathbb{Z}$: any ring R has a map from $\mathbb{Z}$ into it in precisely one way. For fields, there is no such initial object. Nonetheless, there is a family of objects such that every field can be mapped into in exactly one way by exactly one of them, and in no way by the others.

Let F be a field. As $\mathbb{Z}$ is the initial object of the category of rings, there is a ring map $f : \mathbb{Z} \to F$, see 2.1.17. The image of this ring map is an integral domain (as a subring of a field) hence the kernel of f is a prime ideal in $\mathbb{Z}$, see 2.6.15. Hence the kernel of f is either (0) or (p) for some prime number p, see 2.6.2.

In the first case we see that f is injective, and in this case we think of $\mathbb{Z}$ as a subring of F. Moreover, since every nonzero element of F is invertible we see that it makes sense to talk about $p/q \in F$ for $p, q \in \mathbb{Z}$ with $q \neq 0$. Hence in this case we may and we do think of $\mathbb{Q}$ as a subring of F. One can easily see that this is the smallest subfield of F in this case.

In the second case, i.e., when $\mathrm{Ker}(f) = (p)$ we see that $\mathbb{Z}/(p) = \mathbb{F}_p$ is a subring of F. Clearly it is the smallest subfield of F.

Arguing in this way we see that every field contains a smallest subfield which is either $\mathbb{Q}$ or finite equal to $\mathbb{F}_p$ for some prime number p.

3.2.11 Definition The **characteristic** of a field F is 0 if $\mathbb{Z} \subset F$, or is a prime p if $p = 0$ in F. The **prime subfield of** F is the smallest subfield of F which is either $\mathbb{Q} \subset F$ if the characteristic is zero, or $\mathbb{F}_p \subset F$ if the characteristic is $p > 0$.

It is easy to see that if E is a field containing k, then the characteristic of E is the same as the characteristic of k.

3.2.12 Example The characteristic of $\mathbb{Z}/p$ is p, and that of $\mathbb{Q}$ is 0. This is obvious from the definitions.

3.3. Field extensions

Preliminaries

In general, though, we are interested not so much in fields by themselves but in field *extensions.* This is perhaps analogous to studying not rings but *algebras* over a fixed ring. The nice thing for fields is that the notion of a "field over another field" just recovers the notion of a field extension, by the next result.

3.3.1 Proposition *If F is a field and R is any ring, then any ring homomorphism $f : F \to R$ is either injective or the zero map (in which case $R = 0$).*

Proof. Indeed, $\ker(f)$ is an ideal in F. But there are only two ideals in F, namely (0) and (1). If f is identically zero, then $1 = f(1) = 0$ in R, so $R = 0$ too. ☐

3.3.2 Definition If F is a field contained in a field G, then G is said to be a **field extension** of F. We shall write G/F to indicate that G is an extension of F.

So if F, F' are fields, and $F \to F'$ is any ring-homomorphism, we see by proposition 3.3.1 that it is injective,[2] and F' can be regarded as an extension of F, by a slight abuse of notation. Alternatively, a field extension of F is just an F-algebra that happens to be a field. This is completely different than the situation for general rings, since a ring homomorphism is not necessarily injective.

Let k be a field. There is a *category* of field extensions of k. An object of this category is an extension E/k, that is a (necessarily injective) morphism of fields

$$k \to E,$$

while a morphism between extensions $E/k, E'/k$ is a k-algebra morphism $E \to E'$; alternatively, it is a commutative diagram

$$\begin{tikzcd} E \arrow[rr] & & E' \\ & k \arrow[ul] \arrow[ur] & \end{tikzcd}.$$

3.3.3 Definition A **tower** of field extensions $E'/E/k$ consists of an extension E/k and an extension E'/E.

[2]The zero ring is not a field!

It is easy to see that any morphism $E \to E'$ in the category of k-extensions gives a tower.

Let us give a few examples of field extensions.

3.3.4 Example Let k be a field, and $P \in k[x]$ an irreducible polynomial. We have seen that $k[x]/(P)$ is a field (3.2.7). Since it is also a k-algebra in the obvious way, it is an extension of k.

3.3.5 Example If X is a Riemann surface, then the field of meromorphic functions $\mathbb{C}(X)$ (see example 3.2.8) is an extension field of $\mathbb{C}$, because any element of $\mathbb{C}$ induces a meromorphic—indeed, holomorphic—constant function on X.

Let F/k be a field extension. Let $S \subset F$ be any subset. Then there is a *smallest* subextension of F (that is, a subfield of F containing k) that contains S. To see this, consider the family of subfields of F containing S and k, and take their intersection; one easily checks that this is a field. It is easy to see, in fact, that this is the set of elements of F that can be obtained via a finite number of elementary algebraic operations (addition, multiplication, subtraction, and division) involving elements of k and S.

3.3.6 Definition If F/k is an extension and $S \subset F$, we write $k(S)$ for the smallest subextension of F containing S. We will say that S **generates** the extension $k(S)/k$.

For instance, $\mathbb{C}$ is generated by i over $\mathbb{R}$.

3.3.7 Remark Show that $\mathbb{C}$ does not have a countable set of generators over $\mathbb{Q}$.

Let us now classify extensions generated by one element.

3.3.8 Proposition (Simple extensions of a field) *If an extension F/k is generated by one element, then it is F is k-isomorphic either to the rational function field $k(t)/k$ or to one of the extensions $k[t]/(P)$ for $P \in k[t]$ irreducible.*

We will see that many of the most important cases of field extensions are generated by one element, so this is actually useful.

Proof. Let $\alpha \in F$ be such that $F = k(\alpha)$; by assumption, such an α exists. There is a morphism of rings

$$k[t] \to F$$

sending the indeterminate t to α. The image is a domain, so the kernel is a prime ideal. Thus, it is either (0) or (P) for $P \in k[t]$ irreducible.

If the kernel is (P) for $P \in k[t]$ irreducible, then the map factors through $k[t]/(P)$, and induces a morphism of fields $k[t]/(P) \to F$. Since the image contains α, we see easily that the map is surjective, hence an isomorphism. In this case, $k[t]/(P) \simeq F$.

If the kernel is trivial, then we have an injection $k[t] \to F$. One may thus define a morphism of the quotient field $k(t)$ into F; given a quotient $R(t)/Q(t)$ with $R(t), Q(t) \in k[t]$, we map this to $R(\alpha)/Q(\alpha)$. The hypothesis that $k[t] \to F$ is injective implies that $Q(\alpha) \neq 0$ unless Q is the zero polynomial. The quotient field of $k[t]$ is the rational function field $k(t)$, so we get a morphism $k(t) \to F$ whose image contains α. It is thus surjective, hence an isomorphism. ∎

Finite extensions

If F/E is a field extension, then evidently F is also a vector space over E (the scalar action is just multiplication in F).

3.3.9 Definition The dimension of F considered as an E-vector space is called the **degree** of the extension and is denoted $[F : E]$. If $[F : E] < \infty$ then F is said to be a **finite** extension.

3.3.10 Example $\mathbb{C}$ is obviously a finite extension of $\mathbb{R}$ (of degree 2).

Let us now consider the degree in the most important special example, that given by proposition 3.3.8, in the next two examples.

3.3.11 Example (Degree of a simple transcendental extension) If k is any field, then the rational function field $k(t)$ is *not* a finite extension. The elements $\{t^n, n \in \mathbb{Z}\}$ are linearly independent over k.

In fact, if k is uncountable, then $k(t)$ is *uncountably* dimensional as a k-vector space. To show this, we claim that the family of elements $\{1/(t-\alpha), \alpha \in k\} \subset k(t)$ is linearly independent over k. A nontrivial relation between them would lead to a contradiction: for instance, if one works over $\mathbb{C}$, then this follows because $\frac{1}{t-\alpha}$, when considered as a meromorphic function on $\mathbb{C}$, has a pole at α and nowhere else. Consequently any sum $\sum c_i \frac{1}{t-\alpha_i}$ for the $c_i \in k^*$, and $\alpha_i \in k$ distinct, would have poles at each of the α_i. In particular, it could not be zero.

(Amusingly, this leads to a quick if suboptimal proof of the Hilbert Nullstellensatz; see **??**.)

3.3.12 Example (Degree of a simple algebraic extension) Consider a monogenic field extension E/k of the form in 3.2.7, say $E = k[t]/(P)$ for $P \in k[t]$ an irreducible polynomial. Then the degree $[E : k]$ is just the degree $\deg P$. Indeed, without loss of generality, we can assume P monic, say

$$P = t^n + a_1 t^{n-1} + \cdots + a_0. \tag{3.3.1}$$

It is then easy to see that the images of $1, t, \dots, t^{n-1}$ in $k[t]/(P)$ are linearly independent over k, because any relation involving them would have degree strictly smaller than that of P, and P is the element of smallest degree in the ideal (P).

Conversely, the set $S = \{1, t, \dots, t^{n-1}\}$ (or more properly their images) spans $k[t]/(P)$ as a vector space. Indeed, we have by (3.3.1) that t^n lies in the span of S. Similarly, the relation $tP(t) = 0$ shows that the image of t^{n+1} lies in the span of $\{1, t, \dots, t^n\}$—by what was just shown, thus in the span of S. Working upward inductively, we find that the image of t^M for $M \geqslant n$ lies in the span of S.

This confirms the observation that $[\mathbb{C} : \mathbb{R}] = 2$, for instance. More generally, if k is a field, and $\alpha \in k$ is not a square, then the irreducible polynomial $x^2 - \alpha \in k[x]$ allows one to construct an extension $k[x]/(x^2 - \alpha)$ of degree two. We shall write this as $k(\sqrt{\alpha})$. Such extensions will be called **quadratic,** for obvious reasons.

The basic fact about the degree is that it is *multiplicative in towers.*

3.3.13 Proposition (Multiplicativity) *Suppose given a tower $F/E/k$. Then*

$$[F : k] = [F : E][E : k].$$

Proof. Let $\alpha_1, \ldots, \alpha_n \in F$ be an E-basis for F. Let $\beta_1, \ldots, \beta_m \in E$ be a k-basis for E. Then the claim is that the set of products $\{\alpha_i\beta_j, 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$ is a k-basis for F. Indeed, let us check first that they span F over k.

By assumption, the $\{\alpha_i\}$ span F over E. So if $f \in F$, there are $a_i \in E$ with

$$f = \sum a_i \alpha_i,$$

and, for each i, we can write $a_i = \sum b_{ij}\beta_j$ for some $b_{ij} \in k$. Putting these together, we find

$$f = \sum_{i,j} b_{ij}\alpha_i\beta_j,$$

proving that the $\{\alpha_i\beta_j\}$ span F over k.

Suppose now that there existed a nontrivial relation

$$\sum_{i,j} c_{ij}\alpha_i\beta_j = 0$$

for the $c_{ij} \in k$. In that case, we would have

$$\sum_i \alpha_i \left(\sum_j c_{ij}\beta_j \right) = 0,$$

and the inner terms lie in E as the β_j do. Now E-linear independence of the $\{\alpha_i\}$ shows that the inner sums are all zero. Then k-linear independence of the $\{\beta_j\}$ shows that the c_{ij} all vanish. □

We sidetrack to a slightly tangential definition:

3.3.14 Definition A field extensions K of $\mathbb{Q}$ is said to be a **number field** if it is a finite extension of $\mathbb{Q}$.

Number fields are the basic objects in algebraic number theory. We shall see later that, for the analog of the integers $\mathbb{Z}$ in a number field, something kind of like unique factorization still holds (though strict unique factorization generally does not!).

Algebraic extensions

Consider a field extension F/E.

3.3.15 Definition An element $\alpha \in F$ is said to be **algebraic** over E if α is the root of some polynomial with coefficients in E. If all elements of F are **algebraic** then F is said to be an algebraic extension.

By proposition 3.3.8, the subextension $E(\alpha)$ is isomorphic either to the rational function field $E(t)$ or to a quotient ring $E[t]/(P)$ for $P \in E[t]$ an irreducible polynomial. In the latter case, α is algebraic over E (in fact, it satisfies the polynomial P!); in the former case, it is not.

3.3.16 Example $\mathbb{C}$ is algebraic over $\mathbb{R}$.

3.3.17 Example Let X be a compact Riemann surface, and $f \in \mathbb{C}(X) - \mathbb{C}$ any nonconstant meromorphic function on X (see example 3.2.8). Then it is known that $\mathbb{C}(X)$ is algebraic over the subextension $\mathbb{C}(f)$ generated by f. We shall not prove this.

We now show that there is a deep connection between finiteness and being algebraic.

3.3.18 Proposition *A finite extension is algebraic. In fact, an extension E/k is algebraic if and only if every subextension $k(\alpha)/k$ generated by some $\alpha \in E$ is finite.*

In general, it is very false that an algebraic extension is finite.

Proof. Let E/k be finite, say of degree n. Choose $\alpha \in E$. Then the elements $\{1, \alpha, \dots, \alpha^n\}$ are linearly dependent over E, or we would necessarily have $[E : k] > n$. A relation of linear dependence now gives the desired polynomial that α must satisfy.

For the last assertion, note that a monogenic extension $k(\alpha)/k$ is finite if and only α is algebraic over k, by example 3.3.11 and example 3.3.12. So if E/k is algebraic, then each $k(\alpha)/k, \alpha \in E$, is a finite extension, and conversely. ☐

We can extract a corollary of the last proof (really of example 3.3.11 and example 3.3.12): a monogenic extension is finite if and only if it is algebraic. We shall use this observation in the next result.

3.3.19 Corollary *Let k be a field, and let $\alpha_1, \alpha_2, \dots, \alpha_n$ be elements of some extension field such that each α_i is finite over k. Then the extension $k(\alpha_1, \dots, \alpha_n)/k$ is finite. That is, a finitely generated algebraic extension is finite.*

Proof. Indeed, each $k(\alpha_1, \dots, \alpha_{i+1})/k(\alpha_1, \dots, \alpha_i)$ is monogenic and algebraic, hence finite. ☐

The set of complex numbers that are algebraic over $\mathbb{Q}$ are simply called the **algebraic numbers.** For instance, $\sqrt{2}$ is algebraic, i is algebraic, but π is not. It is a basic fact that the algebraic numbers form a field, although it is not obvious how to prove this from the definition that a number is algebraic precisely when it satisfies a nonzero polynomial equation with rational coefficients (e.g. by polynomial equations).

3.3.20 Corollary *Let E/k be a field extension. Then the elements of E algebraic over k form a field.*

Proof. Let $\alpha, \beta \in E$ be algebraic over k. Then $k(\alpha, \beta)/k$ is a finite extension by corollary 3.3.19. It follows that $k(\alpha + \beta) \subset k(\alpha, \beta)$ is a finite extension, which implies that $\alpha + \beta$ is algebraic by proposition 3.3.18. ☐

Many nice properties of field extensions, like those of rings, will have the property that they will be preserved by towers and composita.

3.3.21 Proposition (Towers) *Let E/k and F/E be algebraic. Then F/k is algebraic.*

Proof. Choose $\alpha \in F$. Then α is algebraic over E. The key observation is that α is algebraic over a *finitely generated* subextension of k. That is, there is a finite set $S \subset E$ such that α is algebraic over $k(S)$: this is clear because being algebraic means that a certain polynomial in $E[x]$ that α satisfies exists, and as S we can take the coefficients of this polynomial.

It follows that α is algebraic over $k(S)$. In particular, $k(S, \alpha)/k(S)$ is finite. Since S is a finite set, and $k(S)/k$ is algebraic, corollary 3.3.19 shows that $k(S)/k$ is finite. Together we find that $k(S, \alpha)/k$ is finite, so α is algebraic over k. ∎

The method of proof in the previous argument—that being algebraic over E was a property that *descended* to a finitely generated subextension of E—is an idea that recurs throughout algebra, and will be put to use more generality in **??**.

Minimal polynomials

Let E/k be a field extension, and let $\alpha \in E$ be algebraic over k. Then α satisfies a (nontrivial) polynomial equation in $k[x]$. Consider the set of polynomials $P(x) \in k[x]$ such that $P(\alpha) = 0$; by hypothesis, this set does not just contain the zero polynomial. It is easy to see that this set is an *ideal.* Indeed, it is the kernel of the map

$$k[x] \to E, \quad x \mapsto \alpha.$$

Since $k[x]$ is a PID, there is a *generator* $m(x) \in k[x]$ of this ideal. If we assume m monic, without loss of generality, then m is uniquely determined.

3.3.22 Definition $m(x)$ as above is called the **minimal polynomial** of α over k.

The minimal polynomial has the following characterization: it is the monic polynomial, of smallest degree, that annihilates α. (Any nonconstant multiple of $m(x)$ will have larger degree, and only multiples of $m(x)$ can annihilate α.) This explains the name *minimal.*

Clearly the minimal polynomial is *irreducible.* This is equivalent to the assertion that the ideal in $k[x]$ consisting of polynomials annihilating α is prime. But this follows from the fact that the map $k[x] \to E, x \mapsto \alpha$ is a map into a domain (even a field), so the kernel is a prime ideal.

3.3.23 Proposition *The degree of the minimal polynomial is* $[k(\alpha) : k]$.

Proof. This is just a restatement of the argument in **??**: the observation is that if $m(x)$ is the minimal polynomial of α, then the map

$$k[x]/(m(x)) \to k(\alpha), \quad x \mapsto \alpha$$

is an isomorphism as in the aforementioned proof, and we have counted the degree of such an extension (see example 3.3.12). ∎

So the observation of the above proof is that if $\alpha \in E$ is algebraic, then $k(\alpha) \subset E$ is isomorphic to $k[x]/(m(x))$.

Algebraic closure

Now we want to define a "universal" algebraic extension of a field. Actually, we should be careful: the algebraic closure is *not* a universal object. That is, the algebraic closure is not unique up to *unique* isomorphism: it is only unique up to isomorphism. But still, it will be very handy, if not functorial.

3.3.24 Definition Let F be a field. An **algebraic closure** of F is a field $\overline{F}$ containing F such that:

(AC1) $\overline{F}$ is algebraic over F.

(AC2) $\overline{F}$ is **algebraically closed** (that is, every non-constant polynomial in $\overline{F}[X]$ has a root in $\overline{F}$).

The "fundamental theorem of algebra" states that $\mathbb{C}$ is algebraically closed. While the easiest proof of this result uses Liouville's theorem in complex analysis, we shall give a mostly algebraic proof below (**??**).

We now prove the basic existence result.

3.3.25 Theorem *Every field has an algebraic closure.*

The proof will mostly be a red herring to the rest of the chapter. However, we will want to know that it is *possible* to embed a field inside an algebraically closed field, and we will often assume it done.

Proof. Let K be a field and Σ be the set of all monic irreducibles in $K[x]$. Let $A = K[\{x_f : f \in \Sigma\}]$ be the polynomial ring generated by indeterminates x_f, one for each $f \in \Sigma$. Then let $\mathfrak{a}$ be the ideal of A generated by polynomials of the form $f(x_f)$ for each $f \in \Sigma$.

Claim 1. $\mathfrak{a}$ is a proper ideal.

Proof of claim 1. Suppose $\mathfrak{a} = (1)$, so there exist finitely many polynomials $f_i \in \Sigma$ and $g_i \in A$ such that $1 = f_1(x_{f_1})g_1 + \cdots + f_k(x_{f_k})g_k$. Each g_i uses some finite collection of indeterminates $V_i\{x_{f_{i_1}}, \ldots, x_{f_{i_{k_i}}}\}$. This notation is ridiculous, so we simplify it.

We can take the union of all the V_i, together with the indeterminates $x_{f_1}, \ldots, x_{f_k}$ to get a larger but still finite set of indeterminates $V = \{x_{f_1}, \ldots, x_{f_n}\}$ for some $n \geqslant k$ (ordered so that the original $x_{f_1}, \ldots, x_{f_k}$ agree the first k elements of V). Now we can regard each g_i as a polynomial in this new set of indeterminates V. Then, we can write $1 = f_1(x_{f_1})g_1 + \cdots + f_n(x_{f_n})g_n$ where for each $i > k$, we let $g_i = 0$ (so that we've adjoined a few zeroes to the right hand side of the equality). Finally, we define $x_i = x_{f_i}$, so that we have $1 = f_1(x_1)g_1(x_1, \ldots, x_n) + \cdots + f_n(x_n)g_n(x_1, \ldots, x_n)$.

Suppose n is the minimal integer such that there exists an expression of this form, so that

$$\mathfrak{b} = (f_1(x_1), \ldots, f_{n-1}(x_{n-1}))$$

is a proper ideal of $B = K[x_1, \ldots, x_{n-1}]$, but

$$(f_1(x_1), \ldots, f_n(x_n))$$

is the unit ideal in $B[x_n]$. Let $\hat{B} = B/\mathfrak{b}$ (observe that this ring is nonzero). We have a composition of maps

$$B[x_n] \to \hat{B}[x_n] \to \hat{B}[x_n]/(\widehat{f_n(x_n)})$$

where the first map is reduction of coefficients modulo $\mathfrak{b}$, and the second map is the quotient by the principal ideal generated by the image $\widehat{f_n(x_n)}$ of $f_n(x_n)$ in $\hat{B}[x_n]$. We know $\hat{B}$ is a nonzero ring, so since f_n is monic, the top coefficient of $\widehat{f_n(x_n)}$ is still $1 \in \hat{B}$. In particular, the top coefficient cannot be nilpotent. Furthermore, since f_n was irreducible, it is not a constant polynomial, so by the characterization of units in polynomial rings, $\widehat{f_n(x_n)}$ is not a unit, so it does not generate the unit ideal. Thus the quotient $\hat{B}[x_n]/(\widehat{f_n(x_n)})$ should not be the zero ring.

On the other hand, observe that each $f_i(x_i)$ is in the kernel of this composition, so in fact the entire ideal $(f_1(x_1), \ldots, f_n(x_n))$ is contained in the kernel. But this ideal is the unit ideal, so all of $B[x_n]$ is in the kernel of this composition. In particular, $1 \in B[x_n]$ is in the kernel, and since ring maps preserve identity, this forces $1 = 0$ in $\hat{B}[x_n]/(\widehat{f_n(x_n)})$, which makes this the the zero ring. This contradicts our previous observation, and proves the claim that $\mathfrak{a}$ is a proper ideal.

Now, given claim 1, there exists a maximal ideal $\mathfrak{m}$ of A containing $\mathfrak{a}$. Let $K_1 = A/\mathfrak{m}$. This is an extension field of K via the inclusion given by

$$K \to A \to A/\mathfrak{m}$$

(this map is automatically injective as it is a map between fields). Furthermore every $f \in \Sigma$ has a root in K_1. Specifically, the coset $x_f + \mathfrak{m}$ in $A/\mathfrak{m} = K_1$ is a root of f since

$$f(x_f + \mathfrak{m}) = f(x_f) + \mathfrak{m} = 0.$$

Inductively, given K_n for some $n \geqslant 1$, repeat the construction with K_n in place of K to get an extension field K_{n+1} of K_n in which every irreducible $f \in K_n[x]$ has a root. Let $L = \bigcup_{n=1}^{\infty} K_n$.

Claim 2. Every $f \in L[x]$ splits completely into linear factors in L.

Proof of claim 2. We induct on the degree of f. In the base case, when f itself is linear, there is nothing to prove. Inductively, suppose every polynomial in $L[x]$ of degree less than n splits completely into linear factors, and suppose

$$f = a_0 + a_1 x + \cdots + a_n x^n \in L[x]$$

has degree n. Then each $a_i \in K_{n_i}$ for some n_i, so let $n = \max n_i$ and regard f as a polynomial in $K_n[x]$. If f is reducible in $K_n[x]$, then we have a factorization $f = gh$ with the degree of g, h strictly less than n. Therefore, inductively, they both split into linear factors in $L[x]$, so f must also. On the other hand, if f is irreducible, then by our construction, it has a root $a \in K_{n+1}$, so we have $f = (x - a)g$ for some $g \in K_{n+1}[x]$ of degree $n - 1$. Again inductively, we can split g into linear factors in L, so clearly we can do the same with f also. This completes the proof of claim 2.

Let $\bar{K}$ be the set of algebraic elements in L. Clearly $\bar{K}$ is an algebraic extension of K. If $f \in \bar{K}[x]$, then we have a factorization of f in $L[x]$ into linear factors

$$f = b(x - a_1)(x - a_2)\cdots(x - a_n). \qquad \square$$

for $b \in \bar{K}$ and, a priori, $a_i \in L$. But each a_i is a root of f, which means it is algebraic over $\bar{K}$, which is an algebraic extension of K; so by transitivity of "being algebraic," each a_i is algebraic over K. So in fact we conclude that $a_i \in \bar{K}$ already, since $\bar{K}$ consisted of all elements algebraic over K. Therefore, since $\bar{K}$ is an algebraic extension of K such that every $f \in \bar{K}[x]$ splits into linear factors in $\bar{K}$, $\bar{K}$ is the algebraic closure of K.

add: two algebraic closures are isomorphic

Let K be an algebraically closed field. Then the ring $K[x]$ has a very simple ideal structure. Since every polynomial $P \in K[x]$ has a root, it follows that there is always a decomposition (by dividing repeatedly)

$$P = c(x - \alpha_1)\dots(x - \alpha_n),$$

where c is the constant term and the $\{\alpha_i\} \subset k$ are the roots of P. In particular:

3.3.26 Proposition *For K algebraically closed, the only irreducible polynomials in $K[x]$ are the linear polynomials $c(x - \alpha)$, $c, \alpha \in K$ (and $c \neq 0$).*

In particular, two polynomials in $K[x]$ are **relatively prime** (i.e., generate the unit ideal) if and only if they have no common roots. This follows because the maximal ideals of $K[x]$ are of the form $(x - \alpha), \alpha \in K$. So if $F, G \in K[x]$ have no common root, then (F, G) cannot be contained in any $(x - \alpha)$ (as then they would have a common root at α).

If k is *not* algebraically closed, then this still gives information about when two polynomials in $k[x]$ generate the unit ideal.

3.3.27 Definition If k is any field, we say that two polynomials in $k[x]$ are **relatively prime** if they generate the unit ideal in $k[x]$.

3.3.28 Proposition *Two polynomials in $k[x]$ are relatively prime precisely when they have no common roots in an algebraic closure $\overline{k}$ of k.*

Proof. The claim is that any two polynomials P, Q generate (1) in $k[x]$ if and only if they generate (1) in $\overline{k}[x]$. This is a piece of linear algebra: a system of linear equations with coefficients in k has a solution if and only if it has a solution in any extension of k. Consequently, we can reduce to the case of an algebraically closed field, in which case the result is clear from what we have already proved. $\square$

3.4. Separability and normality

Separable extensions

Throughout, $F \subset K$ is a finite field extension. We fix once and for all an algebraic closure $\overline{F}$ for F and an embedding of F in M.

3.4.1 Definition For an element $\alpha \in K$ with minimal polynomial $q \in F[x]$, we say q and α are **separable** if q has distinct roots (in some algebraic closure $\overline{F}$!), and we say K is separable if this holds for all $\alpha \in K$.

By proposition 3.3.28, separability of a polynomial $P \in F[x]$ is equivalent to $(P, P') = 1$ in $F[x]$. Indeed, this follows from the fact that P has no multiple roots if and only if P, P' have no common roots.

3.4.2 Lemma *$q(x) \in F[x]$ is separable if and only if $\gcd(q, q') = 1$, where q' is the formal derivative of q.*

Purely inseparable extensions

3.4.3 Definition For an element $\alpha \in K$ with minimal polynomial q, we say α is **purely inseparable** if q has only one root. We say K is splitting if each q splits in K.

3.4.4 Definition If $K = F(\alpha)$ for some α with minimal polynomial $q(x) \in F[x]$, then by 3.5.3, $q(x) = r(x^{p^d})$, where $p = \operatorname{char} F$ (or 1 if $\operatorname{char} F = 0$) and r is separable; in this case we also denote $\deg_s(K/F) = \deg(r), \deg_i(K/F) = p^d$.

3.5. Galois theory

Definitions

Throughout, $F \subset K$ is a finite field extension. We fix once and for all an algebraic closure M for both and an embedding of F in M. When necessary, we write $K = F(\alpha_1, \ldots, \alpha_n)$, and $K_0 = F, K_i = F(\alpha_1, \ldots, \alpha_i)$, q_i the minimal polynomial of α_i over F_{i-1}, Q_i that over F.

3.5.1 Definition $\mathsf{Aut}(K/F)$ denotes the group of automorphisms of K which fix F (pointwise!). $\operatorname{Emb}(K/F)$ denotes the set of embeddings of K into M respecting the chosen embedding of F.

3.5.2 Definition By $\deg(K/F)$ we mean the dimension of K as an F-vector space. We denote K_s/F the set of elements of K whose minimal polynomials over F have distinct roots; by 3.5.13 this is a subfield, and $\deg(K_s/F) = \deg_s(K/F)$ and $\deg(K/K_s) = \deg_i(K/F)$ by definition.

Theorems

3.5.3 Lemma *If $\operatorname{char} F = 0$ then $K_s = K$. If $\operatorname{char} F = p > 0$, then for any irreducible $q(x) \in K[x]$, there is some $d \geqslant 0$ and polynomial $r(x) \in K[x]$ such that $q(x) = r(x^{p^d})$, and r is separable and irreducible.*

Proof. By formal differentiation, $q'(x)$ has positive degree unless each exponent is a multiple of p; in characteristic zero this never occurs. If this is not the case, since q is irreducible, it can have no factor in common with q' and therefore has distinct roots by 3.4.2.

If $p > 0$, let d be the largest integer such that each exponent of q is a multiple of p^d, and define r by the above equation. Then by construction, r has at least one exponent which is not a multiple of p, and therefore has distinct roots. □

3.5.4 Corollary *In the statement of 3.5.3, q and r have the same number of roots.*

Proof. α is a root of q if and only if α^{p^d} is a root of r; i.e. the roots of q are the roots of $x^{p^d} - \beta$, where β is a root of r. But if α is one such root, then $(x-\alpha)^{p^d} = x^{p^d} - \alpha^{p^d} = x^{p^d} - \beta$ since $\operatorname{char} K = p$, and therefore α is the only root of $x^{p^d} - \beta$. □

3.5.5 Lemma *The correspondence which to each $g \in \mathrm{Emb}(K/F)$ assigns the n-tuple $(g(\alpha_1), \ldots, g(\alpha_n))$ of elements of M is a bijection from $\mathrm{Emb}(K/F)$ to the set of tuples of $\beta_i \in M$, such that β_i is a root of q_i over $K(\beta_1, \ldots, \beta_{i-1})$.*

Proof. First take $K = F(\alpha) = F[x]/(q)$, in which case the maps $g\colon K \to M$ over F are identified with the elements $\beta \in M$ such that $q(\beta) = 0$ (where $g(\alpha) = \beta$).

Now, considering the tower $K = K_n/K_{n-1}/\ldots/K_0 = F$, each extension of which is primitive, and a given embedding g, we define recursively $g_1 \in \mathrm{Emb}(K_1/F)$ by restriction and subsequent g_i by identifying K_{i-1} with its image and restricting g to K_i. By the above paragraph each g_i corresponds to the image $\beta_i = g_i(\alpha_i)$, each of which is a root of q_i. Conversely, given such a set of roots of the q_i, we define g recursively by this formula. □

3.5.6 Corollary $|\mathrm{Emb}(K/F)| = \prod_{i=1}^{n} \deg_s(q_i)$.

Proof. This follows immediately by induction from 3.5.5 by 3.5.4. □

3.5.7 Lemma *For any $f \in \mathrm{Emb}(K/F)$, the map $\mathsf{Aut}(K/F) \to \mathrm{Emb}(K/F)$ given by $\sigma \mapsto f \circ \sigma$ is injective.*

Proof. This is immediate from the injectivity of f. □

3.5.8 Corollary $\mathsf{Aut}(K/F)$ *is finite.*

Proof. By 3.5.7, $\mathsf{Aut}(K/F)$ injects into $\mathrm{Emb}(K/F)$, which by 3.5.6 is finite. □

3.5.9 Proposition *The inequality*

$$|\mathsf{Aut}(K/F)| \leqslant |\mathrm{Emb}(K/F)|$$

is an equality if and only if the q_i all split in K.

Proof. The inequality follows from 3.5.7 and from 3.5.8. Since both sets are finite, equality holds if and only if the injection of 3.5.7 is surjective (for fixed $f \in \mathrm{Emb}(K/F)$).

If surjectivity holds, let $\beta_1, \ldots, \beta_n$ be arbitrary roots of $q_1, \ldots, q_n$ in the sense of 3.5.5, and extract an embedding $g\colon K \to M$ with $g(\alpha_i) = \beta_i$. Since the correspondence $f \mapsto f \circ \sigma$ ($\sigma \in \mathsf{Aut}(K/F)$) is a bijection, there is some σ such that $g = f \circ \sigma$, and therefore f and g have the same image. Therefore the image of K in M is canonical, and contains $\beta_1, \ldots, \beta_n$ for any choice thereof.

If the q_i all split, let $g \in \mathrm{Emb}(K/F)$ be arbitrary, so the $g(\alpha_i)$ are roots of q_i in M as in 3.5.5. But the q_i have all their roots in K, hence in the image $f(K)$, so f and g again have the same image, and $f^{-1} \circ g \in \mathsf{Aut}(K/F)$. Thus $g = f \circ (f^{-1} \circ g)$ shows that the map of 3.5.7 is surjective. □

3.5.10 Corollary *Define*

$$D(K/F) = \prod_{i=1}^{n} \deg_s(K_i/K_{i-1}).$$

Then the chain of equalities and inequalities

$$|\mathsf{Aut}(K/F)| \leqslant |\operatorname{Emb}(K/F)| = D(K/F) \leqslant \deg(K/F)$$

holds; the first inequality is an equality if and only if each q_i splits in K, and the second if and only if each q_i is separable.

Proof. The statements concerning the first inequality are just 3.5.9; the interior equality is just 3.5.6; the latter inequality is obvious from the multiplicativity of the degrees of field extensions; and the deduction for equality follows from the definition of $\deg_s$. □

3.5.11 Corollary *The q_i respectively split and are separable in K if and only if the Q_i do and are.*

Proof. The ordering of the α_i is irrelevant, so we may take each $i = 1$ in turn. Then $Q_1 = q_1$ and if either of the equalities in 3.5.10 holds then so does the corresponding statement here. Conversely, clearly each q_i divides Q_i, so splitting or separability for the latter implies that for the former. □

3.5.12 Corollary *Let $\alpha \in K$ have minimal polynomial q; if the Q_i are respectively split, separable, and purely inseparable over F then q is as well.*

Proof. We may take α as the first element of an alternative generating set for K/F. The numerical statement of 3.5.10 does not depend on the particular generating set, hence the conditions given hold of the set containing α if and only if they hold of the canonical set $\alpha_1, \ldots, \alpha_n$.

For purely inseparable, if the Q_i all have only one root then $|\operatorname{Emb}(K/F)| = 1$ by 3.5.10, and taking α as the first element of a generating set as above shows that q must have only one root as well for this to hold. □

3.5.13 Corollary *K_s is a field and* $\deg(K_s/F) = D(K/F)$.

Proof. Assume $\operatorname{char} F = p > 0$, for otherwise $K_s = K$. Using 3.5.3, write each $Q_i = R_i(x^{p^{d_i}})$, and let $\beta_i = \alpha_i^{p^{d_i}}$. Then the β_i have R_i as minimal polynomials and the α_i satisfy $s_i = x^{p^{d_i}} - \beta_i$ over $K' = F(\beta_1, \ldots, \beta_n)$. Therefore the α_i have minimal polynomials over K' dividing the s_i and hence those polynomials have but one distinct root.

By 3.5.12, the elements of K' are separable, and those of K' purely inseparable over K'. In particular, since these minimal polynomials divide those over F, none of these elements is separable, so $K' = K_s$.

The numerical statement follows by computation:

$$\deg(K/K') = \prod_{i=1}^{n} p^{d_i} = \prod_{i=1}^{n} \frac{\deg(K_i/K_{i-1})}{\deg_s(K_i/K_{i-1})} = \frac{\deg(K/F)}{D(K/F)}. \qquad \square$$

3.5.14 Theorem *The following inequality holds:*

$$|\mathsf{Aut}(K/F)| \leqslant |\operatorname{Emb}(K/F)| = \deg_s(K/F) \leqslant \deg(K/F).$$

Equality holds on the left if and only if K/F is splitting; it holds on the right if and only if K/F is separable.

Proof. The numerical statement combines 3.5.10 and 3.5.13. The deductions combine 3.5.11 and 3.5.12. $\square$

Definitions

Throughout, we will denote as before K/F a finite field extension, and $G = \mathsf{Aut}(K/F)$, H a subgroup of G. L/F is a subextension of K/F.

3.5.15 Definition When K/F is separable and splitting, we say it is Galois and write $G = \operatorname{Gal}(K/F)$, the Galois group of K over F.

3.5.16 Definition The fixed field of H is the field K^H of elements fixed by the action of H on K. Conversely, G_L is the fixing subgroup of L, the subgroup of G whose elements fix L.

Theorems

3.5.17 Lemma *A polynomial $q(x) \in K[x]$ which splits in K lies in $K^H[x]$ if and only if its roots are permuted by the action of H. In this case, the sets of roots of the irreducible factors of q over K^H are the orbits of the action of H on the roots of q (counting multiplicity).*

Proof. Since H acts by automorphisms, we have $\sigma q(x) = q(\sigma x)$ as a functional equation on K, so σ permutes the roots of q. Conversely, since the coefficients of σ are the elementary symmetric polynomials in its roots, H permuting the roots implies that it fixes the coefficients.

Clearly q is the product of the polynomials q_i whose roots are the orbits of the action of H on the roots of q, counting multiplicities, so it suffices to show that these polynomials are defined over K^H and are irreducible. Since H acts on the roots of the q_i by construction, the former is satisfied. If some q_i factored over K^H, its factors would admit an action of H on their roots by the previous paragraph. The roots of q_i are distinct by construction, so its factors do not share roots; hence the action on the roots of q_i would not be transitive, a contradiction. $\square$

3.5.18 Corollary *Let $q(x) \in K[x]$; if it is irreducible, then H acts transitively on its roots; conversely, if q is separable and H acts transitively on its roots, then $q(x) \in K^H[x]$ is irreducible.*

Proof. Immediate from 3.5.17. ☐

3.5.19 Lemma *If K/F is Galois, so is K/L, and* $\operatorname{Gal}(K/L) = G_L$..

Proof. K/F Galois means that the minimal polynomial over F of every element of K is separable and splits in K; the minimal polynomials over $L = K^H$ divide those over F, and therefore this is true of K/L as well; hence K/L is likewise a Galois extension. $\operatorname{Gal}(K/L) = \mathsf{Aut}(K/L)$ consists of those automorphisms σ of K which fix L; since $F \subset L$ we have *a fortiori* that σ fixes F, hence $\operatorname{Gal}(K/L) \subset G$ and consists of the subgroup which fixes L; i.e. G_L. ☐

3.5.20 Corollary *If K/F and L/F are Galois, then the action of G on elements of L defines a surjection of G onto* $\operatorname{Gal}(L/F)$. *Thus G_L is normal in G and* $\operatorname{Gal}(L/F) \cong G/G_L$. *Conversely, if $N \subset G$ is normal, then K^N/F is Galois.*

Proof. L/F is splitting, so by 3.5.17 the elements of G act as endomorphisms (hence automorphisms) of L/F, and the kernel of this action is G_L. By 3.5.19, we have $G_L = \operatorname{Gal}(K/L)$, so $|G_L| = |\operatorname{Gal}(K/L)| = [K : L] = [K : F]/[L : F]$, or rearranging and using that K/F is Galois, we get $|G|/|G_L| = [L : F] = |\operatorname{Gal}(L/F)|$. Thus the map $G \to \operatorname{Gal}(L/F)$ is surjective and thus the induced map $G/G_L \to \operatorname{Gal}(L/F)$ is an isomorphism.

Conversely, let N be normal and take $\alpha \in K^N$. For any conjugate β of α, we have $\beta = g(\alpha)$ for some $g \in G$; let $n \in N$. Then $n(\beta) = (ng)(\alpha) = g(g^{-1}ng)(\alpha) = g(\alpha) = \beta$, since $g^{-1}ng \in N$ by normality of N. Thus $\beta \in K^N$, so K^N is splitting, i.e., Galois. ☐

3.5.21 Proposition *If K/F is Galois and $H = G_L$, then $K^H = L$.*

Proof. By 3.5.19, K/L and K/K^H are both Galois. By definition, $\operatorname{Gal}(K/L) = G_L = H$; since H fixes K^H we certainly have $H < \operatorname{Gal}(K/K^H)$, but since $L \subset K^H$ we have *a fortiori* that $\operatorname{Gal}(K/K^H) < \operatorname{Gal}(K/L) = H$, so $\operatorname{Gal}(K/K^H) = H$ as well. It follows from 3.5.14 that $\deg(K/L) = |H| = \deg(K/K^H)$, so that $K^H = L$. ☐

3.5.22 Lemma *If K is a finite field, then K^* is cyclic.*

Proof. K is then a finite extension of $\mathbb{F}_p$ for $p = \operatorname{char} K$, hence has order p^n, $n = \deg(K/\mathbb{F}_p)$. Thus $\alpha^{p^n} = \alpha$ for all $\alpha \in K$, since $|K^*| = p^n - 1$. It follows that every element of K is a root of $q_n(x) = x^{p^n} - x$. For any $d < n$, the elements of order at most $p^d - 1$ satisfy $q_d(x)$, which has p^d roots. It follows that there are at least $p^n(p - 1) > 0$ elements of order exactly $p^n - 1$, so K^* is cyclic. ☐

3.5.23 Corollary *If K is a finite field, then* $\operatorname{Gal}(K/F)$ *is cyclic, generated by the Frobenius automorphism.*

Proof. First take $F = \mathbb{F}_p$. Then the map $f_i(\alpha) = \alpha^{p^i}$ is an endomorphism, injective since K is a field, and surjective since it is finite, hence an automorphism. Since every α satisfies $\alpha^{p^n} = \alpha$,

$f_n = 1$, but by 3.5.22, f_{n-1} is nontrivial (applied to the generator). Since $n = \deg(K/F)$, $f = f_1$ generates $\mathrm{Gal}(K/F)$.

If F is now arbitrary, by 3.5.21 we have $\mathrm{Gal}(K/F) = \mathrm{Gal}(K/\mathbb{F}_p)_F$, and every subgroup of a cyclic group is cyclic. □

3.5.24 Corollary *If K is finite, K/F is primitive.*

Proof. No element of G fixes the generator α of K^*, so it cannot lie in any proper subfield. Therefore $F(\alpha) = K$. □

3.5.25 Proposition *If F is infinite and K/F has only finitely many subextensions, then it is primitive.*

Proof. We proceed by induction on the number of generators of K/F.

If $K = F(\alpha)$ we are done. If not, $K = F(\alpha_1, \ldots, \alpha_n) = F(\alpha_1, \ldots, \alpha_{n-1})(\alpha_n) = F(\beta, \alpha_n)$ by induction, so we may assume $n = 2$. There are infinitely many subfields $F(\alpha_1 + t\alpha_2)$, with $t \in F$, hence two of them are equal, say for t_1 and t_2. Thus, $\alpha_1 + t_2\alpha_2 \in F(\alpha_1 + t_1\alpha_2)$. Then $(t_2 - t_1)\alpha_2 \in F(\alpha_1 + t_1\alpha_2)$, hence α_2 lies in this field, hence α_1 does. Therefore $K = F(\alpha_1 + t_1\alpha_2)$. □

3.5.26 Corollary *If K/F is separable, it is primitive, and the generator may be taken to be a linear combination of any finite set of generators of K/F.*

Proof. We may embed K/F in a Galois extension M/F by adjoining all the conjugates of its generators. Subextensions of K/F are as well subextensions of K'/F and by 3.5.21 the map $H \mapsto (K')^H$ is a surjection from the subgroups of G to the subextensions of K'/F, which are hence finite in number. By 3.5.24 we may assume F is infinite. The result now follows from 3.5.25. □

3.5.27 Corollary *If K/F is Galois and $H \subset G$, then if $L = K^H$, we have $H = G_L$.*

Proof. Let α be a primitive element for K/L. The polynomial $\prod_{h \in H}(x - h(\alpha))$ is fixed by H, and therefore has coefficients in L, so α has $|H|$ conjugate roots over L. But since α is primitive, we have $K = L(\alpha)$, so the minimal polynomial of α has degree $\deg(K/L)$, which is the same as the number of its roots. Thus $|H| = \deg(K/L)$. Since $H \subset G_L$ and $|G_L| = \deg(K/L)$, we have equality. □

3.5.28 Theorem *The correspondences $H \mapsto K^H$, $L \mapsto G_L$ define inclusion-reversing inverse maps between the set of subgroups of G and the set of subextensions of K/F, such that normal subgroups and Galois subfields correspond.*

Proof. This combines 3.5.21, 3.5.27, and 3.5.20. □

3.6. Transcendental Extensions

There is a distinguished type of transcendental extension: those that are "purely transcendental."

3.6.1 Definition A field extension E'/E is purely transcendental if it is obtained by adjoining a set B of algebraically independent elements. A set of elements is algebraically independent over E if there is no nonzero polynomial$P$ with coefficients in E such that $P(b_1, b_2, \cdots b_n) = 0$ for any finite subset of elements $b_1, \ldots, b_n \in B$.

3.6.2 Example The field $\mathbb{Q}(\pi)$ is purely transcendental; in particular, $\mathbb{Q}(\pi) \cong \mathbb{Q}(x)$ with the isomorphism fixing $\mathbb{Q}$.

Similar to the degree of an algebraic extension, there is a way of keeping track of the number of algebraically independent generators that are required to generate a purely transcendental extension.

3.6.3 Definition Let E'/E be a purely transcendental extension generated by some set of algebraically independent elements B. Then the transcendence degree $trdeg(E'/E) = \#(B)$ and B is called a transcendence basis for E'/E (we will see later that $trdeg(E'/E)$ is independent of choice of basis).

In general, let F/E be a field extension, we can always construct an intermediate extension $F/E'/E$ such that F/E' is algebraic and E'/E is purely transcendental. Then if B is a transcendence basis for E', it is also called a transcendence basis for F. Similarly, $trdeg(F/E)$ is defined to be $trdeg(E'/E)$.

3.6.4 Theorem *Let F/E be a field extension, a transcendence basis exists.*

Proof. Let A be an algebraically independent subset of F. Now pick a subset $G \subset F$ that generates F/E, we can find a transcendence basis B such that $A \subset B \subset G$. Define a collection of algebraically independent sets $\mathcal{B}$ whose members are subsets of G that contain A. The set can be partially ordered inclusion and contains at least one element, A. The union of elements of $\mathcal{B}$ is algebraically independent since any algebraic dependence relation would have occurred in one of the elements of $\mathcal{B}$ since the polynomial is only allowed to be over finitely many variables. The union also satisfies $A \subset \bigcup \mathcal{B} \subset G$ so by Zorn's lemma, there is a maximal element $B \in \mathcal{B}$. Now we claim F is algebraic over $E(B)$. This is because if it wasn't then there would be a transcendental element $f \in G$ (since $E(G) = F$)such that $B \cup \{f\}$ wold be algebraically independent contradicting the maximality of B. Thus B is our transcendence basis. □

Now we prove that the transcendence degree of a field extension is independent of choice of basis.

3.6.5 Theorem *Let F/E be a field extension. Any two transcendence bases for F/E have the same cardinality. This shows that the $trdeg(E/F)$ is well defined.*

Proof. Let B and B' be two transcendence bases. Without loss of generality, we can assume that $\#(B') \leqslant \#(B)$. Now we divide the proof into two cases: the first case is that B is an infinite set. Then for each $\alpha \in B'$, there is a finite set B_α such that α is algebraic over $E(B_\alpha)$ since any algebraic dependence relation only uses finitely many indeterminates. Then we define $B^* = \bigcup_{\alpha \in B'} B_\alpha$. By construction, $B^* \subset B$, but we claim that in fact the two sets are equal. To see this, suppose that they are not equal, say there is an element $\beta \in B \backslash B^*$. We know β

is algebraic over $E(B')$ which is algebraic over $E(B^*)$. Therefor β is algebraic over $E(B^*)$, a contradiction. So $\#(B) \leqslant \sum_{\alpha \in B'} \#(B_\alpha)$. Now if B' is finite, then so is B so we can assume B' is infinite; this means

$$\#(B) \leqslant \sum_{\alpha \in B'} \#(B_\alpha) = \#(\coprod B_\alpha) \leqslant \#(B' \times \mathbb{Z}) = \#(B') \tag{3.6.1}$$

with the inequality $\#(\coprod B_\alpha) \leqslant \#(B' \times \mathbb{Z})$ given by the correspondence $b_{\alpha_i} \mapsto (\alpha, i) \in B' \times \mathbb{Z}$ with $B_\alpha = \{b_{\alpha_1}, b_{\alpha_2} \cdots b_{\alpha_{n_\alpha}}\}$ Therefore in the infinite case, $\#(B) = \#(B')$.

Now we need to look at the case where B is finite. In this case, B' is also finite, so suppose $B = \{\alpha_1, \cdots \alpha_n\}$ and $B' = \{\beta_1, \cdots \beta_m\}$ with $m \leqslant n$. We perform induction on m: if $m = 0$ then F/E is algebraic so $B =$ so $n = 0$, otherwise there is an irreducible polynomial $f \in E[x, y_1, \cdots y_n]$ such that $f(\beta_1, \alpha_1, \cdots \alpha_n) = 0$. Since β_1 is not algebraic over E, f must involve some y_i so without loss of generality, assume f uses y_1. Let $B^* = \{\beta_1, \alpha_2, \cdots \alpha_n\}$. We claim that B^* is a basis for F/E. To prove this claim, we see that we have a tower of algebraic extensions $F/E(B^*, \alpha_1)/E(B^*)$ since α_1 is algebraic over $E(B^*)$. Now we claim that B^* (counting multiplicity of elements) is algebraically independent over E because if it weren't, then there would be an irreducible $g \in E[x, y_2, \cdots y_n]$ such that $g(\beta_1, \alpha_2, \cdots \alpha_n) = 0$ which must involve x making β_1 algebraic over $E(\alpha_2, \cdots \alpha_n)$ which would make α_1 algebraic over $E(\alpha_2, \cdots \alpha_n)$ which is impossible. So this means that $\{\alpha_2, \cdots \alpha_n\}$ and $\{\beta_2, \cdots \beta_m\}$ are bases for F over $E(\beta_1)$ which means by induction, $m = n$. □

3.6.6 Example Consider the field extension $\mathbb{Q}(e, \pi)$ formed by adjoining the numbers e and π. This field extension has transcendence degree at least 1 since both e and π are transcendental over the rationals. However, this field extension might have transcendence degree 2 if e and π are algebraically independent. Whether or not this is true is unknown and the problem of determining $trdeg(\mathbb{Q}(e, \pi))$ is an open problem.

3.6.7 Example let E be a field and $F = E(t)/E$. Then $\{t\}$ is a transcendence basis since $F = E(t)$. However, $\{t^2\}$ is also a transcendence basis since $E(t)/E(t^2)$ is algebraic. This illustrates that while we can always decompose an extension F/E into an algebraic extension F/E' and a purely transcendental extension E'/E, this decomposition is not unique and depends on choice of transcendence basis.

3.6.8 Remark If we have a tower of fields $G/F/E$, then $trdeg(G/E) = trdeg(F/E) + trdeg(G/F)$.

3.6.9 Example Let X be a compact Riemann surface. Then the function field $\mathbb{C}(X)$ (see example 3.2.8) has transcendence degree one over $\mathbb{C}$. In fact, *any* finitely generated extension of $\mathbb{C}$ of transcendence degree one arises from a Riemann surface. There is even an equivalence of categories between the category of compact Riemann surfaces and (non-constant) holomorphic maps and the opposite category of finitely generated extensions of $\mathbb{C}$ and morphisms of $\mathbb{C}$-algebras. See **?**.

There is an algebraic version of the above statement as well. Given an (irreducible) algebraic curve in projective space over an algebraically closed field k (e.g. the complex numbers), one can consider its "field of rational functions:" basically, functions that look like quotients of polynomials, where the denominator does not identically vanish on the curve. There is a similar anti-equivalence of categories between smooth projective curves and non-constant morphisms of curves and finitely generated extensions of k of transcendence degree one. See **?**.

Linearly Disjoint Field Extensions

Let k be a field, K and L field extensions of k. Suppose also that K and L are embedded in some larger field Ω.

3.6.10 Definition The compositum of K and L written KL is $k(K \cup L) = L(K) = K(L)$.

3.6.11 Definition K and L are said to be linearly disjoint over k if the following map is injective:

$$\theta : K \otimes_k L \to KL \tag{3.6.2}$$

defined by $x \otimes y \mapsto xy$.

I.4. Three important functors

There are three functors that will be integral to our study of commutative algebra in the future: localization, the tensor product, and hom. While localization is an *exact* functor, the tensor product and hom are not. The failure of exactness in those cases leads to the theory of flatness and projectivity (and injectivity), and eventually the *derived functors* Tor and Ext that crop up in commutative algebra.

4.1. Localization

Localization is the process of making invertible a collection of elements in a ring. It is a generalization of the process of forming a quotient field of an integral domain.

Geometric intuition

We first start off with some of the geometric intuition behind the idea of localization. Suppose we have a Riemann surface X (for example, the Riemann sphere). Let $A(U)$ be the ring of holomorphic functions over some neighborhood $U \subset X$. Now, for holomorphicity to hold, all that is required is that a function doesn't have a pole inside of U, thus when $U = X$, this condition is the strictest and as U gets smaller functions begin to show up that may not arise from the restriction of a holomorphic function over a larger domain. For example, if we want to study holomorphicity "near a point z_0" all that we should require is that the function doesn't pole at z_0. This means that we should consider quotients of holomorphic functions f/g where $g(z_0) \neq 0$. This process of inverting a collection of elements is expressed through the algebraic construction known as "localization."

Localization at a multiplicative subset

Let R be a commutative ring. We start by constructing the notion of *localization* in the most general sense.

We have already implicitly used this definition, but nonetheless, we make it formally:

4.1.1 Definition A subset $S \subset R$ is a **multiplicative subset** if $1 \in S$ and if $x, y \in S$ implies $xy \in S$.

We now define the notion of *localization*. Formally, this means inverting things. This will give us a functor from R-modules to R-modules.

4.1.2 Definition If M is an R-module, we define the module $S^{-1}M$ as the set of formal fractions

$$\{m/s, m \in M, s \in S\}$$

modulo an equivalence relation: where $m/s \sim m'/s'$ if and only if

$$t(s'm - m's) = 0$$

for some $t \in S$. The reason we need to include the t in the definition is that otherwise the relation would not be transitive (i.e. would not be an equivalence relation).

So two fractions agree if they agree when clearing denominators and multiplication.

It is easy to check that this is indeed an equivalence relation. Moreover $S^{-1}M$ is an abelian group with the usual addition of fractions

$$\frac{m}{s} + \frac{m'}{s'} = \frac{s'm + sm'}{ss'}$$

and it is easy to check that this is a legitimate abelian group.

4.1.3 Definition Let M be an R-module and $S \subset R$ a multiplicative subset. The abelian group $S^{-1}M$ is naturally an R-module. We define

$$x(m/s) = (xm)/s, \quad x \in R.$$

It is easy to check that this is well-defined and makes it into a module.

Finally, we note that localization is a *functor* from the category of R-modules to itself. Indeed, given $f : M \to N$, there is a naturally induced map $S^{-1}M \stackrel{S^{-1}f}{\to} S^{-1}N$.

We now consider the special case when the localized module is the initial ring itself. Let $M = R$. Then $S^{-1}R$ is an R-module, and it is in fact a commutative ring in its own right. The ring structure is quite tautological:

$$(x/s)(y/s') = (xy/ss').$$

There is a map $R \to S^{-1}R$ sending $x \to x/1$, which is a ring-homomorphism.

4.1.4 Definition For $S \subset R$ a multiplicative set, the localization $S^{-1}R$ is a commutative ring as above. In fact, it is an R-algebra; there is a natural map $\phi : R \to S^{-1}R$ sending $r \to r/1$.

We can, in fact, describe $\phi : R \to S^{-1}R$ by a *universal property.* Note that for each $s \in S$, $\phi(s)$ is invertible. This is because $\phi(s) = s/1$ which has a multiplicative inverse $1/s$. This property characterizes $S^{-1}R$.

For any commutative ring B, $\hom(S^{-1}R, B)$ is naturally isomorphic to the subset of $\hom(R, B)$ that send S to units. The map takes $S^{-1}R \to B$ to the pull-back $R \to S^{-1}R \to B$. The proof of this is very simple. Suppose that $f : R \to B$ is such that $f(s) \in B$ is invertible for each $s \in S$. Then we must define $S^{-1}R \to B$ by sending r/s to $f(r)f(s)^{-1}$. It is easy to check that this is well-defined and that the natural isomorphism as claimed is true.

Let R be a ring, M an R-module, $S \subset R$ a multiplicatively closed subset. We defined a ring of fractions $S^{-1}R$ and an R-module $S^{-1}M$. But in fact this is a module over the ring $S^{-1}R$. We just multiply $(x/t)(m/s) = (xm/st)$.

In particular, localization at S gives a *functor* from R-modules to $S^{-1}R$-modules.

4.1.5 Remark (exercise) Let R be a ring, S a multiplicative subset. Let T be the R-algebra $R[\{x_s\}_{s\in S}]/(\{sx_s - 1\})$. This is the polynomial ring in the variables x_s, one for each $s \in S$, modulo the ideal generated by $sx_s = 1$. Prove that this R-algebra is naturally isomorphic to $S^{-1}R$, using the universal property.

4.1.6 Remark (exercise) Define a functor **Rings** $\to$ **Sets** sending a ring to its set of units, and show that it is corepresentable (use $\mathbb{Z}[X, X^{-1}]$).

Local rings

A special case of great importance in the future is when the multiplicative subset is the complement of a prime ideal, and we study this in the present subsec. Such localizations will be "local rings" and geometrically correspond to the process of zooming at a point.

4.1.7 Example Let R be an integral domain and let $S = R - \{0\}$. This is a multiplicative subset because R is a domain. In this case, $S^{-1}R$ is just the ring of fractions by allowing arbitrary nonzero denominators; it is a field, and is called the **quotient field**. The most familiar example is the construction of $\mathbb{Q}$ as the quotient field of $\mathbb{Z}$.

We'd like to generalize this example.

4.1.8 Example Let R be arbitrary and $\mathfrak{p}$ is a prime ideal. This means that $1 \notin \mathfrak{p}$ and $x, y \in R - \mathfrak{p}$ implies that $xy \in R - \mathfrak{p}$. Hence, the complement $S = R - \mathfrak{p}$ is multiplicatively closed. We get a ring $S^{-1}R$.

4.1.9 Definition This ring is denoted $R_{\mathfrak{p}}$ and is called the **localization at $\mathfrak{p}$.** If M is an R-module, we write $M_{\mathfrak{p}}$ for the localization of M at $R - \mathfrak{p}$.

This generalizes the previous example (where $\mathfrak{p} = (0)$).

There is a nice property of the rings $R_{\mathfrak{p}}$. To elucidate this, we start with a lemma.

4.1.10 Lemma *Let R be a nonzero commutative ring. The following are equivalent:*

1. *R has a unique maximal ideal.*
2. *If $x \in R$, then either x or $1 - x$ is invertible.*

4.1.11 Definition In this case, we call R **local**. A local ring is one with a unique maximal ideal.

Proof of the lemma. First we prove (2) $\implies$ (1).

Assume R is such that for each x, either x or $1-x$ is invertible. We will find the maximal ideal. Let $\mathfrak{M}$ be the collection of noninvertible elements of R. This is a subset of R, not containing 1, and it is closed under multiplication. Any proper ideal must be a subset of $\mathfrak{M}$, because otherwise that proper ideal would contain an invertible element.

We just need to check that $\mathfrak{M}$ is closed under addition. Suppose to the contrary that $x, y \in \mathfrak{M}$ but $x+y$ is invertible. We get (with $a = x/(x+y)$)

$$1 = \frac{x}{x+y} + \frac{y}{x+y} = a + (1-a).$$

Then one of $a, 1-a$ is invertible. So either $x(x+y)^{-1}$ or $y(x+y)^{-1}$ is invertible, which implies that either x, y is invertible, contradiction.

Now prove the reverse direction. Assume R has a unique maximal ideal $\mathfrak{M}$. We claim that $\mathfrak{M}$ consists precisely of the noninvertible elements. To see this, first note that $\mathfrak{M}$ can't contain any invertible elements since it is proper. Conversely, suppose x is not invertible, i.e. $(x) \subsetneq R$. Then (x) is contained in a maximal ideal by 2.6.8, so $(x) \subset \mathfrak{M}$ since $\mathfrak{M}$ is unique among maximal ideals. Thus $x \in \mathfrak{M}$.

Suppose $x \in R$; we can write $1 = x + (1-x)$. Since $1 \notin \mathfrak{M}$, one of $x, 1-x$ must not be in $\mathfrak{M}$, so one of those must not be invertible. So (1) $\implies$ (2). The lemma is proved. □

Let us give some examples of local rings.

4.1.12 Example Any field is a local ring because the unique maximal ideal is (0).

4.1.13 Example Let R be any commutative ring and $\mathfrak{p} \subset R$ a prime ideal. Then $R_{\mathfrak{p}}$ is a local ring.

We state this as a result.

4.1.14 Proposition *$R_{\mathfrak{p}}$ is a local ring if $\mathfrak{p}$ is prime.*

Proof. Let $\mathfrak{m} \subset R_{\mathfrak{p}}$ consist of elements x/s for $x \in \mathfrak{p}$ and $s \in R - \mathfrak{p}$. It is left as an exercise (using the primality of $\mathfrak{p}$) to the reader to see that whether the numerator belongs to $\mathfrak{p}$ is *independent* of the representation x/s used for it.

Then I claim that $\mathfrak{m}$ is the unique maximal ideal. First, note that $\mathfrak{m}$ is an ideal; this is evident since the numerators form an ideal. If $x/s, y/s'$ belong to $\mathfrak{m}$ with appropriate expressions, then the numerator of

$$\frac{xs' + ys}{ss'}$$

belongs to $\mathfrak{p}$, so this sum belongs to $\mathfrak{m}$. Moreover, $\mathfrak{m}$ is a proper ideal because $\frac{1}{1}$ is not of the appropriate form.

I claim that $\mathfrak{m}$ contains all other proper ideals, which will imply that it is the unique maximal ideal. Let $I \subset R_{\mathfrak{p}}$ be any proper ideal. Suppose $x/s \in I$. We want to prove $x/s \in \mathfrak{m}$. In other words, we have to show $x \in \mathfrak{p}$. But if not x/s would be invertible, and $I = (1)$, contradiction. This proves locality. □

4.1.15 Remark (exercise) Any local ring is of the form $R_\mathfrak{p}$ for some ring R and for some prime ideal $\mathfrak{p} \subset R$.

4.1.16 Example Let $R = \mathbb{Z}$. This is not a local ring; the maximal ideals are given by (p) for p prime. We can thus construct the localizations $\mathbb{Z}_{(p)}$ of all fractions $a/b \in \mathbb{Q}$ where $b \notin (p)$. Here $\mathbb{Z}_{(p)}$ consists of all rational numbers that don't have powers of p in the denominator.

4.1.17 Remark (exercise) A local ring has no idempotents other than 0 and 1. (Recall that $e \in R$ is *idempotent* if $e^2 = e$.) In particular, the product of two rings is never local.

It may not yet be clear why localization is such a useful process. It turns out that many problems can be checked on the localizations at prime (or even maximal) ideals, so certain proofs can reduce to the case of a local ring. Let us give a small taste.

4.1.18 Proposition *Let $f : M \to N$ be a homomorphism of R-modules. Then f is injective if and only if for every maximal ideal $\mathfrak{m} \subset R$, we have that $f_\mathfrak{m} : M_\mathfrak{m} \to N_\mathfrak{m}$ is injective.*

Recall that, by definition, $M_\mathfrak{m}$ is the localization at $R - \mathfrak{m}$.

There are many variants on this (e.g. replace with surjectivity, bijectivity). This is a general observation that lets you reduce lots of commutative algebra to local rings, which are easier to work with.

Proof. Suppose first that each $f_\mathfrak{m}$ is injective. I claim that f is injective. Suppose $x \in M - \{0\}$. We must show that $f(x) \neq 0$. If $f(x) = 0$, then $f_\mathfrak{m}(x) = 0$ for every maximal ideal $\mathfrak{m}$. Then by injectivity it follows that x maps to zero in each $M_\mathfrak{m}$. We would now like to get a contradiction.

Let $I = \{a \in R : ax = 0 \in M\}$. This is proper since $x \neq 0$. So I is contained in some maximal ideal $\mathfrak{m}$. Then x maps to zero in $M_\mathfrak{m}$ by the previous paragraph; this means that there is $s \in R - \mathfrak{m}$ with $sx = 0 \in M$. But $s \notin I$, contradiction.

Now let us do the other direction. Suppose f is injective and $\mathfrak{m}$ a maximal ideal; we prove $f_\mathfrak{m}$ injective. Suppose $f_\mathfrak{m}(x/s) = 0 \in N_\mathfrak{m}$. This means that $f(x)/s = 0$ in the localized module, so that $f(x) \in M$ is killed by some $t \in R - \mathfrak{m}$. We thus have $f(tx) = t(f(x)) = 0 \in M$. This means that $tx = 0 \in M$ since f is injective. But this in turn means that $x/s = 0 \in M_\mathfrak{m}$. This is what we wanted to show. $\square$

Localization is exact

Localization is to be thought of as a very mild procedure.

The next result says how inoffensive localization is. This result is a key tool in reducing problems to the local case.

4.1.19 Proposition *Suppose $f : M \to N, g : N \to P$ and $M \to N \to P$ is exact. Let $S \subset R$ be multiplicatively closed. Then*

$$S^{-1}M \to S^{-1}N \to S^{-1}P$$

is exact.

Or, as one can alternatively express it, localization is an *exact functor.*

Before proving it, we note a few corollaries:

4.1.20 Corollary *If $f : M \to N$ is surjective, then $S^{-1}M \to S^{-1}N$ is too.*

Proof. To say that $A \to B$ is surjective is the same as saying that $A \to B \to 0$ is exact. From this the corollary is evident. ☐

Similarly:

4.1.21 Corollary *If $f : M \to N$ is injective, then $S^{-1}M \to S^{-1}N$ is too.*

Proof. To say that $A \to B$ is injective is the same as saying that $0 \to A \to B$ is exact. From this the corollary is evident. ☐

Proof of the proposition. We adopt the notation of the proposition. If the composite $g \circ f$ is zero, clearly the localization $S^{-1}M \to S^{-1}N \to S^{-1}P$ is zero too. Call the maps $S^{-1}M \to S^{-1}N, S^{-1}N \to S^{-1}P$ as ϕ, ψ. We know that $\psi \circ \phi = 0$ so $\ker(\psi) \supset \mathrm{im}(\phi)$. Conversely, suppose something belongs to $\ker(\psi)$. This can be written as a fraction

$$x/s \in \ker(\psi)$$

where $x \in N, s \in S$. This is mapped to

$$g(x)/s \in S^{-1}P,$$

which we're assuming is zero. This means that there is $t \in S$ with $tg(x) = 0 \in P$. This means that $g(tx) = 0$ as an element of P. But $tx \in N$ and its image of g vanishes, so tx must come from something in M. In particular,

$$tx = f(y) \text{ for some } y \in M.$$

In particular,

$$\frac{x}{s} = \frac{tx}{ts} = \frac{f(y)}{ts} = \phi(y/ts) \in \mathrm{im}(\phi).$$

This proves that anything belonging to the kernel of ψ lies in $\mathrm{im}(\phi)$. ☐

Nakayama's lemma

We now state a very useful criterion for determining when a module over a *local* ring is zero.

4.1.22 Lemma (Nakayama's lemma) *If R is a local ring with maximal ideal $\mathfrak{m}$. Let M be a finitely generated R-module. If $\mathfrak{m}M = M$, then $M = 0$.*

Note that $\mathfrak{m}M$ is the submodule generated by products of elements of $\mathfrak{m}$ and M.

4.1.23 Remark Once one has the theory of the tensor product, this equivalently states that if M is finitely generated, then

$$M \otimes_R R/\mathfrak{m} = M/\mathfrak{m}M \neq 0.$$

So to prove that a finitely generated module over a local ring is zero, you can reduce to studying the reduction to $R/\mathfrak{m}$. This is thus a very useful criterion.

Nakayama's lemma highlights why it is so useful to work over a local ring. Thus, it is useful to reduce questions about general rings to questions about local rings. Before proving it, we note a corollary.

4.1.24 Corollary *Let R be a local ring with maximal ideal $\mathfrak{m}$, and M a finitely generated module. If $N \subset M$ is a submodule such that $N + \mathfrak{m}N = M$, then $N = M$.*

Proof. Apply Nakayama above (lemma 4.1.22) to M/N. ☐

We shall prove more generally:

4.1.25 Proposition *Suppose M is a finitely generated R-module, $J \subset R$ an ideal. Suppose $JM = M$. Then there is $a \in 1 + J$ such that $aM = 0$.*

If J is the maximal ideal of a local ring, then a is a unit, so that $M = 0$.

Proof. Suppose M is generated by $\{x_1, \dots, x_n\} \subset M$. This means that every element of M is a linear combination of elements of x_i. However, each $x_i \in JM$ by assumption. In particular, each x_i can be written as

$$x_i = \sum a_{ij} x_j, \text{ where } a_{ij} \in \mathfrak{m}.$$

If we let A be the matrix $\{a_{ij}\}$, then A sends the vector (x_i) into itself. In particular, $I - A$ kills the vector (x_i).

Now $I - A$ is an n-by-n matrix in the ring R. We could, of course, reduce everything modulo J to get the identity; this is because A consists of elements of J. It follows that the determinant must be congruent to 1 modulo J.

In particular, $a = \det(I - A)$ lies in $1 + J$. Now by familiar linear algebra, aI can be represented as the product of A and the matrix of cofactors; in particular, aI annihilates the vector (x_i), so that $aM = 0$. ☐

Before returning to the special case of local rings, we observe the following useful fact from ideal theory:

4.1.26 Proposition *Let R be a commutative ring, $I \subset R$ a finitely generated ideal such that $I^2 = I$. Then I is generated by an idempotent element.*

Proof. We know that there is $x \in 1 + I$ such that $xI = 0$. If $x = 1 + y, y \in I$, it follows that

$$yt = t$$

for all $t \in I$. In particular, y is idempotent and $(y) = I$. ☐

4.1.27 Remark (exercise) 4.1.26 fails if the ideal is not finitely generated.

4.1.28 Remark (exercise) Let M be a finitely generated module over a ring R. Suppose $f : M \to M$ is a surjection. Then f is an isomorphism. To see this, consider M as a module over $R[t]$ with t acting by f; since $(t)M = M$, argue that there is a polynomial $Q(t) \in R[t]$ such that $Q(t)t$ acts as the identity on M, i.e. $Q(f)f = 1_M$.

4.1.29 Remark (exercise) Give a counterexample to the conclusion of Nakayama's lemma when the module is not finitely generated.

4.1.30 Remark (exercise) Let M be a finitely generated module over the ring R. Let $\mathfrak{J}$ be the Jacobson radical of R (cf. 2.6.19). If $\mathfrak{J}M = M$, then $M = 0$.

4.1.31 Remark (exercise) [A converse to Nakayama's lemma] Suppose conversely that R is a ring, and $\mathfrak{a} \subset R$ an ideal such that $\mathfrak{a}M \neq M$ for every nonzero finitely generated R-module. Then $\mathfrak{a}$ is contained in every maximal ideal of R.

4.1.32 Remark (exercise) Here is an alternative proof of Nakayama's lemma. Let R be local with maximal ideal $\mathfrak{m}$, and let M be a finitely generated module with $\mathfrak{m}M = M$. Let n be the minimal number of generators for M. If $n > 0$, pick generators $x_1, \dots, x_n$. Then write $x_1 = a_1x_1 + \dots + a_nx_n$ where each $a_i \in \mathfrak{m}$. Deduce that x_1 is in the submodule generated by the $x_i, i \geqslant 2$, so that n was not actually minimal, contradiction.

Let M, M' be finitely generated modules over a local ring $(R, \mathfrak{m})$, and let $\phi : M \to M'$ be a homomorphism of modules. Then Nakayama's lemma gives a criterion for ϕ to be a surjection: namely, the map $\overline{\phi} : M/\mathfrak{m}M \to M'/\mathfrak{m}M'$ must be a surjection. For injections, this is false. For instance, if ϕ is multiplication by any element of $\mathfrak{m}$, then $\overline{\phi}$ is zero but ϕ may yet be injective. Nonetheless, we give a criterion for a map of *free* modules over a local ring to be a *split* injection.

4.1.33 Proposition *Let R be a local ring with maximal ideal $\mathfrak{m}$. Let F, F' be two finitely generated free R-modules, and let $\phi : F \to F'$ be a homomorphism. Then ϕ is a split injection if and only if the reduction $\overline{\phi}$*

$$F/\mathfrak{m}F \xrightarrow{\overline{\phi}} F'/\mathfrak{m}F'$$

is an injection.

Proof. One direction is easy. If ϕ is a split injection, then it has a left inverse $\psi : F' \to F$ such that $\psi \circ \phi = 1_F$. The reduction of ψ as a map $F'/\mathfrak{m}F' \to F/\mathfrak{m}F$ is a left inverse to $\overline{\phi}$, which is thus injective.

Conversely, suppose $\overline{\phi}$ injective. Let $e_1, \dots, e_r$ be a "basis" for F, and let $f_1, \dots, f_r$ be the images under ϕ in F'. Then the reductions $\overline{f_1}, \dots, \overline{f_r}$ are linearly independent in the $R/\mathfrak{m}$-vector space $F'/\mathfrak{m}F'$. Let us complete this to a basis of $F'/\mathfrak{m}F'$ by adding elements $\overline{g_1}, \dots, \overline{g_s} \in F'/\mathfrak{m}F'$, which we can lift to elements $g_1, \dots, g_s \in F'$. It is clear that F' has rank $r + s$ since its reduction $F'/\mathfrak{m}F'$ does.

We claim that the set $\{f_1, \dots, f_r, g_1, \dots, g_s\}$ is a basis for F'. Indeed, we have a map

$$R^{r+s} \to F'$$

of free modules of rank $r + s$. It can be expressed as an $r + s$-by-$r + s$ matrix M; we need to show that M is invertible. But if we reduce modulo $\mathfrak{m}$, it is invertible since the reductions of $f_1, \dots, f_r, g_1, \dots, g_s$ form a basis of $F'/\mathfrak{m}F'$. Thus the determinant of M is not in $\mathfrak{m}$, so by locality it is invertible. The claim about F' is thus proved.

We can now define the left inverse $F' \to F$ of ϕ. Indeed, given $x \in F'$, we can write it uniquely as a linear combination $\sum a_i f_i + \sum b_j g_j$ by the above. We define $\psi(\sum a_i f_i + \sum b_j g_j) = \sum a_i e_i \in F$. It is clear that this is a left inverse ◻

We next note a slight strenghtening of the above result, which is sometimes useful. Namely, the first module does not have to be free.

4.1.34 Proposition *Let R be a local ring with maximal ideal $\mathfrak{m}$. Let M, F be two finitely generated R-modules with F free, and let $\phi : M \to F'$ be a homomorphism. Then ϕ is a split injection if and only if the reduction $\overline{\phi}$*

$$M/\mathfrak{m}M \xrightarrow{\overline{\phi}} F/\mathfrak{m}F$$

is an injection.

It will in fact follow that M is itself free, because M is projective (see **??** below) as it is a direct summand of a free module.

Proof. Let L be a "free approximation" to M. That is, choose a basis $\overline{x_1}, \dots, \overline{x_n}$ for $M/\mathfrak{m}M$ (as an $R/\mathfrak{m}$-vector space) and lift this to elements $x_1, \dots, x_n \in M$. Define a map

$$L = R^n \to M$$

by sending the ith basis vector to x_i. Then $L/\mathfrak{m}L \to M/\mathfrak{m}M$ is an isomorphism. By Nakayama's lemma, $L \to M$ is surjective.

Then the composite map $L \to M \to F$ is such that the $L/\mathfrak{m}L \to F/\mathfrak{m}F$ is injective, so $L \to F$ is a split injection (by proposition 4.1.33). It follows that we can find a splitting $F \to L$, which when composed with $L \to M$ is a splitting of $M \to F$. ◻

4.1.35 Remark (exercise) Let A be a local ring, and B a ring which is finitely generated and free as an A-module. Suppose $A \to B$ is an injection. Then $A \to B$ is a *split injection.* (Note that any nonzero morphism mapping out of a field is injective.)

4.2. The functor hom

In any category, the morphisms between two objects form a set.[1] In many categories, however, the hom-sets have additional structure. For instance, the hom-sets between abelian groups are themselves abelian groups. The same situation holds for the category of modules over a commutative ring.

[1]Strictly speaking, this may depend on your set-theoretic foundations.

4.2.1 Definition Let R be a commutative ring and M, N to be R-modules. We write $\hom_R(M, N)$ for the set of all R-module homomorphisms $M \to N$. $\hom_R(M, N)$ is an R-module because one can add homomorphisms $f, g : M \to N$ by adding them pointwise: if f, g are homomorphisms $M \to N$, define $f + g : M \to N$ via $(f + g)(m) = f(m) + g(m)$; similarly, one can multiply homomorphisms $f : M \to N$ by elements $a \in R$: one sets $(af)(m) = a(f(m))$.

Recall that in any category, the hom-sets are *functorial*. For instance, given $f : N \to N'$, post-composition with f defines a map $\hom_R(M, N) \to \hom_R(M, N')$ for any M. Similarly precomposition gives a natural map $\hom_R(N', M) \to \hom_R(N, M)$. In particular, we get a bifunctor hom, contravariant in the first variable and covariant in the second, of R-modules into R-modules.

Left-exactness of hom

We now discuss the exactness properties of this construction of forming hom-sets. The following result is basic and is, in fact, a reflection of the universal property of the kernel.

4.2.2 Proposition *If M is an R-module, then the functor*

$$N \to \hom_R(M, N)$$

is left exact (but not exact *in general).*

This means that if

$$0 \to N' \to N \to N''$$

is exact, then

$$0 \to \hom_R(M, N') \to \hom_R(M, N) \to \hom_R(M, N'')$$

is exact as well.

Proof. First, we have to show that the map $\hom_R(M, N') \to \hom_R(M, N)$ is injective; this is because $N' \to N$ is injective, and composition with $N' \to N$ can't kill any nonzero $M \to N'$. Similarly, exactness in the middle can be checked easily, and follows from 2.5.11; it states simply that a map $M \to N$ has image landing inside N' (i.e. factors through N') if and only if it composes to zero in N''. $\square$

This functor $\hom_R(M, \cdot)$ is not exact in general. Indeed:

4.2.3 Example Suppose $R = \mathbb{Z}$, and consider the R-module (i.e. abelian group) $M = \mathbb{Z}/2\mathbb{Z}$. There is a short exact sequence

$$0 \to 2\mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0.$$

Let us apply $\hom_R(M, \cdot)$. We get a *complex*

$$0 \to \hom(\mathbb{Z}/2\mathbb{Z}, 2\mathbb{Z}) \to \hom(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) \to \hom(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \to 0.$$

The second-to-last term is $\mathbb{Z}/2\mathbb{Z}$; everything else is zero. Thus the sequence is not exact, and in particular the functor $\hom_{\mathbb{Z}}(\mathbb{Z}/2, -)$ is not an exact functor.

We have seen that homming out of a module is left-exact. Now, we see the same for homming *into* a module.

4.2.4 Proposition *If M is a module, then* $\hom_R(-, M)$ *is a left-exact contravariant functor.*

We write this proof in slightly more detail than proposition 4.2.2, because of the contravariance.

Proof. We want to show that $\hom(\cdot, M)$ is a left-exact contravariant functor, which means that if $A \xrightarrow{u} B \xrightarrow{v} C \to 0$ is exact, then so is

$$0 \to \hom(C, M) \xrightarrow{\mathbf{v}} \hom(B, M) \xrightarrow{\mathbf{u}} \hom(A, M)$$

is exact. Here, the bold notation refers to the induced maps of u, v on the hom-sets: if $f \in \hom(B, M)$ and $g \in \hom(C, M)$, we define $\mathbf{u}$ and $\mathbf{v}$ via $\mathbf{v}(g) = g \circ v$ and $\mathbf{u}(f) = f \circ u$.

Let us show first that $\mathbf{v}$ is injective. Suppose that $g \in \hom(C, M)$. If $\mathbf{v}(g) = g \circ v = 0$ then $(g \circ v)(b) = 0$ for all $b \in B$. Since v is a surjection, this means that $g(C) = 0$ and hence $g = 0$. Therefore, $\mathbf{v}$ is injective, and we have exactness at $\hom(C, M)$.

Since $v \circ u = 0$, it is clear that $\mathbf{u} \circ \mathbf{u} = 0$.

Now, suppose that $f \in \ker(\mathbf{u}) \subset \hom(B, M)$. Then $\mathbf{u}(f) = f \circ u = 0$. Thus $f : B \to M$ factors through $B/\operatorname{im}(u)$. However, $\operatorname{im}(u) = \ker(v)$, so f factors through $B/\ker(v)$. Exactness shows that there is an isomorphism $B/\ker(v) \simeq C$. In particular, we find that f factors through C. This is what we wanted. ∎

4.2.5 Remark (exercise) Come up with an example where $\hom_R(-, M)$ is not exact.

4.2.6 Remark (exercise) Over a *field*, hom is always exact.

Projective modules

Let M be an R-module for a fixed commutative ring R. We have seen that $\hom_R(M, -)$ is generally only a left-exact functor. Sometimes, however, we do have exactness. We axiomatize this with the following.

4.2.7 Definition An R-module M is called **projective** if the functor $\hom_R(M, \cdot)$ is exact.[2]

One may first observe that a free module is projective. Indeed, let $F = R^I$ for an indexing set. Then the functor $N \to \hom_R(F, N)$ is naturally isomorphic to $N \to N^I$. It is easy to see that this functor preserves exact sequences (that is, if $0 \to A \to B \to C \to 0$ is exact, so is $0 \to A^I \to B^I \to C^I \to 0$). Thus F is projective. One can also easily check that a *direct summand* of a projective module is projective.

It turns out that projective modules have a very clean characterization. They are *precisely* the direct summands in free modules.

add: check this

[2] It is possible to define a projective module over a noncommutative ring. The definition is the same, except that the hom-sets are no longer modules, but simply abelian groups.

4.2.8 Proposition *The following are equivalent for an R-module M:*

1. *M is projective.*
2. *Given any map $M \to N/N'$ from M into a quotient of R-module N/N', we can lift it to a map $M \to N$.*
3. *There is a module M' such that $M \oplus M'$ is free.*

Proof. The equivalence of 1 and 2 is just unwinding the definition of projectivity, because we just need to show that $\hom_R(M, \cdot)$ preserves surjective maps, i.e. quotients. ($\hom_R(M, \cdot)$ is already left-exact, after all.) To say that $\hom_R(M, N) \to \hom_R(M, N/N')$ is surjective is just the statement that any map $M \to N/N'$ can be lifted to $M \to N$.

Let us show that 2 implies 3. Suppose M satisfies 2. Then choose a surjection $P \twoheadrightarrow M$ where P is free, by proposition 2.8.6. Then we can write $M \simeq P/P'$ for a submodule $P' \subset P$. The isomorphism map $M \to P/P'$ leads by 2 to a lifting $M \to P$. In particular, there is a section of $P \to M$, namely this lifting. Since a section leads to a split exact sequence by **??**, we find then that $P \simeq \ker(P \to M) \oplus \operatorname{im}(M \to P) \simeq \ker(P \to M) \oplus M$, verifying 3 since P is free.

Now let us show that 3 implies 2. Suppose $M \oplus M'$ is free, isomorphic to P. Then a map $M \to N/N'$ can be extended to

$$P \to N/N'$$

by declaring it to be trivial on M'. But now $P \to N/N'$ can be lifted to N because P is free, and we have observed that a free module is projective above; alternatively, we just lift the image of a basis. This defines $P \to N$. We may then compose this with the inclusion $M \to P$ to get the desired map $M \to P \to N$, which is a lifting of $M \to N/N'$. ∎

Of course, the lifting $P \to N$ of a given map $P \to N/N'$ is generally not unique, and in fact is unique precisely when $\hom_R(P, N') = 0$.

So projective modules are precisely those with the following lifting property. Consider a diagram

$$\begin{tikzcd} & P \arrow[d] & \\ M \arrow[r] & M'' \arrow[r] & 0 \end{tikzcd}$$

where the bottom row is exact. Then, if P is projective, there is a lifting $P \to M$ making commutative the diagram

$$\begin{tikzcd} & P \arrow[d] \arrow[dl, dashed] & \\ M \arrow[r] & M'' \arrow[r] & 0 \end{tikzcd}$$

4.2.9 Corollary *Let M be a module. Then there is a surjection $P \twoheadrightarrow M$, where P is projective.*

Proof. Indeed, we know (2.8.6) that we can always get a surjection from a free module. Since free modules are projective by 4.2.8, we are done. ∎

4.2.10 Remark (exercise) Let R be a principal ideal domain, F' a submodule of a free module F. Show that F' is free. (Hint: well-order the set of generators of F, and climb up by transfinite induction.) In particular, any projective modules is free.

Example: the Serre-Swan theorem

We now briefly digress to describe an important correspondence between projective modules and vector bundles. The material in this section will not be used in the sequel.

Let X be a compact space. We shall not recall the topological notion of a *vector bundle* here.

We note, however, that if E is a (complex) vector bundle, then the set $\Gamma(X, E)$ of global sections is naturally a module over the ring $C(X)$ of complex-valued continuous functions on X.

4.2.11 Proposition *If E is a vector bundle on a compact Hausdorff space X, then there is a surjection $\mathcal{O}^N \twoheadrightarrow E$ for some N.*

Here $\mathcal{O}^N$ denotes the trivial bundle.

It is known that in the category of vector bundles, every epimorphism splits. In particular, it follows that E can be viewed as a *direct summand* of the bundle $\mathcal{O}^N$. Since $\Gamma(X, E)$ is then a direct summand of $\Gamma(X, \mathcal{O}^N) = C(X)^N$, we find that $\Gamma(X, E)$ is a direct summand of a projective $C(X)$-module. Thus:

4.2.12 Proposition $\Gamma(X, E)$ *is a finitely generated projective $C(X)$-module.*

4.2.13 Theorem (Serre-Swan) *The functor $E \mapsto \Gamma(X, E)$ induces an equivalence of categories between vector bundles on X and finitely generated projective modules over $C(X)$.*

Injective modules

We have given a complete answer to the question of when the functor $\hom_R(M, -)$ is exact. We have shown that there are a lot of such *projective* modules in the category of R-modules, enough that any module admits a surjection from one such. However, we now have to answer the dual question: when is the functor $\hom_R(-, Q)$ exact?

Let us make the dual definition:

4.2.14 Definition An R-module Q is **injective** if the functor $\hom_R(-, Q)$ is exact.

Thus, a module Q over a ring R is injective if whenever $M \to N$ is an injection, and one has a map $M \to Q$, it can be extended to $N \to Q$: in other words, $\hom_R(N, Q) \to \hom_R(M, Q)$ is surjective. We can visualize this by a diagram

$$\begin{tikzcd} 0 \arrow[r] & M \arrow[r] \arrow[d] & N \arrow[dl, dashed] \\ & Q & \end{tikzcd}$$

where the dotted arrow always exists if Q is injective.

The notion is dual to projectivity, in some sense, so just as every module M admits an epimorphic map $P \to M$ for P projective, we expect by duality that every module admits a monomorphic map $M \to Q$ for Q injective. This is in fact true, but will require some work. We start, first, with a fact about injective abelian groups.

4.2.15 Theorem *A divisible abelian group (i.e. one where the map $x \to nx$ for any $n \in \mathbb{N}$ is surjective) is injective as a $\mathbb{Z}$-module (i.e. abelian group).*

Proof. The actual idea of the proof is rather simple, and similar to the proof of the Hahn-Banach theorem. Namely, we extend bit by bit, and then use Zorn's lemma.

The first step is that we have a subgroup M of a larger abelian group N. We have a map of $f : M \to Q$ for Q some divisible abelian group, and we want to extend it to N.

Now we can consider the poset of pairs $(\tilde{f}, M')$ where $M' \supset M$, and $\tilde{f} : M' \to N$ is a map extending f. Naturally, we make this into a poset by defining the order as "$(\tilde{f}, M') \leqslant (\tilde{f}', M'')$ if M'' contains M' and $\tilde{f}'$ is an extension of $\tilde{f}$. It is clear that every chain has an upper bound, so Zorn's lemma implies that we have a submodule $M' \subset N$ containing M, and a map $\tilde{f} : M' \to N$ extending f, such that there is no proper extension of $\tilde{f}$. From this we will derive a contradiction unless $M' = N$.

So suppose we have $M' \neq N$, for M' the maximal submodule to which f can be extended, as in the above paragraph. Pick $m \in N - M'$, and consider the submodule $M' + \mathbb{Z}m \subset N$. We are going to show how to extend $\tilde{f}$ to this bigger submodule. First, suppose $\mathbb{Z}m \cap M' = \{0\}$, i.e. the sum is direct. Then we can extend $\tilde{f}$ because $M' + \mathbb{Z}m$ is a direct sum: just define it to be zero on $\mathbb{Z}m$.

The slightly harder part is what happens if $\mathbb{Z}m \cap M' \neq \{0\}$. In this case, there is an ideal $I \subset \mathbb{Z}$ such that $n \in I$ if and only if $nm \in M'$. This ideal, however, is principal; let $g \in \mathbb{Z} - \{0\}$ be a generator. Then $gm = p \in M'$. In particular, $\tilde{f}(gm)$ is defined. We can "divide" this by g, i.e. find $u \in Q$ such that $gu = \tilde{f}(gm)$.

Now we may extend to a map $\tilde{f}'$ from $\mathbb{Z}m + M'$ into Q as follows. Choose $m' \in M', k \in \mathbb{Z}$. Define $\tilde{f}'(m' + km) = \tilde{f}(m') + ku$. It is easy to see that this is well-defined by the choice of u, and gives a proper extension of $\tilde{f}$. This contradicts maximality of M' and completes the proof. ☐

4.2.16 Remark (exercise) theorem 4.2.15 works over any principal ideal domain.

4.2.17 Remark (exercise) [Baer] Let N be an R-module such that for any ideal $I \subset R$, any morphism $I \to N$ can be extended to $R \to N$. Then N is injective. (Imitate the above argument.)

From this, we may prove:

4.2.18 Theorem *Any R-module M can be imbedded in an injective R-module Q.*

Proof. First of all, we know that any R-module M is a quotient of a free R-module. We are going to show that the dual (to be defined shortly) of a free module is injective. And so since every module admits a surjection from a free module, we will use a dualization argument to prove the present theorem.

First, for any abelian group G, define the **dual group** as $G^\vee = \hom_{\mathbb{Z}}(G, \mathbb{Q}/\mathbb{Z})$. Dualization is clearly a contravariant functor from abelian groups to abelian groups. By proposition 4.2.4 and theorem 4.2.15, an exact sequence of groups

$$0 \to A \to B \to C \to 0$$

induces an exact sequence

$$0 \to C^\vee \to B^\vee \to A^\vee \to 0.$$

In particular, dualization is an exact functor:

4.2.19 Proposition *Dualization preserves exact sequences (but reverses the order).*

Now, we are going to apply this to R-modules. The dual of a left R-module is acted upon by R. The action, which is natural enough, is as follows. Let M be an R-module, and $f : M \to \mathbb{Q}/\mathbb{Z}$ be a homomorphism of abelian groups (since $\mathbb{Q}/\mathbb{Z}$ has in general no R-module structure), and $r \in R$; then we define rf to be the map $M \to \mathbb{Q}/\mathbb{Z}$ defined via

$$(rf)(m) = f(rm).$$

It is easy to check that $M^\vee$ is thus made into an R-module.[3] In particular, dualization into $\mathbb{Q}/\mathbb{Z}$ gives a contravariant exact functor from *R-modules* to *R-modules*.

Let M be as before, and now consider the R-module $M^\vee$. By proposition 2.8.6, we can find a free module F and a surjection

$$F \to M^\vee \to 0.$$

Now dualizing gives an exact sequence of R-modules

$$0 \to M^{\vee\vee} \to F^\vee.$$

However, there is a natural map (of R-modules) $M \to M^{\vee\vee}$: given $m \in M$, we can define a functional $\hom(M, \mathbb{Q}/\mathbb{Z}) \to \mathbb{Q}/\mathbb{Z}$ by evaluation at m. One can check that this is a homomorphism. Moreover, this morphism $M \to M^{\vee\vee}$ is actually injective: if $m \in M$ were in the kernel, then by definition every functional $M \to \mathbb{Q}/\mathbb{Z}$ must vanish on m. It is easy to see (using $\mathbb{Z}$-injectivity of $\mathbb{Q}/\mathbb{Z}$) that this cannot happen if $m \neq 0$: we could just pick a nontrivial functional on the monogenic *subgroup* $\mathbb{Z}m$ and extend to M.

We claim now that $F^\vee$ is injective. This will prove the theorem, as we have the composite of monomorphisms $M \hookrightarrow M^{\vee\vee} \hookrightarrow F^\vee$ that embeds M inside an injective module.

4.2.20 Lemma *The dual of a free R-module F is an injective R-module.*

Proof. Let $0 \to A \to B$ be exact; we have to show that

$$\hom_R(B, F^\vee) \to \hom_R(A, F^\vee) \to 0.$$

is exact. Now we can reduce to the case where F is the R-module R itself. Indeed, F is a direct sum of R's by assumption, and taking hom's turns them into direct products; moreover the direct product of exact sequences is exact.

So we are reduced to showing that $R^\vee$ is injective. Now we claim that

$$\hom_R(B, R^\vee) = \hom_{\mathbb{Z}}(B, \mathbb{Q}/\mathbb{Z}). \tag{4.2.1}$$

∎

In particular, $\hom_R(-, R^\vee)$ is an exact functor because $\mathbb{Q}/\mathbb{Z}$ is an injective abelian group. The proof of eq. (4.2.1) is actually "trivial." For instance, a R-homomorphism $f : B \to R^\vee$ induces $\tilde{f} : B \to \mathbb{Q}/\mathbb{Z}$ by sending $b \to (f(b))(1)$. One checks that this is bijective.

[3] If R is noncommutative, this would not work: instead $M^\vee$ would be an *right* R-module. For commutative rings, we have no such distinction between left and right modules.

The small object argument

There is another, more set-theoretic approach to showing that any R-module M can be imbedded in an injective module. This approach, which constructs the injective module by a transfinite colimit of push-outs, is essentially analogous to the "small object argument" that one uses in homotopy theory to show that certain categories (e.g. the category of CW complexes) are model categories in the sense of Quillen; see **?**. While this method is somewhat abstract and more complicated than the one of section 4.2, it is also more general. Apparently this method originates with Baer, and was revisited by Cartan & Eilenberg (1999) and by Grothendieck (1957b). There, Grothendieck uses it to show that many other abelian categories have enough injectives.

We first begin with a few remarks on smallness. Let $\{B_\alpha\}, \alpha \in \mathcal{A}$ be an inductive system of objects in some category $\mathcal{C}$, indexed by an ordinal $\mathcal{A}$. Let us assume that $\mathcal{C}$ has (small) colimits. If A is an object of $\mathcal{C}$, then there is a natural map

$$\varinjlim \hom(A, B_\alpha) \to \hom(A, \varinjlim B_\alpha) \tag{4.2.2}$$

because if one is given a map $A \to B_\beta$ for some β, one naturally gets a map from A into the colimit by composing with $B_\beta \to \varinjlim B_\alpha$. (Note that the left colimit is one of sets!)

In general, the map eq. (4.2.2) is neither injective or surjective.

4.2.21 Example Consider the category of sets. Let $A = \mathbb{N}$ and $B_n = \{1, \dots, n\}$ be the inductive system indexed by the natural numbers (where $B_n \to B_m, n \leqslant m$ is the obvious map). Then $\varinjlim B_n = \mathbb{N}$, so there is a map

$$A \to \varinjlim B_n,$$

which does not factor as

$$A \to B_m$$

for any m. Consequently, $\varinjlim \hom(A, B_n) \to \hom(A, \varinjlim B_n)$ is not surjective.

4.2.22 Example Next we give an example where the map fails to be injective. Let $B_n = \mathbb{N}/\{1, 2, \dots, n\}$, that is, the quotient set of $\mathbb{N}$ with the first n elements collapsed to one element. There are natural maps $B_n \to B_m$ for $n \leqslant m$, so the $\{B_n\}$ form an inductive system. It is easy to see that the colimit $\varinjlim B_n = \{*\}$: it is the one-point set. So it follows that $\hom(A, \varinjlim B_n)$ is a one-element set.

However, $\varinjlim \hom(A, B_n)$ is *not* a one-element set. Consider the family of maps $A \to B_n$ which are just the natural projections $\mathbb{N} \to \mathbb{N}/\{1, 2, \dots, n\}$ and the family of maps $A \to B_n$ which map the whole of A to the class of 1. These two families of maps are distinct at each step and thus are distinct in $\varinjlim \hom(A, B_n)$, but they induce the same map $A \to \varinjlim B_n$.

Nonetheless, if A is a *finite set*, it is easy to see that for any sequence of sets $B_1 \to B_2 \to \dots$, we have

$$\varinjlim \hom(A, B_n) = \hom(A, \varinjlim B_n).$$

Proof. Let $f : A \to \varinjlim B_n$. The range of A is finite, containing say elements $c_1, \dots, c_r \in \varinjlim B_n$. These all come from some elements in B_N for N large by definition of the colimit. Thus we can define $\widetilde{f} : A \to B_N$ lifting f at a finite stage.

Next, suppose two maps $f_n : A \to B_m, g_n : A \to B_m$ define the same map $A \to \varinjlim B_n$. Then each of the finitely many elements of A gets sent to the same point in the colimit. By definition of the colimit for sets, there is $N \geqslant m$ such that the finitely many elements of A get sent to the same points in B_N under f and g. This shows that $\varinjlim \hom(A, B_n) \to \hom(A, \varinjlim B_n)$ is injective. ∎

The essential idea is that A is "small" relative to the long chain of compositions $B_1 \to B_2 \to \dots$, so that it has to factor through a finite step.

Let us generalize this.

4.2.23 Definition Let $\mathcal{C}$ be a category, I a class of maps, and ω an ordinal. An object $A \in \mathcal{C}$ is said to be ω-**small** (with respect to I) if whenever $\{B_\alpha\}$ is an inductive system parametrized by ω with maps in I, then the map

$$\varinjlim \hom(A, B_\alpha) \to \hom(A, \varinjlim B_\alpha)$$

is an isomorphism.

Our definition varies slightly from that of **?**, where only "nice" transfinite sequences $\{B_\alpha\}$ are considered.

In our applications, we shall begin by restricting ourselves to the category of R-modules for a fixed commutative ring R. We shall also take I to be the set of *monomorphisms,* or injections.[4] Then each of the maps

$$B_\beta \to \varinjlim B_\alpha$$

is an injection, so it follows that $\hom(A, B_\beta) \to \hom(A, \varinjlim B_\alpha)$ is one, and in particular the canonical map

$$\varinjlim \hom(A, B_\alpha) \to \hom(A, \varinjlim B_\alpha) \tag{4.2.3}$$

is an *injection.* We can in fact interpret the B_α's as subobjects of the big module $\varinjlim B_\alpha$, and think of their union as $\varinjlim B_\alpha$. (This is not an abuse of notation if we identify B_α with the image in the colimit.)

We now want to show that modules are always small for "large" ordinals ω. For this, we have to digress to do some set theory:

4.2.24 Definition Let ω be a *limit* ordinal, and κ a cardinal. Then ω is κ-**filtered** if every collection C of ordinals strictly less than ω and of cardinality at most κ has an upper bound strictly less than ω.

4.2.25 Example A limit ordinal (e.g. the natural numbers ω_0) is κ-filtered for any finite cardinal κ.

[4] There are, incidentally, categories, such as the category of rings, where a categorical epimorphism may not be a surjection of sets.

4.2.26 Proposition *Let κ be a cardinal. Then there exists a κ-filtered ordinal ω.*

Proof. If κ is finite, example 4.2.25 shows that any limit ordinal will do. Let us thus assume that κ is infinite.

Consider the smallest ordinal ω whose cardinality is strictly greater than that of κ. Then we claim that ω is κ-filtered. Indeed, if C is a collection of at most κ ordinals strictly smaller than ω, then each of these ordinals is of size at most κ. Thus the union of all the ordinals in C (which is an ordinal) is of size at most κ, so is strictly smaller than ω, and it provides an upper bound as in the definition. ☐

4.2.27 Proposition *Let M be a module, κ the cardinality of the set of its submodules. Then if ω is κ-filtered, then M is ω-small (with respect to injections).*

The proof is straightforward, but let us first think about a special case. If M is finite, then the claim is that for any inductive system $\{B_\alpha\}$ with injections between them, parametrized by a limit ordinal, any map $M \to \varinjlim B_\alpha$ factors through one of the B_α. But this is clear. M is finite, so since each element in the image must land inside one of the B_α, so all of M lands inside some finite stage.

Proof. We need only show that the map eq. (4.2.3) is a surjection when ω is κ-filtered. Let $f : A \to \varinjlim B_\alpha$ be a map. Consider the subobjects $\{f^{-1}(B_\alpha)\}$ of A, where B_α is considered as a subobject of the colimit. If one of these, say $f^{-1}(B_\beta)$, fills A, then the map factors through B_β.

So suppose to the contrary that all of the $f^{-1}(B_\alpha)$ were proper subobjects of A. However, we know that
$$\bigcup f^{-1}(B_\alpha) = f^{-1}\left(\bigcup B_\alpha\right) = A.$$

Now there are at most κ different subobjects of A that occur among the $f^{-1}(B_\alpha)$, by hypothesis. Thus we can find a set A of cardinality at most κ such that as α' ranges over A, the $f^{-1}(B_{\alpha'})$ range over *all* the $f^{-1}(B_\alpha)$.

However, A has an upper bound $\widetilde{\omega} < \omega$ as ω is κ-filtered. In particular, all the $f^{-1}(B_{\alpha'})$ are contained in $f^{-1}(B_{\widetilde{\omega}})$. It follows that $f^{-1}(B_{\widetilde{\omega}}) = A$. In particular, the map f factors through $B_{\widetilde{\omega}}$. ☐

From this, we will be able to deduce the existence of lots of injectives. Let us recall the criterion of Baer (remark 4.2.17): a module Q is injective if and only if in every commutative diagram

$$\begin{tikzcd} \mathfrak{a} \arrow[r] \arrow[d] & Q \\ R \arrow[ur, dashed] & \end{tikzcd}$$

for $\mathfrak{a} \subset R$ an ideal, the dotted arrow exists. In other words, we are trying to solve an *extension problem* with respect to the inclusion $\mathfrak{a} \hookrightarrow R$ into the module M.

If M is an R-module, then in general we may have a semi-complete diagram as above. In it, we can form the *push-out*

$$\begin{tikzcd} \mathfrak{a} \arrow[r] \arrow[d] & Q \arrow[d] \\ R \arrow[r] & R \oplus_{\mathfrak{a}} Q \end{tikzcd} .$$

Here the vertical map is injective, and the diagram commutes. The point is that we can extend $\mathfrak{a} \to Q$ to R *if* we extend Q to the larger module $R \oplus_{\mathfrak{a}} Q$.

The point of the small object argument is to repeat this procedure transfinitely many times.

4.2.28 Theorem *Let M be an R-module. Then there is an embedding $M \hookrightarrow Q$ for Q injective.*

Proof. We start by defining a functor $\mathbf{M}$ on the category of R-modules. Given N, we consider the set of all maps $\mathfrak{a} \to N$ for $\mathfrak{a} \subset R$ an ideal, and consider the push-out

$$\begin{tikzcd} \bigoplus \mathfrak{a} \arrow[r] \arrow[d] & N \arrow[d] \\ \bigoplus R \arrow[r] & N \oplus_{\bigoplus \mathfrak{a}} \bigoplus R \end{tikzcd} \tag{4.2.4}$$

where the direct sum of copies of R is taken such that every copy of an ideal $\mathfrak{a}$ corresponds to one copy of R. We define $\mathbf{M}(N)$ to be this push-out. Given a map $N \to N'$, there is a natural morphism of diagrams eq. (4.2.4), so $\mathbf{M}$ is a functor. Note furthermore that there is a natural transformation

$$N \to \mathbf{M}(N),$$

which is *always an injection.*

The key property of $\mathbf{M}$ is that if $\mathfrak{a} \to N$ is any morphism, it can be extended to $R \to \mathbf{M}(N)$, by the very construction of $\mathbf{M}(N)$. The idea will now be to apply $\mathbf{M}$ a transfinite number of times and to use the small object property.

We define for each ordinal ω a functor $\mathbf{M}_\omega$ on the category of R-modules, together with a natural injection $N \to \mathbf{M}_\omega(N)$. We do this by transfinite induction. First, $\mathbf{M}_1 = \mathbf{M}$ is the functor defined above. Now, suppose given an ordinal ω, and suppose $\mathbf{M}_{\omega'}$ is defined for $\omega' < \omega$. If ω has an immediate predecessor $\widetilde{\omega}$, we let

$$\mathbf{M}_\omega = \mathbf{M} \circ \mathbf{M}_{\widetilde{\omega}}.$$

If not, we let $\mathbf{M}_\omega(N) = \varinjlim_{\omega' < \omega} \mathbf{M}_{\omega'}(N)$. It is clear (e.g. inductively) that the $\mathbf{M}_\omega(N)$ form an inductive system over ordinals ω, so this is reasonable.

Let κ be the cardinality of the set of ideals in R, and let Ω be a κ-filtered ordinal. The claim is as follows.

4.2.29 Lemma *For any N, $\mathbf{M}_\Omega(N)$ is injective.*

If we prove this, we will be done. In fact, we will have shown that there is a *functorial* embedding of a module into an injective. Thus, we have only to prove this lemma.

Proof. By Baer's criterion (remark 4.2.17), it suffices to show that if $\mathfrak{a} \subset R$ is an ideal, then any map $f : \mathfrak{a} \to \mathbf{M}_\Omega(N)$ extends to $R \to \mathbf{M}_\Omega(N)$. However, we know since Ω is a limit ordinal that

$$\mathbf{M}_\Omega(N) = \varinjlim_{\omega<\Omega} \mathbf{M}_\omega(N),$$

so by proposition 4.2.27, we find that

$$\hom_R(\mathfrak{a}, \mathbf{M}_\Omega(N)) = \varinjlim_{\omega<\Omega} \hom_R(\mathfrak{a}, \mathbf{M}_\omega(N)).$$

This means in particular that there is some $\omega' < \Omega$ such that f factors through the submodule $\mathbf{M}_{\omega'}(N)$, as

$$f : \mathfrak{a} \to \mathbf{M}_{\omega'}(N) \to \mathbf{M}_\Omega(N).$$

However, by the fundamental property of the functor $\mathbf{M}$, we know that the map $\mathfrak{a} \to \mathbf{M}_{\omega'}(N)$ can be extended to

$$R \to \mathbf{M}(\mathbf{M}_{\omega'}(N)) = \mathbf{M}_{\omega'+1}(N), \qquad \square$$

and the last object imbeds in $\mathbf{M}_\Omega(N)$. In particular, f can be extended to $\mathbf{M}_\Omega(N)$. $\square$

Split exact sequences

add: additive functors preserve split exact seq Suppose that $0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{f} N \longrightarrow 0$ is a split short exact sequence. Since $\mathrm{Hom}_R(D, \cdot)$ is a left-exact functor, we see that

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{f'} \mathrm{Hom}_R(D, N)$$

is exact. In addition, $\mathrm{Hom}_R(D, L \oplus N) \cong \mathrm{Hom}_R(D, L) \oplus \mathrm{Hom}_R(D, N)$. Therefore, in the case that we start with a split short exact sequence $M \cong L \oplus N$, applying $\mathrm{Hom}_R(D, \cdot)$ does yield a split short exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{f'} \mathrm{Hom}_R(D, N) \longrightarrow 0.$$

Now, assume that

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{f'} \mathrm{Hom}_R(D, N) \longrightarrow 0$$

is a short exact sequence of abelian groups for all R-modules D. Set $D = R$ and using $\mathrm{Hom}_R(R, N) \cong N$ yields that $0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{f} N \longrightarrow 0$ is a short exact sequence.

Set $D = N$, so we have

$$0 \longrightarrow \mathrm{Hom}_R(N, L) \xrightarrow{\psi'} \mathrm{Hom}_R(N, M) \xrightarrow{f'} \mathrm{Hom}_R(N, N) \longrightarrow 0$$

Here, f' is surjective, so the identity map of $\mathrm{Hom}_R(N, N)$ lifts to a map $g \in \mathrm{Hom}_R(N, M)$ so that $f \circ g = f'(g) = id$. This means that g is a splitting homomorphism for the sequence $0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{f} N \longrightarrow 0$, and therefore the sequence is a split short exact sequence.

4.3. The tensor product

We shall now introduce the third functor of this chapter: the tensor product. The tensor product's key property is that it allows one to "linearize" bilinear maps. When taking the tensor product of rings, it provides a categorical coproduct as well.

Bilinear maps and the tensor product

Let R be a commutative ring, as usual. We have seen that the hom-sets $\hom_R(M, N)$ of R-modules M, N are themselves R-modules. Consequently, if we have three R-modules M, N, P, we can think about module-homomorphisms

$$M \xrightarrow{\lambda} \hom_R(N, P).$$

Suppose $x \in M, y \in N$. Then we can consider $\lambda(x) \in \hom_R(N, P)$ and thus we can consider the element $\lambda(x)(y) \in P$. We denote this element $\lambda(x)(y)$, which depends on the variables $x \in M, y \in N$, by $\lambda(x, y)$ for convenience; it is a function of two variables $M \times N \to P$.

There are certain properties of $\lambda(\cdot, \cdot)$ that we list below. Fix $x, x' \in M$; $y, y' \in N$; $a \in R$. Then:

1. $\lambda(x, y + y') = \lambda(x, y) + \lambda(x, y')$ because $\lambda(x)$ is additive.
2. $\lambda(x, ay) = a\lambda(x, y)$ because $\lambda(x)$ is an R-module homomorphism.
3. $\lambda(x + x', y) = \lambda(x, y) + \lambda(x', y)$ because λ is additive.
4. $\lambda(ax, y) = a\lambda(x, y)$ because λ is an R-module homomorphism.

Conversely, given a function $\lambda : M \times N \to P$ of two variables satisfying the above properties, it is easy to see that we can get a morphism of R-modules $M \to \hom_R(N, P)$.

4.3.1 Definition An **R-bilinear map** $\lambda : M \times N \to P$ is a map satisfying the above listed conditions. In other words, it is required to be R-linear in each variable separately.

The previous discussion shows that there is a *bijection* between R-bilinear maps $M \times N \to P$ with R-module maps $M \to \hom_R(N, P)$. Note that the first interpretation is symmetric in M, N; the second, by contrast, can be interpreted in terms of the old concepts of an R-module map. So both are useful.

4.3.2 Remark (exercise) Prove that a $\mathbb{Z}$-bilinear map out of $\mathbb{Z}/2 \times \mathbb{Z}/3$ is identically zero, whatever the target module.

Let us keep the notation of the previous discussion: in particular, M, N, P will be modules over a commutative ring R.

Given a bilinear map $M \times N \to P$ and a homomorphism $P \to P'$, we can clearly get a bilinear map $M \times N \to P'$ by composition. In particular, given M, N, there is a *covariant functor* from R-modules to **Sets** sending any R-module P to the collection of R-bilinear maps $M \times N \to P$. As usual, we are interested in when this functor is *corepresentable*. As a result, we are interested in *universal* bilinear maps out of $M \times N$.

4.3.3 Definition An R-bilinear map $\lambda : M \times N \to P$ is called **universal** if for all R-modules Q, the composition of $P \to Q$ with $M \times N \xrightarrow{\lambda} P$ gives a **bijection**

$$\hom_R(P, Q) \simeq \{\text{bilinear maps } M \times N \to Q\}$$

So, given a bilinear map $M \times N \to Q$, there is a *unique* map $P \to Q$ making the diagram

$$\begin{tikzcd} & P \arrow[dd] \\ M \times N \arrow[ur, "\lambda"] \arrow[dr] & \\ & Q \end{tikzcd}$$

Alternatively, P *corepresents* the functor $Q \to \{\text{bilinear maps } M \times N \to Q\}$.

General nonsense says that given M, N, an universal R-bilinear map $M \times N \to P$ is **unique** up to isomorphism (if it exists). This follows from *Yoneda's lemma.* For convenience, we give a direct proof.

Suppose $M \times N \xrightarrow{\lambda} P$ was universal and $M \times N \xrightarrow{\lambda'} P'$ is also universal. Then by the universal property, there are unique maps $P \to P'$ and $P' \to P$ making the following diagram commutative:

$$\begin{tikzcd} & P \arrow[dd, shift left] \\ M \times N \arrow[ur, "\lambda"] \arrow[dr, "\lambda'"'] & \\ & P' \arrow[uu, shift left] \end{tikzcd}$$

These compositions $P \to P' \to P, P' \to P \to P'$ have to be the identity because of the uniqueness part of the universal property. As a result, $P \to P'$ is an isomorphism.

We shall now show that this universal object does indeed exist.

4.3.4 Proposition *Given M, N, a universal bilinear map out of $M \times N$ exists.*

Before proving it we make:

4.3.5 Definition We denote the codomain of the universal map out of $M \times N$ by $M \otimes_R N$. This is called the **tensor product** of M, N, so there is a universal bilinear map out of $M \times N$ into $M \otimes_R N$.

Proof of 4.3.4. We will simply give a presentation of the tensor product by "generators and relations." Take the free R-module $M \otimes_R N$ generated by the symbols $\{x \otimes y\}_{x \in M, y \in N}$ and quotient out by the relations forced upon us by the definition of a bilinear map (for $x, x' \in M,\ y, y' \in N,\ a \in R$)

1. $(x + x') \otimes y = x \otimes y + x' \otimes y$.
2. $(ax) \otimes y = a(x \otimes y) = x \otimes (ay)$.
3. $x \otimes (y + y') = x \otimes y + x \otimes y'$.

We will abuse notation and denote $x \otimes y$ for its image in $M \otimes_R N$ (as opposed to the symbol generating the free module).

There is a bilinear map $M \times N \to M \otimes_R N$ sending $(x, y) \to x \otimes y$; the relations imposed imply that this map is a bilinear map. We have to check that it is universal, but this is actually quite direct.

Suppose we had a bilinear map $\lambda : M \times N \to P$. We must construct a linear map $M \otimes_R N \to P$. To do this, we can just give a map on generators, and show that it is zero on each of the relations. It is easy to see that to make the appropriate diagrams commute, the linear map $M \otimes N \to P$ has to send $x \otimes y \to \lambda(x, y)$. This factors through the relations on $x \otimes y$ by bilinearity and leads to an R-linear map $M \otimes_R N \to P$ such that the following diagram commutes:

$$\begin{tikzcd} M \times N \arrow[r] \arrow[dr, "\lambda"] & M \otimes_R N \arrow[d] \\ & P \end{tikzcd} .$$

It is easy to see that $M \otimes_R N \to P$ is unique because the $x \otimes y$ generate it. ∎

The theory of the tensor product allows one to do away with bilinear maps and just think of linear maps.

Given M, N, we have constructed an object $M \otimes_R N$. We now wish to see the functoriality of the tensor product. In fact, $(M, N) \to M \otimes_R N$ is a *covariant functor* in two variables from R-modules to R-modules. In particular, if $M \to M', N \to N'$ are morphisms, there is a canonical map

$$M \otimes_R N \to M' \otimes_R N'. \tag{4.3.1}$$

To obtain eq. (4.3.1), we take the natural bilinear map $M \times N \to M' \times N' \to M' \otimes_R N'$ and use the universal property of $M \otimes_R N$ to get a map out of it.

Basic properties of the tensor product

We make some observations and prove a few basic properties. As the proofs will show, one powerful way to prove things about an object is to reason about its universal property. If two objects have the same universal property, they are isomorphic.

4.3.6 Proposition *The tensor product is symmetric: for R-modules M, N, we have $M \otimes_R N \simeq N \otimes_R M$ canonically.*

Proof. This is clear from the universal properties: giving a bilinear map out of $M \times N$ is the same as a bilinear map out $N \times M$. Thus $M \otimes_R N$ and $N \otimes_R N$ have the same universal property. It is also clear from the explicit construction. ∎

4.3.7 Proposition *For an R-module M, there is a canonical isomorphism $M \to M \otimes_R R$.*

Proof. If we think in terms of bilinear maps, this statement is equivalent to the statement that a bilinear map $\lambda : M \times R \to P$ is the same as a linear map $M \to N$. Indeed, to do this, restrict λ to $\lambda(\cdot, 1)$. Given $f : M \to N$, similarly, we take for λ as $\lambda(x, a) = af(x)$. This gives a bijection as claimed. ∎

4.3.8 Proposition *The tensor product is associative. There are canonical isomorphisms $M \otimes_R (N \otimes_R P) \simeq (M \otimes_R N) \otimes_R P$.*

Proof. There are a few ways to see this: one is to build it explicitly from the construction given, sending $x \otimes (y \otimes z) \to (x \otimes y) \otimes z$.

More conceptually, both have the same universal property: by general categorical nonsense (Yoneda's lemma), we need to show that for all Q, there is a canonical bijection

$$\hom_R(M \otimes (N \otimes P)), Q) \simeq \hom_R((M \otimes N) \otimes P, Q)$$

where the R's are dropped for simplicity. But both of these sets can be identified with the set of trilinear maps[5] $M \times N \times P \to Q$. Indeed

$$\begin{aligned} \hom_R(M \otimes (N \otimes P), Q) &\simeq \text{bilinear } M \times (N \otimes P) \to Q \\ &\simeq \hom(N \otimes P, \hom(M, Q)) \\ &\simeq \text{bilinear } N \times P \to \hom(M, Q) \\ &\simeq \hom(N, \hom(P, \hom(M, Q)) \\ &\simeq \text{trilinear maps.} \end{aligned}$$

∎

The adjoint property

Finally, while we defined the tensor product in terms of a "universal bilinear map," we saw earlier that bilinear maps could be interpreted as maps into a suitable hom-set. In particular, fix R-modules M, N, P. We know that the set of bilinear maps $M \times N \to P$ is naturally in bijection with

$$\hom_R(M, \hom_R(N, P))$$

as well as with

$$\hom_R(M \otimes_R, N, P).$$

As a result, we find:

4.3.9 Proposition *For R-modules M, N, P, there is a natural bijection*

$$\hom_R(M, \hom_R(N, P)) \simeq \hom_R(M \otimes_R N, P).$$

[5] Easy to define.

There is a more evocative way of phrasing the above natural bijection. Given N, let us define the functors F_N, G_N via

$$F_N(M) = M \otimes_R N, \quad G_N(P) = \hom_R(N, P).$$

Then the above proposition states that there is a natural isomorphism

$$\hom_R(F_N(M), P) \simeq \hom_R(M, G_N(P)).$$

In particular, F_N and G_N are *adjoint functors.* So, in a sense, the operations of hom and $\otimes$ are dual to each other.

4.3.10 Proposition *Tensoring commutes with colimits.*

In particular, it follows that if $\{N_\alpha\}$ is a family of modules, and M is a module, then

$$M \otimes_R \bigoplus N_\alpha = \bigoplus M \otimes_R N_\alpha.$$

4.3.11 Remark (exercise) Give an explicit proof of the above relation.

Proof. This is a formal consequence of the fact that the tensor product is a left adjoint and consequently commutes with all colimits. **add: proof** □

In particular, by proposition 4.3.10, the tensor product commutes with *cokernels.* That is, if $A \to B \to C \to 0$ is an exact sequence of R-modules and M is an R-module, $A \otimes_R M \to B \otimes_R M \to C \otimes_R M \to 0$ is also exact, because exactness of such a sequence is precisely a condition on the cokernel. That is, the tensor product is *right exact.*

We can thus prove a simple result on finite generation:

4.3.12 Proposition *If M, N are finitely generated, then $M \otimes_R N$ is finitely generated.*

Proof. Indeed, if we have surjections $R^m \to M, R^n \to N$, we can tensor them; we get a surjection since the tensor product is right-exact. So have a surjection $R^{mn} = R^m \otimes_R R^n \to M \otimes_R N$. □

The tensor product as base-change

Before this, we have considered the tensor product as a functor within a fixed category. Now, we shall see that when one takes the tensor product with a *ring*, one gets additional structure. As a result, we will be able to get natural functors between *different* module categories.

Suppose we have a ring-homomorphism $\phi : R \to R'$. In this case, any R'-module can be regarded as an R-module. In particular, there is a canonical functor of *restriction*

$$R'\text{-modules} \to R\text{-modules}.$$

We shall see that the tensor product provides an *adjoint* to this functor. Namely, if M has an R-module structure, then $M \otimes_R R'$ has an R' module structure where R' acts on the right. Since the tensor product is functorial, this gives a functor in the opposite direction:

$$R\text{-modules} \to R'\text{-modules}.$$

Let M' be an R'-module and M an R-module. In view of the above, we can talk about

$$\hom_R(M, M')$$

by thinking of M' as an R-module.

4.3.13 Proposition *There is a canonical isomorphism between*

$$\hom_R(M, M') \simeq \hom_{R'}(M \otimes_R R', M').$$

In particular, the restriction functor and the functor $M \to M \otimes_R R'$ are adjoints to each other.

Proof. We can describe the bijection explicitly. Given an R'-homomorphism $f : M \otimes_R R' \to M'$, we get a map

$$f_0 : M \to M'$$

sending

$$m \to m \otimes 1 \to f(m \otimes 1).$$

This is easily seen to be an R-module-homomorphism. Indeed,

$$f_0(ax) = f(ax \otimes 1) = f(\phi(a)(x \otimes 1)) = af(x \otimes 1) = af_0(x)$$

since f is an R'-module homomorphism.

Conversely, if we are given a homomorphism of R-modules

$$f_0 : M \to M'$$

then we can define

$$f : M \otimes_R R' \to M'$$

by sending $m \otimes r' \to r' f_0(m)$, which is a homomorphism of R' modules. This is well-defined because f_0 is a homomorphism of R-modules. We leave some details to the reader. ∎

4.3.14 Example In the representation theory of finite groups, the operation of tensor product corresponds to the procedure of *inducing* a representation. Namely, if $H \subset G$ is a subgroup of a group G, then there is an obvious restriction functor from G-representations to H-representations. The adjoint to this is the induction operator. Since a H-representation (resp. a G-representation) is just a module over the group ring, the operation of induction is really a special case of the tensor product. Note that the group rings are generally not commutative, so this should be interpreted with some care.

Some concrete examples

We now present several concrete computations of tensor products in explicit cases to illuminate what is happening.

4.3.15 Example Let us compute $\mathbb{Z}/10 \otimes_{\mathbb{Z}} \mathbb{Z}/12$. Since 1 spans $\mathbb{Z}/(10)$ and 1 spans $\mathbb{Z}/(12)$, we see that $1 \otimes 1$ spans $\mathbb{Z}/(10) \otimes \mathbb{Z}/(12)$ and this tensor product is a cyclic group.

Note that $1 \otimes 0 = 1 \otimes (10 \cdot 0) = 10 \otimes 0 = 0 \otimes 0 = 0$ and $0 \otimes 1 = (12 \cdot 0) \otimes 1 = 0 \otimes 12 = 0 \otimes 0 = 0$. Now, $10(1 \otimes 1) = 10 \otimes 1 = 0 \otimes 1 = 0$ and $12(1 \otimes 1) = 1 \otimes 12 = 1 \otimes 0 = 0$, so the cyclic group $\mathbb{Z}/(10) \otimes \mathbb{Z}/(12)$ has order dividing both 10 and 12. This means that the cyclic group has order dividing $\gcd(10, 12) = 2$.

To show that the order of $\mathbb{Z}/(10) \otimes \mathbb{Z}/(12)$, define a bilinear map $g : \mathbb{Z}/(10) \times \mathbb{Z}/(12) \to \mathbb{Z}/(2)$ via $g : (x, y) \mapsto xy$. The universal property of tensor products then says that there is a unique linear map $f : \mathbb{Z}/(10) \otimes \mathbb{Z}/(12) \to \mathbb{Z}/(2)$ making the diagram

$$\begin{tikzcd}
\mathbb{Z}/(10) \times \mathbb{Z}/(12) \arrow[r, "\otimes"] \arrow[dr, "g"'] & \mathbb{Z}/(10) \otimes \mathbb{Z}/(12) \arrow[d, "f"] \\
& \mathbb{Z}/(2).
\end{tikzcd}$$

commute. In particular, this means that $f(x \otimes y) = g(x, y) = xy$. Hence, $f(1 \otimes 1) = 1$, so f is surjective, and therefore, $\mathbb{Z}/(10) \otimes \mathbb{Z}/(12)$ has size at least two. This allows us to conclude that $\mathbb{Z}/(10) \otimes \mathbb{Z}/(12) = \mathbb{Z}/(2)$.

We now generalize the above example to tensor products of cyclic groups.

4.3.16 Example Let $d = \gcd(m, n)$. We will show that $(\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z}) \simeq (\mathbb{Z}/d\mathbb{Z})$, and thus in particular if m and n are relatively prime, then $(\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z}) \simeq (0)$. First, note that any $a \otimes b \in (\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z})$ can be written as $ab(1 \otimes 1)$, so that $(\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z})$ is generated by $1 \otimes 1$ and hence is a cyclic group. We know from elementary number theory that $d = xm + yn$ for some $x, y \in \mathbb{Z}$. We have $m(1 \otimes 1) = m \otimes 1 = 0 \otimes 1 = 0$ and $n(1 \otimes 1) = 1 \otimes n = 1 \otimes 0 = 0$. Thus $d(1 \otimes 1) = (xm + yn)(1 \otimes 1) = 0$, so that $1 \otimes 1$ has order dividing d.

Conversely, consider the map $f : (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}) \to (\mathbb{Z}/d\mathbb{Z})$ defined by $f(a + m\mathbb{Z}, b + n\mathbb{Z}) = ab + d\mathbb{Z}$. This is well-defined, since if $a' + m\mathbb{Z} = a + m\mathbb{Z}$ and $b' + n\mathbb{Z} = b + n\mathbb{Z}$ then $a' = a + mr$ and $b' = b + ns$ for some r, s and thus $a'b' + d\mathbb{Z} = ab + (mrb + nsa + mnrs) + d\mathbb{Z} = ab + d\mathbb{Z}$ (since $d = \gcd(m, n)$ divides m and n). This is obviously bilinear, and hence induces a map $\tilde{f} : (\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z}) \to (\mathbb{Z}/d\mathbb{Z})$, which has $\tilde{f}(1 \otimes 1) = 1 + d\mathbb{Z}$. But the order of $1 + d\mathbb{Z}$ in $\mathbb{Z}/d\mathbb{Z}$ is d, so that the order of $1 \otimes 1$ in $(\mathbb{Z}/m\mathbb{Z}) \otimes (\mathbb{Z}/n\mathbb{Z})$ must be at least d. Thus $1 \otimes 1$ is in fact of order d, and the map $\tilde{f}$ is an isomorphism between cyclic groups of order d.

Finally, we present an example involving the interaction of hom and the tensor product.

4.3.17 Example Given an R-module M, let us use the notation $M^* = \hom_R(M, R)$. We shall define a functorial map

$$M^* \otimes_R N \to \hom_R(M, N),$$

and show that it is an isomorphism when M is finitely generated and free.

Define $\rho' : M^* \times N \to \hom_R(M,N)$ by $\rho'(f,n)(m) = f(m)n$ (note that $f(m) \in R$, and the multiplication $f(m)n$ is that between an element of R and an element of N). This is bilinear,

$$\rho'(af+bg,n)(m) = (af+bg)(m)n = (af(m)+bg(m))n = af(m)n+bg(m)n = a\rho'(f,n)(m)+b\rho'(g,n)(m)$$

$$\rho'(f,an_1+bn_2)(m) = f(m)(an_1+bn_2) = af(m)n_1+bf(m)n_2 = a\rho'(f,n_1)(m)+b\rho'(f,n_2)(m)$$

so it induces a map $\rho : M^* \otimes N \to \hom(M,N)$ with $\rho(f \otimes n)(m) = f(m)n$. This homomorphism is unique since the $f \otimes n$ generate $M^* \otimes N$.

Suppose M is free on the set $\{a_1,\dots,a_k\}$. Then $M^* = \hom(M,R)$ is free on the set $\{f_i : M \to R,$ $f_i(r_1a_1+\cdots+r_ka_k) = r_i\}$, because there are clearly no relations among the f_i and because any $f : M \to R$ has $f = f(a_1)f_1+\cdots+f(a_n)f_n$. Also note that any element $\sum h_j \otimes p_j \in M^* \otimes N$ can be written in the form $\sum_{i=1}^k f_i \otimes n_i$, by setting $n_i = \sum h_j(a_i)p_j$, and *that this is unique* because the f_i are a basis for M^*.

We claim that the map $\psi : \hom_R(M,N) \to M^* \otimes N$ defined by $\psi(g) = \sum_{i=1}^k f_i \otimes g(a_i)$ is inverse to ρ. Given any $\sum_{i=1}^k f_i \otimes n_i \in M^* \otimes N$, we have

$$\rho(\sum_{i=1}^k f_i \otimes n_i)(a_j) = \sum_{i=1}^k \rho(f_i \otimes n_i)(a_j) = \sum_{i=1}^k f_i(a_j)n_i = n_j$$

Thus, $\rho(\sum_{i=1}^k f_i \otimes n_i)(a_i) = n_i$, and thus $\psi(\rho(\sum_{i=1}^k f_i \otimes n_i)) = \sum_{i=1}^k f_i \otimes n_i$. Thus, $\psi \circ \rho = \mathrm{id}_{M^* \otimes N}$.

Conversely, recall that for $g : M \to N \in \hom_R(M,N)$, we defined $\psi(g) = \sum_{i=1}^k f_i \otimes g(a_i)$. Thus,

$$\rho(\psi(g))(a_j) = \rho(\sum_{i=1}^k f_i \otimes g(a_i))(a_j) = \sum_{i=1}^k \rho(f_i \otimes g(a_i))(a_j) = \sum_{i=1}^k f_i(a_j)g(a_i) = g(a_j)$$

and because $\rho(\psi(g))$ agrees with g on the a_i, it is the same element of $\hom_R(M,N)$ because the a_i generate M. Thus, $\rho \circ \psi = \mathrm{id}_{\hom_R(M,N)}$.

Thus, ρ is an isomorphism.

We now interpret localization as a tensor product.

4.3.18 Proposition *Let R be a commutative ring, $S \subset R$ a multiplicative subset. Then there exists a canonical isomorphism of functors:*

$$\phi : S^{-1}M \simeq S^{-1}R \otimes_R M.$$

Proof. Here is a construction of ϕ. If $x/s \in S^{-1}M$ where $x \in M, s \in S$, we define

$$\phi(x/s) = (1/s) \otimes m.$$

Let us check that this is well-defined. Suppose $x/s = x'/s'$; then this means there is $t \in S$ with

$$xs't = x'st.$$

From this we need to check that $\phi(x/s) = \phi(x'/s')$, i.e. that $1/s \otimes x$ and $1/s' \otimes x'$ represent the same elements in the tensor product. But we know from the last statement that

$$\frac{1}{ss't} \otimes xs't = \frac{1}{ss't} x'st \in S^{-1}R \otimes M$$

and the first is just

$$s't(\frac{1}{ss't} \otimes x) = \frac{1}{s} \otimes x$$

by linearity, while the second is just

$$\frac{1}{s'} \otimes x'$$

similarly. One next checks that ϕ is an R-module homomorphism, which we leave to the reader.

Finally, we need to describe the inverse. The inverse $\psi : S^{-1}R \otimes M \to S^{-1}M$ is easy to construct because it's a map out of the tensor product, and we just need to give a bilinear map

$$S^{-1}R \times M \to S^{-1}M,$$

and this sends $(r/s, m)$ to mr/s.

It is easy to see that ϕ, ψ are inverses to each other by the definitions. □

It is, perhaps, worth making a small categorical comment, and offering an alternative argument. We are given two functors F, G from R-modules to $S^{-1}R$-modules, where $F(M) = S^{-1}R \otimes_R M$ and $G(M) = S^{-1}M$. By the universal property, the map $M \to S^{-1}M$ from an R-module to a tensor product gives a natural map

$$S^{-1}R \otimes_R M \to S^{-1}M,$$

that is a natural transformation $F \to G$. Since it is an isomorphism for free modules, it is an isomorphism for all modules by a standard argument.

Tensor products of algebras

There is one other basic property of tensor products to discuss before moving on: namely, what happens when one tensors a ring with another ring. We shall see that this gives rise to *push-outs* in the category of rings, or alternatively, coproducts in the category of R-algebras. Let R be a commutative ring and suppose R_1, R_2 are R-algebras. That is, we have ring homomorphisms $\phi_0 : R \to R_0, \quad \phi_1 : R \to R_1$.

4.3.19 Proposition *$R_0 \otimes_R R_1$ has the structure of a commutative ring in a natural way.*

Indeed, this multiplication multiplies two typical elements $x \otimes y, x' \otimes y'$ of the tensor product by sending them to $xx' \otimes yy'$. The ring structure is determined by this formula. One ought to check that this approach respects the relations of the tensor product. We will do so in an indirect way.

Proof. Notice that giving a multiplication law on $R_0 \otimes_R R_1$ is equivalent to giving an R-bilinear map

$$(R_0 \otimes_R R_1) \times (R_0 \otimes R_1) \to R_0 \otimes_R R_1,$$

i.e. an R-linear map

$$(R_0 \otimes_R R_1) \otimes_R (R_0 \otimes R_1) \to R_0 \otimes_R R_1$$

which satisfies certain constraints (associativity, commutativity, etc.). But the left side is isomorphic to $(R_0 \otimes_R R_0) \otimes_R (R_1 \otimes_R R_1)$. Since we have bilinear maps $R_0 \times R_0 \to R_0$ and $R_1 \times R_1 \to R_1$, we get linear maps $R_0 \otimes_R R_0 \to R_0$ and $R_1 \otimes_R R_1 \to R_1$. Tensoring these maps gives the multiplication as a bilinear map. It is easy to see that these two approaches are the same.

We now need to check that this operation is commutative and associative, with $1 \otimes 1$ as a unit; moreover, it distributes over addition. Distributivity over addition is built into the construction (i.e. in view of bilinearity). The rest (commutativity, associativity, units) can be checked directly on the generators, since we have distributivity. We shall leave the details to the reader. □

We can in fact describe the tensor product of R-algebras by a universal property. We will describe a commutative diagram:

$$\begin{tikzcd}
 & R \arrow[dl] \arrow[dr] & \\
R_0 \arrow[dr] & & R_1 \arrow[dl] \\
 & R_0 \otimes_R R_1 &
\end{tikzcd}$$

Here $R_0 \to R_0 \otimes_R R_1$ sends $x \mapsto x \otimes 1$; similarly for $R_1 \mapsto R_0 \otimes_R R_1$. These are ring-homomorphisms, and it is easy to see that the above diagram commutes, since $r \otimes 1 = 1 \otimes r = r(1 \otimes 1)$ for $r \in R$. In fact,

4.3.20 Proposition *$R_0 \otimes_R R_1$ is universal with respect to this property: in the language of category theory, the above diagram is a pushout square.*

This means for any commutative ring B, and every pair of maps $u_0 : R_0 \to B$ and $u_1 : R_1 \to B$ such that the pull-backs $R \to R_0 \to B$ and $R \to R_1 \to B$ are the same, then we get a unique map of rings

$$R_0 \otimes_R R_1 \to B$$

which restricts on R_0, R_1 to the morphisms u_0, u_1 that we started with.

Proof. If B is a ring as in the previous paragraph, we make B into an R-module by the map $R \to R_0 \to B$ (or $R \to R_1 \to B$, it is the same by assumption). This map $R_0 \otimes_R R_1 \to B$ sends

$$x \otimes y \to u_0(x)u_1(y).$$

It is easy to check that $(x, y) \to u_0(x)u_1(y)$ is R-bilinear (because of the condition that the two pull-backs of u_0, u_1 to R are the same), and that it gives a homomorphism of rings $R_0 \otimes_R R_1 \to B$ which restricts to u_0, u_1 on R_0, R_1. One can check, for instance, that this is a homomorphism of rings by looking at the generators.

It is also clear that $R_0 \otimes_R R_1 \to B$ is unique, because we know that the map on elements of the form $x \otimes 1$ and $1 \otimes y$ is determined by u_0, u_1; these generate $R_0 \otimes_R R_1$, though. □

In fact, we now claim that the category of rings has *all* coproducts. We see that the coproduct of any two elements exists (as the tensor product over $\mathbb{Z}$). It turns out that arbitrary coproducts exist. More generally, if $\{R_\alpha\}$ is a family of R-algebras, then one can define an object

$$\bigotimes_\alpha R_\alpha,$$

which is a coproduct of the R_α in the category of R-algebras. To do this, we simply take the generators as before, as formal objects

$$\bigotimes r_\alpha, \quad r_\alpha \in R_\alpha,$$

except that all but finitely many of the r_α are required to be the identity. One quotients by the usual relations.

Alternatively, one may use the fact that filtered colimits exist, and construct the infinite coproduct as a colimit of finite coproducts (which are just ordinary tensor products).

4.4. Exactness properties of the tensor product

In general, the tensor product is not exact; it is only exact on the right, but it can fail to preserve injections. Yet in some important cases it *is* exact. We study that in the present section.

Right-exactness of the tensor product

We will start by talking about extent to which tensor products do preserve exactness under any circumstance. First, let's recall what is going on. If M, N are R-modules over the commutative ring R, we have defined another R-module $\hom_R(M, N)$ of morphisms $M \to N$. This is left-exact as a functor of N. In other words, if we fix M and let N vary, then the construction of homming out of M preserves kernels.

In the language of category theory, this construction $N \to \hom_R(M, N)$ has an adjoint. The other construction we discussed last time was this adjoint, and it is the tensor product. Namely, given M, N we defined a **tensor product** $M \otimes_R N$ such that giving a map $M \otimes_R N \to P$ into some R-module P is the same as giving a bilinear map $\lambda : M \times N \to P$, which in turn is the same as giving an R-linear map

$$M \to \hom_R(N, P).$$

So we have a functorial isomorphism

$$\hom_R(M \otimes_R N, P) \simeq \hom_R(M, \hom_R(N, P)).$$

Alternatively, tensoring is the left-adjoint to the hom functor. By abstract nonsense, it follows that since $\hom(M, \cdot)$ preserves cokernels, the left-adjoint preserves cokernels and is right-exact. We shall see this directly.

4.4.1 Proposition *The functor $N \to M \otimes_R N$ is right-exact, i.e. preserves cokernels.*

In fact, the tensor product is symmetric, so it's right exact in either variable.

Proof. We have to show that if $N' \to N \to N'' \to 0$ is exact, then so is

$$M \otimes_R N' \to M \otimes_R N \to M \otimes_R N'' \to 0.$$

There are a lot of different ways to think about this. For instance, we can look at the direct construction. The tensor product is a certain quotient of a free module.

$M \otimes_R N''$ is the quotient of the free module generated by $m \otimes n'', m \in M, n \in N''$ modulo the usual relations. The map $M \otimes N \to M \otimes N''$ sends $m \otimes n \to m \otimes n''$ if n'' is the image of n in N''. Since each n'' can be lifted to some n, it is obvious that the map $M \otimes_R N \to M \otimes_R N''$ is surjective.

Now we know that $M \otimes_R N''$ is a quotient of $M \otimes_R N$. But which relations do you have to impose on $M \otimes_R N$ to get $M \otimes_R N''$? In fact, each relation in $M \otimes_R N''$ can be lifted to a relation in $M \otimes_R N$, but with some redundancy. So the only thing to quotient out by is the statement that $x \otimes y, x \otimes y'$ have the same image in $M \otimes N''$. In particular, we have to quotient out by

$$x \otimes y - x \otimes y' \ , y - y' \in N'$$

so that if we kill off $x \otimes n'$ for $n' \in N' \subset N$, then we get $M \otimes N''$. This is a direct proof.

One can also give a conceptual proof. We would like to know that $M \otimes N''$ is the cokernel of $M \otimes N' \to M \otimes N''$. In other words, we'd like to know that if we mapped $M \otimes_R N$ into some P and the pull-back to $M \otimes_R N'$, it'd factor uniquely through $M \otimes_R N''$. Namely, we need to show that

$$\hom_R(M \otimes N'', P) = \ker(\hom_R(M \otimes N, P) \to \hom_R(M \otimes N'', P)).$$

But the first is just $\hom_R(N'', \hom_R(M, P))$ by the adjointness property. Similarly, the second is just

$$\ker(\hom_R(N, \hom(M, P)) \to \hom_R(N', \hom_R(M, P))$$

but this last statement is $\hom_R(N'', \hom_R(M, P))$ by just the statement that $N'' = \operatorname{coker}(N' \to N)$. To give a map N'' into some module (e.g. $\hom_R(M, P)$) is the same thing as giving a map out of N which kills N'. So we get the functorial isomorphism. $\square$

4.4.2 Remark Formation of tensor products is, in general, **not** exact.

4.4.3 Example Let $R = \mathbb{Z}$. Let $M = \mathbb{Z}/2\mathbb{Z}$. Consider the exact sequence

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$$

which we can tensor with M, yielding

$$0 \to \mathbb{Z}/2\mathbb{Z} \to \mathbb{Q} \otimes \mathbb{Z}/2\mathbb{Z} \to \mathbb{Q}/\mathbb{Z} \otimes \mathbb{Z}/2\mathbb{Z} \to 0$$

I claim that the second thing $\mathbb{Q} \otimes \mathbb{Z}/2\mathbb{Z}$ is zero. This is because by tensoring with $\mathbb{Z}/2\mathbb{Z}$, we've made multiplication by 2 identically zero. By tensoring with $\mathbb{Q}$, we've made multiplication by 2 invertible. The only way to reconcile this is to have the second term zero. In particular, the sequence becomes

$$0 \to \mathbb{Z}/2\mathbb{Z} \to 0 \to 0 \to 0$$

which is not exact.

4.4.4 Remark (exercise) Let R be a ring, $I, J \subset R$ ideals. Show that $R/I \otimes_R R/J \simeq R/(I+J)$.

A characterization of right-exact functors

Let us consider additive functors on the category of R-modules. So far, we know a very easy way of getting such functors: given an R-module N, we have a functor

$$T_N : M \to M \otimes_R N.$$

In other words, we have a way of generating a functor on the category of R-modules for each R-module. These functors are all right-exact, as we have seen. Now we will prove a converse.

4.4.5 Proposition *Let F be a right-exact functor on the category of R-modules that commutes with direct sums. Then F is isomorphic to some T_N.*

Proof. The idea is that N will be $F(R)$.

Without the right-exactness hypothesis, we shall construct a natural morphism

$$F(R) \otimes M \to F(M)$$

as follows. Given $m \in M$, there is a natural map $R \to M$ sending $1 \to m$. This identifies $M = \hom_R(R, M)$. But functoriality gives a map $F(R) \times \hom_R(R, M) \to F(M)$, which is clearly R-linear; the universal property of the tensor product now produces the desired transformation $T_{F(R)} \to F$.

It is clear that $T_{F(R)}(M) \to F(M)$ is an isomorphism for $M = R$, and thus for M free, as both $T_{F(R)}$ and F commute with direct sums. Now let M be any R-module. There is a "free presentation," that is an exact sequence

$$R^I \to R^J \to M \to 0$$

for some sets I, J; we get a commutative, exact diagram

$$\begin{tikzcd} T_{F(R)}(R^I) \arrow[r] \arrow[d] & T_{F(R)}(R^J) \arrow[r] \arrow[d] & T_{F(R)}(M) \arrow[r] \arrow[d] & 0 \\ F(R^I) \arrow[r] & F(R^J) \arrow[r] & F(M) \arrow[r] & 0 \end{tikzcd}$$

where the leftmost two vertical arrows are isomorphisms. A diagram chase now shows that $T_{F(R)}(M) \to F(M)$ is an isomorphism. In particular, $F \simeq T_{F(R)}$ as functors. $\square$

Without the hypothesis that F commutes with arbitrary direct sums, we could only draw the same conclusion on the category of *finitely presented* modules; the same proof as above goes through, though I and J are required to be finite.[6]

4.4.6 Proposition *Let F be a right-exact functor on the category of finitely presented R-modules that commutes with direct sums. Then F is isomorphic to some T_N.*

From this we can easily see that localization at a multiplicative subset $S \subset R$ is given by tensoring with $S^{-1}R$. Indeed, localization is a right-exact functor on the category of R-modules, so it is given by tensoring with some module M; applying to R shows that $M = S^{-1}R$.

Flatness

In some cases, though, the tensor product is exact.

4.4.7 Definition Let R be a commutative ring. An R-module M is called **flat** if the functor $N \to M \otimes_R N$ is exact. An R-algebra is **flat** if it is flat as an R-module.

We already know that tensoring with anything is right exact, so the only thing to be checked for flatness of M is that the operation of tensoring by M preserves injections.

4.4.8 Example $\mathbb{Z}/2\mathbb{Z}$ is not flat as a $\mathbb{Z}$-module by 4.4.3.

4.4.9 Example If R is a ring, then R is flat as an R-module, because tensoring by R is the identity functor.

More generally, if P is a projective module (i.e., homming out of P is exact), then P is flat.

Proof. If $P = \bigoplus_A R$ is free, then tensoring with P corresponds to taking the direct sum $|A|$ times, i.e.

$$P \otimes_R M = \bigoplus_A M.$$

This is because tensoring with R preserves (finite or direct) infinite sums. The functor $M \to \bigoplus_A M$ is exact, so free modules are flat.

[6] Recall that an additive functor commutes with finite direct sums.

A projective module, as discussed earlier, is a direct summand of a free module. So if P is projective, $P \oplus P' \simeq \bigoplus_A R$ for some P'. Then we have that

$$(P \otimes_R M) \oplus (P' \otimes_R M) \simeq \bigoplus_A M.$$

If we had an injection $M \to M'$, then there is a direct sum decomposition yields a diagram of maps

$$\begin{tikzcd} P \otimes_R M \arrow[r] \arrow[d] & \bigoplus_A M \arrow[d] \\ P \otimes_R M' \arrow[r] & \bigoplus_A M' \end{tikzcd}.$$

A diagram-chase now shows that the vertical map is injective. Namely, the composition $P \otimes_R M \to \bigoplus_A M'$ is injective, so the vertical map has to be injective too. ∎

4.4.10 Example If $S \subset R$ is a multiplicative subset, then $S^{-1}R$ is a flat R-module, because localization is an exact functor.

Let us make a few other comments.

4.4.11 Remark Let $\phi : R \to R'$ be a homomorphism of rings. Then, first of all, any R'-module can be regarded as an R-module by composition with ϕ. In particular, R' is an R-module.

If M is an R-module, we can define

$$M \otimes_R R'$$

as an R-module. But in fact this tensor product is an R'-module; it has an action of R'. If $x \in M$ and $a \in R'$ and $b \in R'$, multiplication of $(x \otimes a) \in M \otimes_R R'$ by $b \in R'$ sends this, *by definition*, to

$$b(x \otimes a) = x \otimes ab.$$

It is easy to check that this defines an action of R' on $M \otimes_R R'$. (One has to check that this action factors through the appropriate relations, etc.)

The following fact shows that the hom-sets behave nicely with respect to flat base change.

4.4.12 Proposition *Let M be a finitely presented R-module, N an R-module. Let S be a flat R-algebra. Then the natural map*

$$\hom_R(M, N) \otimes_R S \to \hom_S(M \otimes_R S, N \otimes_R S)$$

is an isomorphism.

Proof. Indeed, it is clear that there is a natural map

$$\hom_R(M, N) \to \hom_S(M \otimes_R S, N \otimes_R S)$$

of R-modules. The latter is an S-module, so the universal property gives the map $\hom_R(M, N) \otimes_R S \to \hom_S(M \otimes_R S, N \otimes_R S)$ as claimed. If N is fixed, then we have two contravariant functors in M,

$$T_1(M) = \hom_R(M, N) \otimes_R S, \quad T_2(M) = \hom_S(M \otimes_R S, N \otimes_R S).$$

We also have a natural transformation $T_1(M) \to T_2(M)$. It is clear that if M is *finitely generated* and *free*, then the natural transformation is an isomorphism (for example, if $M = R$, then we just have the map $N \otimes_R S \to N \otimes_R S$).

Note moreover that both functors are left-exact: that is, given an exact sequence

$$M' \to M \to M'' \to 0,$$

there are induced exact sequences

$$0 \to T_1(M'') \to T_1(M) \to T_1(M'), \quad 0 \to T_2(M'') \to T_2(M) \to T_2(M').$$

Here we are using the fact that hom is always a left-exact functor and the fact that tensoring with S preserves exactness. (Thus it is here that we use flatness.)

Now the following lemma will complete the proof:

4.4.13 Lemma *Let T_1, T_2 be contravariant, left-exact additive functors from the category of R-modules to the category of abelian groups. Suppose a natural transformation $t : T_1(M) \to T_2(M)$ is given, and suppose this is an isomorphism whenever M is finitely generated and free. Then it is an isomorphism for any finitely presented module M.*

Proof. This lemma is a diagram chase. Fix a finitely presented M, and choose a presentation

$$F' \to F \to M \to 0,$$

with F', F finitely generated and free. Then we have an exact and commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & T_1(M) \arrow[r] \arrow[d] & T_1(F) \arrow[r] \arrow[d, "\simeq"] & T_1(F') \arrow[d, "\simeq"] \\ 0 \arrow[r] & T_2(M) \arrow[r] & T_2(F) \arrow[r] & T_2(F'). \end{tikzcd}$$

□

By hypotheses, the two vertical arrows to the right are isomorphisms, as indicated. A diagram chase now shows that the remaining arrow is an isomorphism, which is what we wanted to prove. □

4.4.14 Example Let us now consider finitely generated flat modules over a principal ideal domain R. By 2.7.4, we know that any such M is isomorphic to a direct sum $\bigoplus R/a_i$ for some $a_i \in R$. But if any of the a_i is not zero, then that a_i would be a nonzero zero divisor on M. However, we know no element of $R - \{0\}$ can be a zero divisor on M. It follows that all the $a_i = 0$. In particular, we have proved:

4.4.15 Proposition *A finitely generated module over a PID is flat if and only if it is free.*

Finitely presented flat modules

In example 4.4.9, we saw that a projective module over any ring R was automatically flat. In general, the converse is flat. For instance, $\mathbb{Q}$ is a flat $\mathbb{Z}$-module (as tensoring by $\mathbb{Q}$ is a form of localization). However, because $\mathbb{Q}$ is divisible (namely, multiplication by n is surjective for any n), $\mathbb{Q}$ cannot be a free abelian group.

Nonetheless:

4.4.16 Theorem *A finitely presented flat module over a ring R is projective.*

Proof. We follow **?**.

Let us define the following contravariant functor from R-modules to R-modules. Given M, we send it to $M^* = \hom_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$. This is made into an R-module in the following manner: given $\phi : M \to \mathbb{Q}/\mathbb{Z}$ (which is just a homomorphism of abelian groups!) and $r \in R$, we send this to $r\phi$ defined by $(r\phi)(m) = \phi(rm)$. Since $\mathbb{Q}/\mathbb{Z}$ is an injective abelian group, we see that $M \mapsto M^*$ is an *exact* contravariant functor from R-modules to R-modules. In fact, we note that $0 \to A \to B \to C \to 0$ is exact implies $0 \to C^* \to B^* \to A^* \to 0$ is exact.

Let F be any R-module. There is a natural homomorphism

$$M^* \otimes_R F \to \hom_R(F, M)^*. \tag{4.4.1}$$

This is defined as follows. Given $\phi : M \to \mathbb{Q}/\mathbb{Z}$ and $x \in F$, we define a new map $\hom(F, M) \to \mathbb{Q}/\mathbb{Z}$ by sending a homomorphism $\psi : F \to M$ to $\phi(\psi(x))$. In other words, we have a natural map

$$\hom_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z}) \otimes_R F \to \hom_{\mathbb{Z}}(\hom_R(F, M)^*, \mathbb{Q}/\mathbb{Z}).$$

Now fix M. This map (4.4.1) is an isomorphism if F is *finitely generated* and free. Both are right-exact (because dualizing is contravariant-exact!). The "finite presentation trick" now shows that the map is an isomorphism if F is finitely presented. **add: this should be elaborated on**

Fix now F finitely presented and flat, and consider the above two quantities in (4.4.1) as functors in M. Then the first functor is exact, so the second one is too. In particular, $\hom_R(F, M)^*$ is an exact functor in M; in particular, if $M \twoheadrightarrow M''$ is a surjection, then

$$\hom_R(F, M'')^* \to \hom_R(F, M)^*$$

is an injection. But this implies that

$$\hom_R(F, M) \to \hom_R(F, M'')$$

is a *surjection,* i.e. that F is projective. Indeed:

4.4.17 Lemma *$A \to B \to C$ is exact if and only if $C^* \to B^* \to A^*$ is exact.*

Proof. Indeed, one direction was already clear (from $\mathbb{Q}/\mathbb{Z}$ being an injective abelian group). Conversely, we note that $M = 0$ if and only if $M^* = 0$ (again by injectivity and the fact that $(\mathbb{Z}/a)^* \neq 0$ for any a). Thus dualizing reflects isomorphisms: if a map becomes an isomorphism after dualized, then it was an isomorphism already. From here it is easy to deduce the result (by applying the above fact to the kernel and image). □

I.5. Algebras and their modules

5.1. The category of algebras over a commutative ring

5.1.1 In Section 2.1 we already got acquainted with unital algebras over a commutative ring. Here we want to generalize this concept to possibly non-unital algebras. Throughout this section, R will denote a commutative ring.

Definitions

5.1.2 Definition An R-module A together with a map $\mu : A \times A \to A$ called *multiplication* is an *R-algebra* if the following properties hold.

(Alg1) The multiplication map is *R-bilinear* that is the equalities

$$\begin{aligned} \mu(a_1 + a_2, b) &= \mu(a_1, b) + \mu(a_2, b) \ , \\ \mu(ra, b) &= r\mu(a, b) \ , \\ \mu(a, b_1 + b_2) &= \mu(a, b_1) + \mu(a, b_2) \ , \quad \text{and} \\ \mu(a, rb) &= r\mu(a, b) \end{aligned}$$

hold for all $a, a_1, a_2, b, b_1, b_2 \in A$ and $r \in R$.

(Alg2) The multiplication map is associative that is

$$\mu(a, \mu(b, c)) = \mu(\mu(a, b), c)$$

for all $a, b, c \in A$.

If in addition Axiom (Alg3) below is satisfied, the algebra A is called *unital*, if Axiom (Alg4) holds true, then A is called *commutative*.

(Alg3) There exists a multiplicative identity in A that is an element $1_A \in A$ such that

$$\mu(1_A, a) = \mu(a, 1_A) = a$$

for all $a \in A$.

(Alg4) Multiplication is commutative that is

$$\mu(a, b) = \mu(b, a)$$

for all $a, b \in A$.

5.1.3 Remarks (a) Given an R-algebra A we usually denote the *product* $\mu(a,b)$ of two elements $a,b \in A$ by $a \cdot_A b$, $a \cdot b$ or just by ab.

(b) Since an R-algebra A is also a pseudo-ring, a multiplicative identity is uniquely determined by Proposition 2.1.2 (iv). If it exists and when no confusion can arise we usually denote the multiplicative identity in A just by the symbol 1.

(c) As pointed out at the beginning of this chapter, we already have a notion of a unital R-algebra. That the one from Definition 2.1.11 is equivalent to the definition of a unital R-algebra above is shown by the following result.

5.1.4 Proposition *Assume that A is a ring with multiplication $\mu : A \times A \to A$, $(a,b) \mapsto \mu(a,b) = a \cdot b$ and multiplicative identity 1_A. Then the following holds true.*

(i) *If $\varphi : R \to A$ a ring homomorphism with image in the center of A, then the action $R \times A \to A$, $(r,a) \mapsto \varphi(r) \cdot a$ gives A the structure of an R-module such that axiom* (Alg1) *is satisfied.*

(ii) *If A carries an R-module structure $R \times A \to A$, $(r,a) \mapsto ra$ which satisfies axiom* (Alg1)*, then the map $\varphi : R \to A$, $r \mapsto r \cdot 1_A$ is a ring homomorphism with image in the center of A.*

Proof. Assume that $\varphi : R \to A$ is a ring homomorphism with image in $\mathrm{Z}(A)$. Since φ is a ring homomorphism and since the left and right distributivity laws hold in A one has for all $r, r_1, r_2, s \in R$ and $a, a_1, a_2 \in A$

$$
\begin{aligned}
(rs)a &= \varphi(rs) \cdot a = (\varphi(r) \cdot \varphi(s)) \cdot a = \varphi(r) \cdot (\varphi(s) \cdot a) = \varphi(r) \cdot (sa) = r(sa) \ , \\
1_R a &= \varphi(1_R) \cdot a = 1_A \cdot a = a \ , \\
(r_1 + r_2)a &= \varphi(r_1 + r_2) \cdot a = (\varphi(r_1) + \varphi(r_2)) \cdot a = \varphi(r_1) \cdot a + \varphi(r_2) \cdot a = ra_1 + r_2 a \ , \\
r(a_1 + a_2) &= \varphi(r) \cdot (a_1 + a_2) = \varphi(r) \cdot a_1 + \varphi(r) \cdot a_2 = ra_1 + ra_2 \ .
\end{aligned}
$$

Hence A becomes an R-module as claimed. That multiplication on A is R-bilinear follows from the equalities

$$
\begin{aligned}
& (ra) \cdot b = (\varphi(r) \cdot a) \cdot b = \varphi(r) \cdot (a \cdot b) = r(a \cdot b) \ , \quad \text{and} \\
& a \cdot (rb) = a \cdot (\varphi(r) \cdot b) = (a \cdot \varphi(r)) \cdot b = (\varphi(r) \cdot a) \cdot b = \varphi(r) \cdot (a \cdot b) = r(a \cdot b) \ .
\end{aligned}
$$

Note that hereby we used that φ has image in the center of A. So (i) is proved.

To verify (ii) assume A to be an R-module with R-bilinear multiplication. Define $\varphi : R \to A$ by $\varphi(r) = r1_A$. Then

$$
\begin{aligned}
\varphi(r + s) &= (r + s)1_A = r1_A + s1_A = \varphi(r) + \varphi(s) \ , \\
\varphi(rs) &= (rs)1_A = r(s1_A) = r\varphi(s) = r(1_A \cdot \varphi(s)) = (r1_A) \cdot \varphi(s) = \varphi(r) \cdot \varphi(s) \ , \\
\varphi(1_R) &= 1_R 1_A = 1_A \ ,
\end{aligned}
$$

so φ is a ring homomorphism. By R-bilinearity of μ, the image of φ lies in the center of A. Namely, for all $r \in R$ and $a \in A$

$$
\varphi(r) \cdot a = r1_A \cdot a = r(1_A \cdot a) = r(a \cdot 1_A) = a \cdot (r1_A) = a \cdot \varphi(r) \qquad \square
$$

5.2. Tensor, symmetric, and exterior algebras

The tensor algebra of an R-module

5.2.1 Let M be a module over a commutative ring. Recall that for every $n \in \mathbb{N}$ the *n-th tensor power* $M^{\otimes n}$ of M is defined as the (up to isomorphisms) unique R-module such that the map

$$\epsilon^* : \mathrm{Hom}_{R\text{-ml}}(M^{\times n}, N) \to \mathrm{Hom}_R(M^{\otimes n}, N)$$

from the n-multilinear maps between $M^{\times n}$ and N to the R-linear maps between $M^{\otimes n}$ and N is a bijection, where $\epsilon_n : M^{\times n} \to M$ is the unit. In other words, $M^{\otimes n}$ is the tensor product of n copies of M. Sometimes we will denote $M^{\otimes n}$ by $\mathsf{T}^n M$. Now form the direct sum

$$\mathsf{T}^\bullet M = \bigoplus_{n\in\mathbb{N}} \mathsf{T}^n M = \bigoplus_{n\in\mathbb{N}} M^{\otimes n}$$

and define on it the binary operation

$$\mu : \mathsf{T}^\bullet M \times \mathsf{T}^\bullet M \to \mathsf{T}^\bullet M, \quad ((v_k)_{k\in\mathbb{N}}, (w_l)_{l\in\mathbb{N}}) \mapsto \left(\sum_{k+l=n} v_k \otimes w_l \right)_{n\in\mathbb{N}} .$$

The symmetric algebra of an R-module

5.2.2 Given an R-module M consider the (two-sided) ideal $\mathsf{I}_\mathrm{s}^\bullet \subset \mathsf{T}^\bullet M$ generated by the set of all elements of the form $v \otimes w - w \otimes v$, where v, w run through the elements of M. That ideal can be expressed in the form

$$\mathsf{I}_\mathrm{s}^\bullet = \mathrm{Span}_R \left\{ a \otimes v \otimes w \otimes b - a \otimes w \otimes v \otimes b \in \mathsf{T}^\bullet M \;\middle|\; v, w \in M \,\&\, a, b \in \mathsf{T}^\bullet M \right\} .$$

Obviously, the right hand side is the direct sum of the R-modules

$$\mathsf{I}_\mathrm{s}^n = \begin{cases} \{0\} & \text{for } n = 0, 1 \,, \\ \mathrm{Span}_R \left\{ v \otimes w - w \otimes v \in \mathsf{T}^n M \;\middle|\; v, w \in M \right\} & \text{for } n = 2 \,, \text{ and} \\ \mathrm{Span}_R \big\{ a \otimes v \otimes w \otimes b - a \otimes w \otimes v \otimes b \in \mathsf{T}^n M \;\big| & \\ \qquad v, w \in M,\; a \in \mathsf{T}^k M,\; b \in \mathsf{T}^l M,\; \&\, k + l = n \big\} & \text{for } n > 2 \,, \end{cases}$$

which means that $\mathsf{I}_\mathrm{s}^\bullet$ is a graded ideal. Hence the quotient algebra

$$\mathsf{S}^\bullet M = \mathsf{T}^\bullet M / \mathsf{I}_\mathrm{s}^\bullet$$

is a unital graded algebra which is called the *symmetric algebra* of M. Moreover, $\mathsf{S}^0 M$ coincides with M by construction, so one has a natural embedding $\varepsilon : M \hookrightarrow \mathsf{S}^\bullet M$. These data satisfy the following universal property.

5.2.3 Proposition *Let M be an R-module over the commutative ring R, A a commutative unital R-algebra and $f : M \to A$ an R-module map. Then there exists a unique morphism of unital R-algebras $\overline{f} : \mathsf{S}^\bullet M \to A$ such that the following diagram commutes.*

Part II.

Fundamentals of Topology and Geometry

II.1. General topology

1.1. The category of topological spaces

Topologies and continuous maps

1.1.1 Definition Let X be a set. By a topology on X on understands a set $\mathcal{T}$ of subsets of X such that:

(Top0) The sets X and $\varnothing$ are both elements of $\mathcal{T}$.

(Top1) The union of any collection of elements of $\mathcal{T}$ is again in $\mathcal{T}$ that means if $(U_i)_{i\in I}$ is a family of elements $U_i \in \mathcal{T}$, then $\bigcup_{i\in I} U_i \in \mathcal{T}$.

(Top2) The intersection of finitely many elements of $\mathcal{T}$ is again in $\mathcal{T}$ that means for every natural n and $U_1, \ldots, U_n \in \mathcal{T}$ one has $\bigcap_{i=1}^n U_i \in \mathcal{T}$.

A pair $(X, \mathcal{T})$ is a called a *topological space* when X is a set and $\mathcal{T}$ a topology on X. Moreover, a subset U of X is called *open* if $U \in \mathcal{T}$ and *closed* if $\complement_X U \in \mathcal{T}$.

1.1.2 Remarks (a) Strictly speaking, Axiom (Top0) can be derived from Axioms (Top1) and (Top2), since the union of an empty family of subsets of X coincides with $\varnothing$, and the intersection of an empty family of subsets of X coincides with X. Nevertheless, it is useful to require it, since in proofs one often shows Axiom (Top1) only for non-empty families of open sets, and Axiom (Top2) only for the case of the intersection of two open subsets. Then it is necessary to verify Axiom (Top0), too, when one wants to prove that a given set of subsets of X is a topology.

(b) When using the notation $\mathcal{T}_X$ for a topology we always mean that $\mathcal{T}_X$ is a topology on the space X.

1.1.3 Examples (a) For every set X the power set $\mathcal{P}(X)$ is a topology on X. It is called the *discrete* or *strongest* topology on X.

(b) The set $\{\varnothing, X\}$ is another topology on a set X called the *indiscrete* or *trivial* or *weakest* topology on X. Unless X is empty or has only one element, the discrete and indiscrete topologies differ.

(c) Let S be a set $\{0,1\}$. Then the set $\{\varnothing, \{1\}, \{0,1\}\}$ is a topology on S which does neither coincide with the discrete nor the indiscrete topology. The set S with this topology is called *Sierpiński space*. The closed sets of the Sierpiński space are $\varnothing$, $\{0\}$ and S.

(d) The *standard topology* on the set of real numbers $\mathbb{R}$ consists of all subsets $U \subset \mathbb{R}$ such that for each $x \in U$ there are real numbers a, b satisfying $a < x < b$ and $(a, b) \subset U$. The standard topology on $\mathbb{R}$ will be denoted by $\mathcal{T}_{\mathbb{R}}$.

Let us show that $\mathcal{T}_{\mathbb{R}}$ is a topology on $\mathbb{R}$ indeed. Obviously $\varnothing$ and $\mathbb{R}$ are elements of $\mathcal{T}_{\mathbb{R}}$. Let $U, V \in \mathcal{T}_{\mathbb{R}}$ and $x \in U \cap V$. Then there are $a, b, c, d \in \mathbb{R}$ such that $x \in (a, b) \subset U$ and $x \in (c, d) \subset V$. Put $e := \max\{a, c\}$ and $f := \min\{b, d\}$. Then $x \in (e, f) \subset U \cap V$, which proves $U \cap V \in \mathcal{T}_{\mathbb{R}}$. If $(U_i)_{i \in I}$ is a family of elements $U_i \in \mathcal{T}_{\mathbb{R}}$ and $x \in \bigcup_{i \in I} U_i$, then there exists an $j \in I$ with $x \in U_j$. Choose $a, b \in \mathbb{R}$ such that $x \in (a, b) \subset U_j$. Then $x \in (a, b) \subset \bigcup_{i \in I} U_i$, which proves $\bigcup_{i \in I} U_i \in \mathcal{T}_{\mathbb{R}}$. If not mentioned differently, we always assume the set of real numbers to be equipped with the standard topology. The standard topology coincides with the metric topology induced by the euclidean metric on $\mathbb{R}$, see **??**. One therefore often calls $\mathcal{T}_{\mathbb{R}}$ the *euclidean topology* on $\mathbb{R}$. We will use these terms interchangeably.

(e) The *standard topology* $\mathcal{T}_{\mathbb{Q}}$ on the set of rational numbers $\mathbb{Q}$ is defined analogously. It consists of all subset $U \subset \mathbb{Q}$ such that for each $x \in U$ there exist rational numbers a, b with $a < x < b$ and $(a, b) \subset U$. Like for the reals one proves that $\mathcal{T}_{\mathbb{Q}}$ is a topology on $\mathbb{Q}$. Unless mentioned differently it is always assumed that $\mathbb{Q}$ comes equipped with the standard topology. Like for $\mathbb{R}$, the standard topology on $\mathbb{Q}$ coincides with the *euclidean topology* on $\mathbb{Q}$ which is the one induced by the euclidean metric.

(f) Let X be a set, and let $\mathcal{T}_{\text{cof}}$ denote the set of all subset of X which are either empty or have finite complement in X. Then $\mathcal{T}_{\text{cof}}$ is a topology on X called the *cofinite topology*.

(g) Let X be a set, and let $\mathcal{T}_{\text{coc}}$ denote the set of all subset of X which are either empty or have countable complement in X. Then $\mathcal{T}_{\text{coc}}$ is a topology on X called the *cocountable topology*.

(h) Let X be a (nonempty) set, $(Y, \mathcal{T})$ be a topological space, and $f : X \to Y$ a function. Define

$$f^*\mathcal{T} := f^{-1}\mathcal{T} := \{f^{-1}(U) \in \mathcal{P}(X) \mid U \in \mathcal{T}\} .$$

Then $(X, f^*\mathcal{T})$ is a topological space. One calls $f^*\mathcal{T}$ the *initial topology on X with respect to f* or the *topology on X induced by f*.

Let us verify that $f^*\mathcal{T}$ is a topology on X indeed. By $f^{-1}(Y) = X$ and $f^{-1}(\varnothing) = \varnothing$ the sets X and $\varnothing$ are in $f^*\mathcal{T}$. Now let $(V_i)_{i \in I}$ be a family of elements of $f^*\mathcal{T}$. In other words we have, for each $i \in I$, $V_i = f^{-1}(U_i)$ for some $U_i \in \mathcal{T}$. Then $U := \bigcup_{i \in I} U_i \in \mathcal{T}$ and

$$\bigcup_{i \in I} V_i = \bigcup_{i \in I} f^{-1}(U_i) = f^{-1}\Big(\bigcup_{i \in I} U_i\Big) = f^{-1}(U) \in f^*\mathcal{T} .$$

Finally, let $V_1, \ldots, V_n \in f^{-1}\mathcal{T}$. Then, by definition, there exist $U_1, \ldots, U_n \in \mathcal{T}$ such that $V_i = f^{-1}(U_i)$ for $i = 1, \ldots, n$. Thus $U := \bigcap_{i=1}^n U_i \in \mathcal{T}$ and

$$\bigcap_{i=1}^n V_i = \bigcap_{i=1}^n f^{-1}(U_i) = f^{-1}\Big(\bigcap_{i=1}^n U_i\Big) = f^{-1}(U) \in f^*\mathcal{T} .$$

(i) Let $(X, \mathcal{T})$ be a topological space, Y a (nonempty) set, and $g : X \to Y$ a function. Define $g_*\mathcal{T} \subset \mathcal{P}(Y)$ as the set of all $U \subset Y$ such that $g^{-1}(U) \in \mathcal{T}$. Then $g_*\mathcal{T}$ is a topology on Y. It is called the *final topology on Y with respect to g* or the *topology on Y induced by g*. If $g : X \to Y$

is a *quotient map* that means that g is surjective, then the final topology on Y induced by g is also called the *quotient topology on X induced by g.*

Let us show why $g_*\mathcal{T}$ is a topology on Y. Obviously, $Y, \varnothing \in g_*\mathcal{T}$. Let $(U_i)_{i\in I}$ be a family of elements of $g_*\mathcal{T}$. Then $g^{-1}(U_i) \in \mathcal{T}$ for all $i \in I$ which entails

$$g^{-1}\Big(\bigcup_{i\in I} U_i\Big) = \bigcup_{i\in I} g^{-1}(U_i) \in \mathcal{T},$$

hence $\bigcup\limits_{i\in I} U_i \in g_*\mathcal{T}$. If $U_1, \dots U_k \in g_*\mathcal{T}$, then

$$g^{-1}(U_1 \cap \ldots \cap U_k) = \bigcap_{i=1}^{k} g^{-1}(U_i) \in \mathcal{T}.$$

So $U_1 \cap \ldots \cap U_k \in g_*\mathcal{T}$ and the claim is proved.

1.1.4 Section 1.2 on fundamental examples collects several more examples of topologies. For now, we will work out a few basic properties of topologies and their structure preserving morphisms, the continuous maps defined below.

1.1.5 Definition Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be two topological spaces and assume that $f : X \to Y$ is a function. One says that f is *continuous* if for all $U \in \mathcal{T}_Y$ the preimage $f^{-1}(U)$ is open in X. The map f is called *open* if $f(V)$ is open in Y for all $V \in \mathcal{T}_X$.

1.1.6 Example Any constant function $c : X \to Y$ between two topological spaces is continuous since the preimage of an open set in Y is either the full set X or empty depending on whether the image of c is contained in the open set or not.

1.1.7 Theorem and Definition (a) *The identity map* id_X *on a topological space* $(X, \mathcal{T}_X)$ *is continuous and open.*

(b) *Let* $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ *and* $(Z, \mathcal{T}_Z)$ *be three topological spaces. Assume that* $f : X \to Y$ *and* $g : Y \to Z$ *are maps. If* f *and* g *are both continuous, so is* $g \circ f$. *If* f *and* g *are both open, then* $g \circ f$ *is open as well.*

(c) *Topological spaces as objects together with continuous maps as morphisms form a category. It is called the* category of topological spaces *and will be denoted by* Top.

Proof. It is obvious by definition that the identity map id_X is continuous and open. Now assume that f and g are continuous and let $U \in \mathcal{T}_Z$. Then $g^{-1}(U) \in \mathcal{T}_Y$ by continuity of g. Hence $f^{-1}(g^{-1}(U)) \in \mathcal{T}_X$ by continuity of f. So $g \circ f$ is continuous. If f and g are open maps, and $V \in \mathcal{T}_X$, then $f(V) \in \mathcal{T}_Y$ and $g \circ f(V) = g(f(V)) \in \mathcal{T}_Z$. Hence the composition of two open maps is open, too. The rest of the claim follows immediately. $\square$

Comparison of topologies

1.1.8 The initial topology $f^*\mathcal{T}_Y$ induced by a function $f : X \to Y$ between topological spaces is a subset of the topology on X if and only if f is continuous. This motivates the following definition.

1.1.9 Definition Let X be a set. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. One says that $\mathcal{T}_1$ is *finer* or *stronger* than $\mathcal{T}_2$ and $\mathcal{T}_2$ is *coarser* or *weaker* than $\mathcal{T}_1$ when $\mathcal{T}_2 \subset \mathcal{T}_1$.

1.1.10 Of course, inclusion induces an order relation on topologies on a given set. A remarkable property is that any nonempty subset of the ordered set of topologies on a given set always admits a greatest lower bound.

1.1.11 Theorem *Let X be a set. Let $\mathfrak{S}$ be a nonempty set of topologies on E. Then the set*

$$\mathcal{T}_{\mathfrak{S}} := \bigcap_{\mathcal{T} \in \mathfrak{S}} \mathcal{T} = \left\{ U \in \mathcal{P}(X) \mid U \in \mathcal{T} \text{ for all } \mathcal{T} \in \mathfrak{S} \right\}$$

is a topology on X and it is the greatest lower bound of $\mathfrak{S}$, where the order between topologies is given by inclusion. In other words, $\mathcal{T}_{\mathfrak{S}}$ is the finest topology contained in each topology from $\mathfrak{S}$.

Proof. We first show that $\mathcal{T}_{\mathfrak{S}}$ is a topology. Since each $\mathcal{T} \in \mathfrak{S}$ is a topology on X, we have $\emptyset, X \in \mathcal{T}$ for all $\mathcal{T} \in \mathfrak{S}$. Hence $\emptyset, X \in \mathcal{T}_{\mathfrak{S}}$.

Let $(U_i)_{i \in I}$ be a nonempty family of elements $U_i \in \mathcal{T}_{\mathfrak{S}}$. Let $\mathcal{T} \in \mathfrak{S}$ be arbitrary. By definition of $\mathcal{T}_{\mathfrak{S}}$, we have $U_i \in \mathcal{T}$ for all $i \in I$. Since $\mathcal{T}$ is a topology, $\bigcup_{i \in I} U_i \in \mathcal{T}$. Hence, as $\mathcal{T}$ was arbitrary, $\bigcup_{i \in I} U_i \in \mathcal{T}_{\mathfrak{S}}$.

Now, let $U_1, \ldots, U_n \in \mathcal{T}_{\mathfrak{S}}$. Let $\mathcal{T} \in \mathfrak{S}$ be arbitrary. By definition of $\mathcal{T}_{\mathfrak{S}}$, we have $U_1, \ldots, U_n \in \mathcal{T}$. Therefore, $U_1 \cap \ldots \cap U_n \in \mathcal{T}$ since $\mathcal{T}$ is a topology. Since $\mathcal{T}$ was arbitrary in $\mathfrak{S}$, we conclude that $U_1 \cap \ldots \cap U_n \in \mathcal{T}_{\mathfrak{S}}$ by definition.

So $\mathcal{T}_{\mathfrak{S}}$ is a topology on X. By construction, $\mathcal{T}_{\mathfrak{S}} \subset \mathcal{T}$ for all $\mathcal{T} \in \mathfrak{S}$, so $\mathcal{T}_{\mathfrak{S}}$ is a lower bound for $\mathfrak{S}$. Assume given a new topology $\mathcal{Q}$ on X such that $\mathcal{Q} \subset \mathcal{T}$ for all $\mathcal{T} \in \mathfrak{S}$. Let $U \in \mathcal{Q}$. Then we have $U \in \mathcal{T}$ for all $\mathcal{T} \in \mathfrak{S}$. Hence by definition $U \in \mathcal{T}_{\mathfrak{S}}$. So $\mathcal{Q} \subset \mathcal{T}_{\mathfrak{S}}$ and thus $\mathcal{T}_{\mathfrak{S}}$ is the greatest lower bound of $\mathfrak{S}$. □

1.1.12 Corollary *Let X be a set, $(Y, \mathcal{T})$ be a topological space, and $f : X \to Y$ a map. The coarsest topology on X which makes f continuous is the initial topology $f^*\mathcal{T}$.*

Proof. Let $\mathfrak{S}$ be the set of all topologies on X such that f is continuous. By definition, $f^*\mathcal{T}$ is a lower bound of $\mathfrak{S}$. Moreover, $f^*\mathcal{T} \in \mathfrak{S}$. Hence $f^*\mathcal{T}$ is the coarsest topology making the function $f : X \to Y$ continuous. □

1.1.13 Proposition *Let $(X, \mathcal{T})$ be a topological space, Y a set, and $g : X \to Y$ a map. The finest topology on Y which makes g continuous is the final topology $g_*\mathcal{T}$.*

Proof. Let $\mathcal{S}$ be a topology on Y so that $g : (X, \mathcal{T}) \to (Y, \mathcal{S})$ is continuous. Let $U \in \mathcal{S}$. Then $g^{-1}(U) \in \mathcal{T}$ by continuity of $g : (X, \mathcal{T}) \to (Y, \mathcal{S})$. Hence $U \in g_*\mathcal{T}$ by definition, and $\mathcal{S} \subset \mathcal{T}$. Since $g : (X, \mathcal{T}) \to (Y, g_*\mathcal{T})$ is continuous by definition, the claim follows. □

1.1.14 We can use Theorem 1.1.11 to define other interesting topologies. Note that trivially $\mathcal{P}(X)$ is a topology on a given set X, so given any $\mathcal{S} \subset \mathcal{P}(X)$ there is at least one topology containing $\mathcal{S}$. From this:

1.1.15 Proposition and Definition *Let X be a set, and $\mathcal{S}$ a subset of $\mathcal{P}(X)$. The greatest lower bound of the set*

$$\mathfrak{S} = \{\mathcal{T} \in \mathcal{P}(\mathcal{P}(X)) \mid \mathcal{T} \textit{ is a topology on } X \,\&\, \mathcal{S} \subset \mathcal{T}\}$$

is the coarsest topology on X containing $\mathcal{S}$. We call it the topology generated by $\mathcal{S}$ on X *and denote it by $\mathcal{T}_\mathcal{S}$. The topology $\mathcal{T}_\mathcal{S}$ consists of unions of finite intersections of elements of $\mathcal{S}$ that means*

$$\mathcal{T}_\mathcal{S} = \left\{ U \in \mathcal{P}(X) \mid \exists J \,\forall j \in J \,\exists n_j \in \mathbb{N} \,\exists U_{j,1}, \ldots, U_{j,n_j} \in \mathcal{S} : \, U = \bigcup_{j \in J} \bigcap_{k=1}^{n_j} U_{j,k} \right\} .$$

Proof. By definition of $\mathfrak{S}$ and Theorem 1.1.11, $\mathcal{T}_\mathfrak{S} = \bigcap_{\mathcal{T} \in \mathfrak{S}} \mathcal{T}$ is a topology on X which contains $\mathcal{S}$. Hence $\mathcal{T}_\mathfrak{S}$ is an element of $\mathfrak{S}$ and a subset of any element of $\mathfrak{S}$. The first claim follows. To verify the second, observe that it suffices to show that

$$\mathcal{R} := \left\{ U \in \mathcal{P}(X) \mid \exists J \,\forall j \in J \,\exists n_j \in \mathbb{N} \,\exists U_{j,1}, \ldots, U_{j,n_j} \in \mathcal{S} : \, U = \bigcup_{j \in J} \bigcap_{k=1}^{n_j} U_{j,k} \right\}$$

is a topology. The set $\mathcal{R}$ being a topology namely entails $\mathcal{T}_\mathcal{S} \subset \mathcal{R}$ because $\mathcal{S} \subset \mathcal{R}$. The inclusion $\mathcal{R} \subset \mathcal{T}_\mathcal{S}$ is clear by definition, since $\mathcal{T}_\mathcal{S}$ is a topology containing $\mathcal{S}$. So let us show that $\mathcal{R}$ is a topology. Obviously $\varnothing$ and X are elements of $\mathcal{R}$ because $\bigcup_{i \in \varnothing} U_i = \varnothing$ and $\bigcap_{k=1}^{0} U_k = X$. Now assume that $(U_i)_{i \in I}$ is a family of elements of $\mathcal{R}$. Then there exists for each $i \in I$ a set J_i and for every $j \in J_i$ a natural number $n_{i,j}$ together with elements $U_{i,j,1}, \ldots, U_{i,j,n_{i,j}} \in \mathcal{S}$ such that

$$U_i = \bigcup_{j \in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} .$$

Put $J := \bigcup_{i \in I} \{i\} \times J_i$. Then

$$U := \bigcup_{i \in I} U_i = \bigcup_{i \in I} \bigcup_{j \in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} = \bigcup_{(i,j) \in J} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} \in \mathcal{R} .$$

Last assume $U_1, \ldots U_n \in \mathcal{T}$ where $n \in \mathbb{N}$. Then one can find for each $i \in \{1, \ldots, n\}$ a set J_i and for every $j \in J_i$ a natural number $n_{i,j}$ together with elements $U_{i,j,1}, \ldots, U_{i,j,n_{i,j}} \in \mathcal{S}$ such that

$$U_i = \bigcup_{j \in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} .$$

Put $J := J_1 \times \ldots \times J_n$. Then

$$U := \bigcap_{i=1}^{n} U_i = \bigcap_{i=1}^{n} \bigcup_{j \in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} = \bigcup_{(j_1, \ldots, j_n) \in J} \bigcap_{k_1=1}^{n_{1,j_1}} U_{1,j_1,k_1} \cap \ldots \cap \bigcap_{k_n=1}^{n_{n,j_n}} U_{n,j_n,k_n} \in \mathcal{R} .$$

Hence $\mathcal{R}$ is a topology, indeed, and the proposition is proved. □

1.1.16 Definition Let X be a set, and $\mathcal{T}$ a topology on X. One calls a subset $\mathcal{S} \subset \mathcal{T}$ a *subbase* (or *subbasis*) of the topology if $\mathcal{T}$ coincides with $\mathcal{T}_\mathcal{S}$. If in addition $X = \bigcup_{S \in \mathcal{S}} S$, the subbase $\mathcal{S}$ is said to be *adequate*.

Bases of topologies

1.1.17 When inducing a topology from a family $\mathcal{B}$ of subsets of some set X, the fact that $\mathcal{B}$ enjoys the following property greatly simplifies the description of the topology $\mathcal{T}_{\mathcal{B}}$ generated by $\mathcal{B}$.

1.1.18 Definition Let X be a set. A *base* (or *basis*) on X is a subset $\mathcal{B}$ of the powerset $\mathcal{P}(X)$ such that

(Bas1) $X = \bigcup_{B \in \mathcal{B}} B$,

(Bas2) For all $B_1, B_2 \in \mathcal{B}$ and all $x \in B_1 \cap B_2$ there exists a $B \in \mathcal{B}$ such that $x \in B$ and $B \subset B_1 \cap B_2$.

The main purpose for this definition stems from the following theorem:

1.1.19 Theorem *Let X be some set. Let $\mathcal{B}$ be a base on X. Then the topology generated by $\mathcal{B}$ coincides with the set of unions of elements of $\mathcal{B}$ that means*

$$\mathcal{T}_{\mathcal{B}} = \left\{ \bigcup_{B \in \mathcal{U}} B \in \mathcal{P}(\mathcal{P}(X)) \;\middle|\; \mathcal{U} \subset \mathcal{B} \right\} .$$

Proof. Denote, for this proof, the set $\left\{ \bigcup_{B \in \mathcal{U}} B \;\middle|\; \mathcal{U} \subset \mathcal{B} \right\}$ by $\mathcal{S}$ and let us abbreviate $\mathcal{T}_{\mathcal{B}}$ by $\mathcal{T}$. We wish to prove that $\mathcal{T} = \mathcal{S}$. First, note that $\mathcal{B} \subset \mathcal{S}$ by construction. By definition, $\mathcal{B} \subset \mathcal{T}$. Since $\mathcal{T}$ is a topology, it is closed under arbitrary unions. Hence $\mathcal{S} \subset \mathcal{T}$. To prove the converse, it is sufficient to show that $\mathcal{S}$ is a topology. As it contains $\mathcal{B}$, and $\mathcal{T}$ is the smallest such topology, this will provide us with the inverse inclusion. By definition, $\bigcup_{B \in \varnothing} B = \varnothing$ and thus $\varnothing \in \mathcal{S}$. By assumption, since $\mathcal{B}$ is a base, $X = \bigcup_{B \in \mathcal{B}} B$ so $X \in \mathcal{S}$. As the union of unions of elements in $\mathcal{B}$ is a union of elements in $\mathcal{B}$, $\mathcal{S}$ is closed under abritrary unions. Now, let B_1, B_2 be elements of $\mathcal{B}$. If $B_1 \cap B_2 = \varnothing$ then $B_1 \cap B_2 \in \mathcal{S}$. Assume that B_1 and B_2 are not disjoints. Then by definition of a base, for all $x \in B_1 \cap B_2$ there exists $B_x \in \mathcal{B}$ such that $x \in B_x$ and $B_x \subset B_1 \cap B_2$. So

$$B_1 \cap B_2 = \bigcup_{x \in B_1 \cap B_2} B_x ,$$

and therefore, by definition, $B_1 \cap B_2 \in \mathcal{S}$. We conclude that the intersection of two arbitary elements in $\mathcal{S}$ is again in $\mathcal{S}$ by using the distributivity of the union with respect to the intersection. □

1.1.20 Definition We shall say that a base $\mathcal{B}$ on a set X is a *base for a topology* $\mathcal{T}$ on X when the smallest topology containing $\mathcal{B}$ coincides with $\mathcal{T}$, in other words when $\mathcal{T} = \mathcal{T}_{\mathcal{B}}$.

The typical usage of the preceding theorem comes from the following result.

1.1.21 Corollary *Let $\mathcal{B}$ be a base for a topology $\mathcal{T}$ on X. A subset U of X is in $\mathcal{T}$ if and only if for evry $x \in U$ there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subset U$.*

Proof. We showed that any open set for the topology $\mathcal{T}$ is a union of elements in $\mathcal{B}$. Hence if $x \in U$ for $U \in \mathcal{T}$ then there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subset U$. Conversely, if U is some subset of X such that for all $x \in U$ there exists $B_x \in \mathcal{B}$ such that $x \in B_x$ and $B_x \subset U$, then $U = \bigcup_{x \in U} B_x$ and thus $U \in \mathcal{T}$. □

The last result in this section is a useful tool for showing continuity of a map.

1.1.22 Proposition *Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be two topological spaces, $\mathcal{A}$ a base for the topology $\mathcal{T}_X$ and $\mathcal{B}$ a base for the topology $\mathcal{T}_Y$. Assume further that $f : X \to Y$ is a map. Then the following are equivalent:*

(i) *The map f is continuous.*

(ii) *For every open $V \subset Y$ and all $x \in f^{-1}(V)$ there exists $A \in \mathcal{A}$ such that $x \in U$ and $f(A) \subset V$.*

(iii) *For every $B \in \mathcal{B}$ the preimage $f^{-1}(B)$ is open in X.*

Proof. Obviously, (i) implies (iii).

Assume that (iii) holds and that $V \subset Y$ is open. Let $x \in f^{-1}(V)$ and put $y = f(x)$. Then $y \in V$. Since $\mathcal{B}$ is a base for the topology $\mathcal{T}_Y$ there exists $B \in \mathcal{B}$ such that $x \in B \subset V$. By assumption $f^{-1}(B)$ is open in X and $x \in f^{-1}(B)$. Since $\mathcal{A}$ is a base for the topology $\mathcal{T}_X$, there exists $A \in \mathcal{A}$ such that $x \in A \subset f^{-1}(B)$. Since $f^{-1}(B) \subset f^{-1}(V)$, (ii) follows.

Now assume that (ii) holds true. Let $V \subset Y$ be open, and choose for every $x \in f^{-1}(V)$ a base element $A_x \in \mathcal{A}$ such that $x \in A_x \subset f^{-1}(V)$. Then $f^{-1}(V) = \bigcup_{x \in f^{-1}(V)} A_x$ which is open in X. Hence f is continuous. $\square$

1.2. Examples and categorical constructions of topological spaces

This section provides various examples and constructions of topological spaces which will be used all along in this monograph.

The order topology

1.2.1 Proposition *Let $(X, \leqslant)$ be a totally ordered set, and assume that $\infty, -\infty$ are two symbols not in X. Define $[-\infty, \infty]_X = X \cup \{-\infty, \infty\}$ and extend $\leqslant$ to $[-\infty, \infty]_X$ by requiring $x \leqslant y$ for $x, y \in [-\infty, \infty]_X$ to hold when $x, y \in X$ and $x \leqslant y$, when $x = -\infty$, or when $y = \infty$. Then $[-\infty, \infty]_X$ together with the relation $\leqslant$ becomes a totally ordered set as well, and the embedding $X \hookrightarrow [-\infty, \infty]_X$ is order-preserving.*

Proof. By definition, the relation $\leqslant$ on $[-\infty, \infty]_X$ is reflexive, and any two elements of $[-\infty, \infty]_X$ are comparable. Also by definition, $x \leqslant -\infty$ is equivalent to $x = -\infty$ and $\infty \leqslant y$ equivalent to $y = \infty$. Since the restriction of $\leqslant$ to X is antisymmetric by assumption, $\leqslant$ therefore is an antisymmetric relation on $[-\infty, \infty]_X$. Using the definition of $\leqslant$ again one finally observes that for $x, y, z \in [-\infty, \infty]_X$ the following implications hold true.

$$\begin{aligned}
-\infty \leqslant y \,\&\, y \leqslant z &\implies -\infty \leqslant z \\
x \leqslant -\infty \,\&\, -\infty \leqslant z &\implies x = -\infty \leqslant z \\
x \leqslant y \,\&\, y \leqslant -\infty &\implies x = y = -\infty \\
x \leqslant y \,\&\, y \leqslant \infty &\implies x \leqslant \infty \\
x \leqslant \infty \,\&\, \infty \leqslant z &\implies x \leqslant \infty = z \\
\infty \leqslant y \,\&\, y \leqslant z &\implies \infty = y = z \ .
\end{aligned}$$

Since its restriction to X is already transitive, transitivity of $\leqslant$ now follows and the proposition is proved. □

1.2.2 Remark For the rest of this paragraph we always assume that an ordered set $(X, \leqslant)$ does not contain the symbols $\infty, -\infty$, and that $[-\infty, \infty]_X$ and the extended order relation $\leqslant$ are defined as in the preceding proposition.

1.2.3 Definition For a totally ordered set $(X, \leqslant)$, define *intervals* with boundaries $x, y \in [-\infty, \infty]$ as follows:

$$\begin{aligned}
(x,y) &:= (x,y)_X := \big\{z \in [-\infty,\infty] \mid x < z < y\big\} ,\\
[x,y) &:= [x,y)_X := \big\{z \in [-\infty,\infty] \mid x \leqslant z < y\big\} ,\\
(x,y] &:= (x,y]_X := \big\{z \in [-\infty,\infty] \mid x < z \leqslant y\big\} ,\\
[x,y] &:= [x,y]_X := \big\{z \in [-\infty,\infty] \mid x \leqslant z \leqslant y\big\} .
\end{aligned}$$

The intervals $(x, y)_X$ are called *open intervals*, intervals of the form $[x, y]_X$ are called *closed intervals*, and intervals of the form $[x, y)_X$ or $(x, y]_X$ are the *half-open intervals*.

1.2.4 Remarks (a) Note that in case $x = y$ only the closed interval $[x, x]_X$ is non-empty. In case $y < x$ all the intervals $(x, y)_X$, $[x, y)_X$, $(x, y]_X$, and $[x, y]_X$ are empty.

(b) We mostly use the notation (x, y), $[x, y]$, etc. for intervals and denote the X in intervals only when otherwise some ambiguity could appear.

1.2.5 Definition Let $(X, \leqslant)$ be a totally ordered set. Then the topology generated by the set

$$\mathcal{I}_X = \big\{(x,y) \in \mathcal{P}(X) \mid x,y \in [-\infty,\infty] \,\&\, x \leqslant y\big\}$$

is called the *order topology* on X. It is usually denoted $\mathcal{T}_{(X,\leqslant)}$.

1.2.6 Proposition *Let $(X, \leqslant)$ be a totally ordered set. Then the set $\mathcal{I}_X$ is a base for the order topology on X. A subbase of the order topology is given by the set $\mathcal{S}_X$ of* rays *(x, ∞) and $(-\infty, y)$, where x, y run through the elements of X.*

Proof. Since X is totally ordered, so is $[-\infty, \infty]$. It is immediate that $(x, y) \cap (x', y') = (w, z)$ if w is the largest of x and x' and z is the smallest of y and y'. Hence $\mathcal{I}_X$ is a base of the order topology.

Since $(x, \infty) \cap (-\infty, y) = (x, y)$ for $x \leqslant y$, the set $\mathcal{S}_X$ is a subbase of the order topology. □

1.2.7 Example The standard topology on $\mathbb{R}$ from Example 1.1.3 (d) is the order topology. Likewise, the standard topology on $\mathbb{Q}$ coincides with the order topology.

1.2.8 Remark If X neither has a minimum nor a maximum, one usually denotes the space $[-\infty, \infty]$ by $\overline{X}$. This notation fits with the understanding that $\overline{}$ denotes the closure operation, because the closure of X in $[-\infty, \infty]$ with respect to the order topology coincides with the full space $[-\infty, \infty]$ under the assumptions made.

Extending the ordered set of real numbers $(\mathbb{R}, \leqslant)$ in that way gives the so-called *extended real number system* $\bar{\mathbb{R}}$.

The subspace topology

1.2.9 Proposition and Definition *Let $(X, \mathcal{T})$ be a topological space. Let $S \subset X$ and $\iota : S \hookrightarrow X$ the canonical embedding. Then initial topology $\iota^*\mathcal{T}$ coincides with*

$$\mathcal{T}_S^X := \{U \cap S \in \mathcal{P}(S) \mid U \in \mathcal{T}\} \ .$$

One calls $\mathcal{T}_Y^X$ the subspace *or* trace topology *on S. Sometimes one says that $\mathcal{T}_Y^X$ is the* topology induced by $(X, \mathcal{T})$.

Proof. The claim follows immediately from the definition of the initial topology $\iota^*\mathcal{T}$. □

Just as easy is the following observation:

1.2.10 Proposition *Let $(X, \mathcal{T})$ be a topological space, and $S \subset X$ a subset. Let $\mathcal{B}$ be a basis for $\mathcal{T}$. Then the set*

$$\mathcal{B}_S^X := \{B \cap S \in \mathcal{P}(S) \mid B \in \mathcal{B}\}$$

is a basis for the subspace topology on S induced by $(X, \mathcal{T})$.

Proof. Trivial exercise. □

1.2.11 Example The default topologies on $\mathbb{N}$ and Z are the subspace topologies induced by the standard topology on $\mathbb{R}$. Since $\{n\} = (n - \frac{1}{2}, n + \frac{1}{2}) \cap \mathbb{Z}$ for all $n \in \mathbb{Z}$, we see that the natural topologies on $\mathbb{N}$ and $\mathbb{Z}$ are in fact the discrete topologies. The topology on $\mathbb{Q}$ induced by the standard topology on $\mathbb{R}$ coincides with the default topology on $\mathbb{Q}$ (which is, as pointed out above, the same as the order topology).

The quotient topology

The product topology

1.2.12 Definition Let I be some nonempty set. Let us assume given a family $(X_i, \mathcal{T}_i)_{i \in I}$ of topological spaces. Consider the cartesian product $X := \prod_{i \in I} X_i$ and denote for each $j \in I$ by $\pi_j : X \to X_j$, $(x_i)_{i \in I} \mapsto x_j$ the projection on the i-th coordinate. The initial topology on X with respect to the

basic open set of the cartesian product $\prod_{i \in I} E_i$ is a set of the form $\prod_{i \in I} U_i$ where $\{i \in I : U_i \neq E_i\}$ is finite and for all $i \in I$, we have $U_i \in \mathcal{T}_i$.

1.2.13 Definition Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{T}_i)_{i \in I}$ of topological spaces. The product topology on $\prod_{i \in I} E_i$ is the smallest topology containing all the basic open sets.

1.2.14 Proposition *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{T}_i)_{i \in I}$ of topological spaces. The collection of all basic open sets is a basis on the set $\prod_{i \in I} E_i$.*

Proof. Trivial exercise. □

1.2.15 Remark The product topology is not just the basic open sets on the cartesian products: there are many more open sets!

1.2.16 Proposition *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{T}_i)_{i\in I}$ of topological spaces. The product topology on $\prod_{i\in I} E_i$ is the initial topology for the the set $\{p_i : i \in I\}$ where $p_i : \prod_{j\in I} E_j \to E_i$ is the canonical surjection for all $i \in I$.*

Proof. Fix $i \in I$. Let $V \in \mathcal{T}_{E_i}$. By definition, $p_i^{-1}(V) = \prod_{j\in I} U_j$ where $U_j = E_j$ for $j \in I\backslash\{i\}$, and $U_i = V$. Hence $p_i^{-1}(V)$ is open in the product topology. As V was an arbitrary open subset of E_i, the map p_i is continuous by definition. Hence, as i was arbitrary in I, the initial topology for $\{p_i : i \in I\}$ is coarser than the product topology.

Conversely, note that the product topology is generated by $\{p_i^{-1}(V) : i \in I, V \in \mathcal{T}_{E_i}\}$, so it is coarser than the initial topology for $\{p_i : i \in I\}$. This concludes this proof. □

1.2.17 Corollary *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{T}_i)_{i\in I}$ of topological spaces. Let $\mathcal{T}$ be the product topology on $F = \prod_{i\in I} E_i$. Let $(D, \mathcal{T}_D)$ be a topological space. Then $f : D \to F$ is continuous if and only if $p_i \circ f$ is continuous from $(D, \mathcal{T}_D)$ to $(E_i, \mathcal{T}_{E_i})$ for all $i \in I$, where p_i is the canonical surjection on E_i for all $i \in I$.*

Proof. We simply applied the fundamental property of initial topologies. □

1.2.18 Remarks (a) The *box topology* on the cartesian product $\prod_{i\in I} X_i$ is the smallest topology containing all possible cartesian products of open sets $U_i \subset X_i$, $i \in I$. The box topology is strictly finer than the product topology when the index set is infinite and infintely many of the X_i carry a topology strictly finer than the indiscrete topology. Of course, the box and product topologies coincide otherwise, in particular when the product is finite.

(b) Since the product topology is the coarsest topology which makes the canonical projections continuous, it is the preferred and default one on cartesian products.

The metric topology

1.2.19 Definition Let X be a set. A function $d : X \times X \to \mathbb{R}_{\geqslant 0}$ is a *distance* or *metric* on X when:

(M1) For all $x, y \in X$ the relation $d(x, y) = 0$ holds true if and only if $x = y$.

(M2) The map d is *symmetric* that is one has $d(x, y) = d(y, x)$ for all $x, y \in X$.

(M3) For all $x, y, z \in X$ the *triangle inequality* $d(x, y) \leqslant d(x, z) + d(z, y)$ is satisfied.

If instead of (M1) the axiom (M1)' below is fulfilled while (M2) and (M3) are still valid, then d is called a *pseudometric* on X.

(M1)' For all $x \in X$ the equality $d(x, x) = 0$ holds true.

A pair (X, d) is a *metric space* when X is a set and d a distance on X. If d is only a pseudometric on X, one calls the pair (X, d) a *pseudometric space*.

The following is often useful.

1.2.20 Lemma *Let (X, d) be a pseudometric space. Let $x, y, z \in X$. Then*

$$|d(x,y) - d(x,z)| \leqslant d(y,z) \ .$$

Proof. Since $d(x,y) \leqslant d(x,z) + d(z,y)$ we have $d(x,y) - d(x,z) \leqslant d(z,y) = d(y,z)$. Since $d(x,z) \leqslant d(x,y) + d(y,z)$ we have $d(x,z) - d(x,y) \leqslant d(y,z)$. Hence the claim holds. □

1.2.21 Definition Let (X, d) be a pseudometric space. Let $x \in E$ and $r \in \mathbb{R}_{>0}$. The *open ball* with *center* x and *radius* r in (X, d) is the set

$$\mathbb{B}(x,r) = \mathbb{B}_r(x) = \{y \in X \mid d(x,y) < r\} \ .$$

The *closed ball* with *center* x and *radius* r is defined by

$$\overline{\mathbb{B}}(x,r) = \overline{\mathbb{B}}_r(x) = \{y \in X \mid d(x,y) \leqslant r\} \ .$$

1.2.22 Definition Let (X, d) be a pseudometric space. The *metric topology* on X induced by d is the smallest topology containing all the open balls of X.

1.2.23 Theorem *Let (X, d) be a pseudometric space. The set of all open balls on X is a basis for the metric topology on X induced by d.*

Proof. It is enough to show that the set of all open balls is a topological basis. By definition, $X = \bigcup_{x \in X} \mathbb{B}(x, 1)$. Now, let us be given $\mathbb{B}(x, r_x)$ and $\mathbb{B}(y, r_y)$ for some $x, y \in X$ and $r_x, r_y > 0$. If the intersection of these two balls is empty, we are done; let us assume that there exists $z \in \mathbb{B}(x, r_x) \cap \mathbb{B}(y, r_y)$. Let r be the smallest of $r_x - d(x,z)$ and $r_y - d(y,z)$. Let $w \in \mathbb{B}(z,r)$. Then

$$d(x,w) \leqslant d(x,z) + d(z,w) < d(x,z) + r_x - d(x,z) = r_x \ ,$$

so $w \in \mathbb{B}(x, r_x)$. Similarly, $w \in \mathbb{B}(y, r_y)$. Hence, $\mathbb{B}(z,r) \subset \mathbb{B}(x, r_x) \cap \mathbb{B}(y, r_y)$ as desired. □

The following result shows that metric topologies are minimal in the sense of making the distance functions continuous.

1.2.24 Proposition *Let (X, d) be a pseudometric space. For all $x \in X$, the function*

$$d_x : X \to \mathbb{R}_{\geqslant 0}, \quad y \mapsto d(x,y)$$

is continuous on X for the metric topology. Moreover, the metric topology is the smallest topology such that all the functions d_x, $x \in X$ are continuous.

Proof. Fix $x \in X$. To verify continuity of the maps d_x it is sufficient to show that the preimages of $[0,r)$ and (r,∞) by d_x are open in the metric topology of X, where $r > 0$ is arbitrary. Indeed, these intervals form a subbasis for the topology of $[0,\infty)$ which we assume to carry the subspace topology induced by the order topology on $\mathbb{R}$. Let $r > 0$ be given. Then $d_x^{-1}([0,r)) = \mathbb{B}(x,r)$ by definition, so it is open. Now, let $y \in X$ such that $d(x,y) > r$. Let $\varrho_y = d(x,y) - r > 0$. If $d(w,y) < \varrho_y$ for some $w \in X$, then $d(x,y) - d(w,y) \leqslant d(x,w)$, so $d(x,w) > r$. Hence

$$\mathbb{B}(y, \rho_y) \subset d_x^{-1}((r,\infty)) \quad \text{for all } y \in d_x^{-1}((r,\infty)) \ .$$

Therefore, $d_x^{-1}((r,\infty))$ is open.

Finally, since $d_x^{-1}([0,r)) = \mathbb{B}(x,r)$ for all $x \in X$ and $r > 0$ the minimal topology making all the maps d_x continuous must indeed contain the metric topology as desired, and our proposition is proven. □

1.2.25 Remark The metric topology is the default topology on a pseudometric space.

1.2.26 There are more examples of continuous functions between metric spaces. More precisely, a natural category for metric spaces consists of metric spaces and Lipschitz maps as arrows, defined as follows.

1.2.27 Definition Let (X, d_X), (Y, d_Y) be pseudometric spaces. A function $f : X \to Y$ for which there exists an $L > 0$ such that

$$d_Y(f(x), f(y)) \leqslant L\, d_X(x,y) \quad \text{for all } x, y \in X$$

is called *Lipschitz*.

1.2.28 Definition Let (X, d_X), (Y, d_Y) be pseudometric spaces. Let $f : X \to Y$ be a Lipschitz function. Then the *Lipschitz constant* of f is defined as

$$\operatorname{Lip}(f) = \sup \left\{ \frac{d_Y(f(x), f(y))}{d_X(x,y)} \,\middle|\, x, y \in X, d(x,y) \neq 0 \right\} .$$

A Lipschitz function with Lipschitz constant $L \leqslant 1$ is called a *metric map*. If its Lipschitz constant is < 1, then the Lipschitz function is called a *contraction*.

1.2.29 Examples (a) A constant map $f : X \to Y$ between pseudometric spaces is Lipschitz with Lipschitz constant 0. If both X and Y are metric spaces and $f : X \to Y$ is Lipschitz, then $\operatorname{Lip}(f) = 0$ if and only if f is constant.

(b) The identity map $\mathrm{id}_X : X \to X$ on a pseudometric space (X,d) is Lipschitz. If d is not the zero pseudometric on X, then $\operatorname{Lip}(id_X) = 0$.

1.2.30 Proposition *Let (X, d_X), (Y, d_Y) be pseudometric spaces. If $f : X \to Y$ is a Lipschitz function, then it is continuous.*

Proof. Let L be the Lipschitz constant for f. Let $y \in Y$ and $\varepsilon > 0$. Let $x \in f^{-1}(\mathbb{B}(y,\varepsilon))$. Put $\delta_x = \frac{\varepsilon - d(f(x),y)}{L+1}$ and observe that $\delta_x > 0$. Then, for $z \in \mathbb{B}(x, \delta_x)$,

$$\begin{aligned} d_Y(f(z), y) &\leqslant d_Y(f(z), f(x)) + d_Y(f(x), y) \leqslant L d_X(z,x) + d_Y(f(x), y) < \\ &< \varepsilon - d_Y(f(x), y) + d_Y(f(x), y) = \varepsilon. \end{aligned}$$

Hence $f^{-1}(\mathbb{B}(y,\varepsilon))$ is open and f is continuous. □

1.2.31 Proposition and Definition *Pseudometric spaces as objects together with metric maps between them form a category* PMet *which is called the* category of pseudometric spaces. *Changing the morphism class to Lipschitz maps between pseudometric spaces gives another category which we denote* PMetLip *and call the* category of pseudometric spaces and Lipschitz functions. *Metric spaces together with metric or Lipschitz maps between them form full subcategories* Met *and* MetLip *of* PMet *and* PMetLip, *respectively. They are called the* category of metric spaces *respectively the* category of metric spaces and Lipschitz functions.

Proof. The claim immediately follows from the observation that the identity map on a pseudometric space is metric and that the composition of two metric respectively Lipschitz maps is again metric respectively Lipschitz. □

1.2.32 Remark Using metric or Lipschitz maps as morphisms for categories of metric or pseudometric spaces is natural. Other, more general type of morphisms, would be uniform continuous maps, which are discussed in later sections.

Co-Finite Topologies

A potential source for counter-examples, the family of cofinite topologies is easily defined:

1.2.33 Proposition *Let E be a set. Let:*

$$\mathcal{T}_{\mathrm{cof}}(E) = \{\varnothing\} \cup \{U \subset E : \complement_E U \text{ is finite }\}.$$

Then $\mathcal{T}_{\mathrm{cof}}(E)$ is a topology on E.

Proof. By definition, $\varnothing \in \mathcal{T}_{\mathrm{cof}}(E)$. Moreover, $\complement_E E = \varnothing$ which is finite, so $E \in \mathcal{T}_{\mathrm{cof}}(E)$. Let $U, V \in \mathcal{T}_{\mathrm{cof}}(E)$. If U or V is empty then $U \cap V = \varnothing$ so $U \cap V \in \mathcal{T}_{\mathrm{cof}}(E)$. Otherwise, $\complement_E(U \cap V) = \complement_E U \cup \complement V$ which is finite, since by definition $\complement_E U$ and $\complement_E V$ are finite. Hence $U \cap V \in \mathcal{T}_{\mathrm{cof}}(E)$. Last, let $\mathcal{U} \subset \mathcal{T}_{\mathrm{cof}}(E)$. Again, if $\mathcal{U} = \{\varnothing\}$ then $\bigcup \mathcal{U} = \varnothing \in \mathcal{T}_{\mathrm{cof}}(E)$. Let us now assume that $\mathcal{U}$ contains at least one nonempty set V. Then:

$$\complement_E \bigcup \mathcal{U} = \bigcap \{\complement_E U : U \in \mathcal{U}\} \subset \complement_E V.$$

Since $\complement_E V$ is finite by definition, so is $\bigcup \mathcal{U}$, which is therefore in $\mathcal{T}_{\mathrm{cof}}(E)$. This completes our proof. □

The one-point compactification of $\mathbb{N}$

Limits of sequences is a central tool in topology and this section introduces the natural topology for this concept. The general notion of limit is the subject of the next chapter.

1.2.34 Definition Let ∞ be some symbol not found in $\mathbb{N}$. We define $\bar{\mathbb{N}}$ to be $\mathbb{N} \cup \{\infty\}$.

1.2.35 Proposition *The set:*

$$\mathcal{T}_{\bar{\mathbb{N}}} = \{U \subset \bar{\mathbb{N}} : (U \subset \mathbb{N}) \vee (\infty \in U \wedge \complement_{\mathbb{N}} U \text{ is finite})\}$$

is a topology on $\bar{\mathbb{N}}$.

Proof. By definition, $\varnothing \subset \mathbb{N}$ so $\varnothing \in \mathcal{T}_{\bar{\mathbb{N}}}$. Moreover $\complement_{\bar{\mathbb{N}}} \bar{\mathbb{N}} = \varnothing$ which has cardinal 0 so $\bar{\mathbb{N}} \in \mathcal{T}_{\bar{\mathbb{N}}}$. Let $U, V \in \mathcal{T}_{\bar{\mathbb{N}}}$. If either U or V is a subset of $\mathbb{N}$ then $U \cap V$ is a subset of $\mathbb{N}$ so $U \cap V \in \mathcal{T}_{\bar{\mathbb{N}}}$. Othwiwse, $\infty \in U \cap V$. Yet $\complement_{\bar{\mathbb{N}}}(U \cap V) = \complement_{\bar{\mathbb{N}}} U \cup \complement_{\bar{\mathbb{N}}} V$ which is finite as a finite union of finite sets. Hence $U \cap V \in \mathcal{T}_{\bar{\mathbb{N}}}$ again.

Last, assume that $\mathcal{U} \subset \mathcal{T}_{\bar{\mathbb{N}}}$. Of course, $\infty \in \bigcup \mathcal{U}$ if and only if $\infty \in U$ for some $U \in \mathcal{U}$. So, if $\infty \notin \bigcup \mathcal{U}$ then $\bigcup \mathcal{U} \in \mathcal{T}_{\bar{\mathbb{N}}}$ by definition. If, on the other hand, $\infty \in \bigcup \mathcal{U}$, then there exists $U \in \mathcal{U}$ with $\complement_{\bar{\mathbb{N}}} U$ finite. Now, $\complement_{\bar{\mathbb{N}}} \bigcup \mathcal{U} = \bigcap \{\complement_{\bar{\mathbb{N}}} V : V \in \mathcal{U}\} \subset \complement_{\bar{\mathbb{N}}} U$ so it is finite, and thus again $\bigcup \mathcal{U} \in \mathcal{T}_{\bar{\mathbb{N}}}$. □

1.3. Separation properties

1.3.1 The general definition of a topology allows for examples where elements of a topological space, seen as a set, can not be distinguished from each other by open sets (for instance if the topology is indiscrete). When points can be topologically differentiated, a topology is in some sense separated. The standard separation axioms allow to subsume topological spaces with certain separability properties in particular classes. One then studies the properties of these clases, often with a view to particular applications, and attempts to create counter examples, meaning examples not satsifying the corresponding separation axioms. The most important separability property goes back to the founder of set-theoretic topology, Felix Hausdorff, who introduced it in 1914. The first full presentation of the separation axioms as we know them today appeared in the classic book *Topologie* by Alexandroff & Hopf (1965) under their German name *Trennungsaxiome.*

Let us note that the literature on separation axioms is not uniform when it comes to the axioms (T3) to (T6) below, so one needs to always check which convention an author follows. Here, we follow the convention by (Steen & Seebach, 1995, Part I, Chap. 2) which coincides with the one of

1.3.2 Definition (The Separation Axioms) Recall that two subsets A, B of a topological space $(X, \mathcal{T})$ are called *disjoint* if $A \cap B = \varnothing$. The two sets are called *separated* if $\bar{A} \cap B = A \cap \bar{B} = \varnothing$. The topological space $(X, \mathcal{T})$ now is said to be

(T0) or *Kolmogorov* if for each pair of distinct points $x, y \in X$ there is an open $U \subset X$ such that $x \in U$ and $y \notin U$ holds true, or $y \in U$ and $x \notin U$,

(T1) or *Fréchet* if for each pair of distinct points $x, y \in X$ there is an open $U \subset X$ such that $x \in U$ and $y \notin U$,

(T2) or *Hausdorff* if for each pair of distinct points $x, y \in X$ there exist disjoint open sets $U, V \subset X$ such that $x \in U$ and $y \in V$,

(T2$\frac{1}{2}$) or *Uryson* or *completely Hausdorff* if for each pair of distinct points $x, y \in X$ there exist distinct closed neigborhoods U of x and V of y,

(T3) if for each point $x \in X$ and closed subset $A \subset X$ with $x \notin A$ there exist disjoint open sets $U, V \subset X$ such that $x \in U$ and $A \subset V$,

(T3$\frac{1}{2}$) if for each point $x \in X$ and closed subset $A \subset X$ with $x \notin A$ there exists a continuous function $f : X \to \mathbb{R}$ such that $f(x) = 0$ and $f(A) = \{1\}$,

(T4) if for each pair of closed disjoint subsets $A, B \subset X$ there exist disjoint open sets $U, V \subset X$ such that $A \subset U$ and $B \subset V$,

(T5) if for each pair of separated subsets $A, B \subset X$ there exist disjoint open sets $U, V \subset X$ such that $A \subset U$ and $B \subset V$,

(T6) if for each pair of disjoint closed subsets $A, B \subset X$ there exists a continuous function $f : X \to \mathbb{R}$ such that $A = f^{-1}(0)$ and $B = f^{-1}(0)$.

A Hausdorff space will be called *regular* if it fulfills (T3) , *completely regular*, if it satisfies ($\mathsf{T3}_{\frac{1}{2}}$), and *normal* if (T4) holds true. Finally we call a Hausdorff space *completely normal* if it is (T5) and *perfectly normal* if it is (T6) .

1.4. Filters and convergence

Filters and ultrafilters

1.4.1 Definition Let X be a set. A subset $\mathcal{F}$ of the powerset $\mathcal{P}(X)$ is called a *filter* on X if it satisfies the following axioms:

(Fil1) The empty set $\varnothing$ is not an element of $\mathcal{F}$.

(Fil2) The set X is an element of $\mathcal{F}$.

(Fil3) If $A \in \mathcal{F}$ and if $B \in \mathcal{P}(X)$ satisfies $A \subset B$, then $B \in \mathcal{F}$.

(Fil4) If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then the intersection $A \cap B$ is an element of $\mathcal{F}$ as well.

If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two filters on X such that $\mathcal{F}_1 \subset \mathcal{F}_2$, then one calls $\mathcal{F}_1$ a *subfilter* of $\mathcal{F}_2$ or says that $\mathcal{F}_2$ is *finer* than $\mathcal{F}_1$. Sometimes one expresses this by saying that $\mathcal{F}_2$ *refines* $\mathcal{F}_1$. Filters maximal with respect to set inclusion are called *ultrafilters.* A filter $\mathcal{F}$ is called *free* if $\bigcap_{A\in\mathcal{F}} A = \varnothing$ otherwise it is called *fixed.*

1.4.2 Examples (a) For every set X, the set $\{X\}$ is a filter. It is the smallest of all filters on X.

(b) Given an element $x \in X$ the set $\mathcal{F}_x := \{A \in \mathcal{P}(X) \mid x \in A\}$ is an ultrafilter on X. More generally, if $Y \subset X$ is a non-empty subset, then $\mathcal{F}_Y := \{A \in \mathcal{P}(X) \mid Y \subset A\}$ is a filter on X. It is an ultrafilter if and only if Y has exactly one element.

(c) If $(X, \mathcal{T})$ is a topological space and $x \in X$ an element, then the *neigborhood filter* $\mathcal{U}_x := \{V \in \mathcal{P}(X) \mid \exists U \in \mathcal{T} : \ x \in U \subset V\}$ is a filter contained in $\mathcal{F}_x$. The filters $\mathcal{U}_x$ and $\mathcal{F}_x$ coincide if and only if x is an isolated point.

(d) Now consider the reals and let $\mathcal{F} = \{A \in \mathcal{P}(\mathbb{R}) \mid \exists\, \varepsilon > 0 : \ [0, \varepsilon) \subset A\}$. Then $\mathcal{F}$ is a filter on $\mathbb{R}$ which is properly contained in the ultrafilter $\mathcal{F}_0$ and which properly contains the neighborhood filter $\mathcal{U}_0$ (where $\mathbb{R}$ carries the standard topology).

1.4.3 Proposition *Let $\mathcal{A} \subset \mathcal{P}(X)$ be a non-empty set of subset of X which has the* finite intersection property *that is that $A_1 \cap \ldots \cap A_n$ is non-empty for all $n \in \mathbb{N}^*$ and all $A_1, \ldots, A_n \in \mathcal{A}$. Then there is an ultrafilter $\mathcal{F}$ containing $\mathcal{A}$.*

Proof. Let P be the set of all $\mathcal{J} \subset \mathcal{P}(X)$ having the finite intersection property and containing $\mathcal{A}$. Then P is non-empty, as it contains at least $\mathcal{A}$, and is ordered by set inclusion. If $C \subset P$ is a chain, then $\mathcal{M} := \bigcup_{\mathcal{J}\in C} \mathcal{J}$ contains $\mathcal{A}$ and fulfills the finite intersection property. To verify the latter let $Y_1, \ldots, Y_n \in \mathcal{M}$. Then there exist $\mathcal{J}_1, \ldots, \mathcal{J}_n \in C$ such that $Y_i \in \mathcal{J}_i$ for $i = 1, \ldots, n$. Hence all Y_i lie in the maximum $\mathcal{J}_{\mathrm{m}}$ of the sets $\mathcal{J}_1, \ldots, \mathcal{J}_n$. But $\mathcal{J}_{\mathrm{m}}$ has the finite intersection property, hence $Y_1 \cap \ldots \cap Y_n \neq \varnothing$. So $\mathcal{M}$ is an upper bound of the chain C. By Zorn's Lemma,

P has a maximal element $\mathcal{F}$. It contains $\mathcal{A}$ and has the finite intersection property. Moreover, if $A \in \mathcal{F}$ and $B \in \mathcal{P}(X)$ contains A as a subset, then $\mathcal{F} \cup \{B\}$ also satisfies the finite intersection property, hence by maximality of $\mathcal{F}$ one concludes $B \in \mathcal{F}$. Again by maximality $\mathcal{F}$ has to be an ultrafilter. □

1.4.4 Corollary *Every filter on X is contained in an ultrafilter.*

Proof. This follows from the preceding proposition since a filter has the finite intersection property. □

1.4.5 Theorem *Let $\mathcal{F}$ be a filter on a set X. Then the following are equivalent:*

(i) *$\mathcal{F}$ is an ultrafilter.*

(ii) *If A is a subset of X and A has non-empty intersection with every element of $\mathcal{F}$, then $A \in \mathcal{F}$.*

(iii) *For all $A \subset X$ either $A \in \mathcal{F}$ or $X \backslash A \in \mathcal{F}$.*

Convergence of filters

1.4.6 Definition

1.5. Nets

Directed sets

Let us first recall that by a *preordered set* one understands a set P together with a binary relation $\leqslant$ which is reflexive and transitive, see Definition 2.1.37.

1.5.1 Definition (Directed sets) By a directed set one understands a preordered set $(P, \leqslant)$ such that the binary relation $\leqslant$ is *directed* which means that

(Dir) for all $x, y \in D$ there exists an element $z \in D$ with $x \leqslant z$ and $y \leqslant z$.

1.5.2 Remark The property that $(P, \leqslant)$ is directed is the same as saying that any two elements of the preordered set P have an upper bound.

1.6. Compactness

Quasi-compact topological spaces

1.6.1 Before we come to defining quasi-compactness let us recall some relevant notation. By a *cover* (or *covering*) of a set X one understands a family $\mathcal{U} = (U_i)_{i \in I}$ of subsets $U_i \subset X$ such that $X \subset \bigcup_{i \in I} U_i$. This terminology also holds for a subset $Y \subset X$. That is a family $\mathcal{U} = (U_i)_{i \in I}$ of subsets $U_i \subset X$ is called a *cover* of Y if $Y \subset \bigcup_{i \in I} U_i$. A *subcover* of a cover $\mathcal{U} = (U_i)_{i \in I}$ of Y or shortly a subcover of $\mathcal{U}$ then is a subfamily $(U_i)_{i\in J}$ which also covers Y which means that $J \subset I$ and $Y \subset \bigcup_{i\in J} U_i$. If J is finite, one calls the subcover $(U_i)_{i\in J}$ a *finite subcover.* If $(X, \mathcal{T})$ is a topological space and all elements U_i of a cover $\mathcal{U} = (U_i)_{i\in I}$ of some $Y \subset X$ are open sets, the cover is called an *open cover* of Y.

1.6.2 Proposition *Let be a topological spaces $(X, \mathcal{T})$. Then the following are equivalent:*

(i) *Every open cover of X has a finite subcover.*

(ii) *For every family $(A_i)_{i\in I}$ of closed subset $A_i \subset X$ such that $\bigcap_{i\in I} A_i = \varnothing$ there exist finitely many elements $A_{i_1}, \ldots, A_{i_n}$ such that $A_{i_1} \cap \ldots \cap A_{i_n} = \varnothing$.*

(iii) *Every filter on X has an accummulation point.*

(iv) *Every ultrafilter on X converges.*

Proof. Assume that (i) holds true and let $(A_i)_{i\in I}$ be a family of closed subset $A_i \subset X$ such that $\bigcap_{i\in I} A_i = \varnothing$. Put $U_i := X \backslash A_i$ for all $i \in I$. Then $(U_i)_{i\in I}$ is an open covering of X, hence by assumption there exist $i_1, \ldots, i_n \in I$ such that $X = U_{i_1} \cup \ldots \cup U_{i_n}$. By de Morgan's laws the relation $A_{i_1} \cap \ldots \cap A_{i_n} = \varnothing$ the follows, hence (ii) follows.

Next assume (ii), and let $\mathcal{F}$ be a filter on X. Then $\overline{A_1} \cap \ldots \cap \overline{A_n} \neq \varnothing$ for all $n \in \mathbb{N}^*$ and $A_1, \ldots, A_n \in \mathcal{F}$, since $\mathcal{F}$ is a filter. Hence $\bigcap_{A\in\mathcal{F}} \bar{A} \neq \varnothing$ by (ii). Every element of $\bigcap_{A\in\mathcal{F}} \bar{A}$ now is an accummulation point of $\mathcal{F}$, so (iii) follows.

By **??**, (iii) implies (iv).

Finally assume that every ultrafilter on X converges, and let $\mathcal{U} = (U_i)_{i\in I}$ be an open cover of X. Assume that $\mathcal{U}$ has no finite subcover. For each finite subset $J \subset I$ the set $B_J := X \backslash \bigcup_{i\in J} U_i$ then is non-empty, hence $\mathcal{B} := \{B_J \in \mathcal{P}(X) \mid J \subset I \;\&\; \#J < \infty\}$ is a filter base. Let $\mathcal{F}$ be an ultrafilter containing $\mathcal{B}$. By assumption $\mathcal{F}$ converges to some $x \in X$. Since $\mathcal{U}$ is an open covering of X there is some U_i with $x \in U_i$, hence U_i since $\mathcal{F}$ converges to x. On the other hand $X \backslash U_i \in \mathcal{B} \subset \mathcal{F}$ by construction. This is a contradiction, so $\mathcal{U}$ must have a finite subcover. ☐

1.6.3 Definition (Bourbaki (1998)[I.§9.1.)]A topological space $(X, \mathcal{T})$ is called *quasi-compact*, if every filter on X has an accummulation point.

1.6.4 Theorem (Alexander Subbase Theorem) *Let $(X, \mathcal{T})$ be a topological space, and $\mathcal{S}$ an* adequate *subbase of the topology that is a subbase of $\mathcal{T}$ such that $X = \bigcup_{S\in\mathcal{S}} S$. If every cover of X by elements of $\mathcal{S}$ has a finite subcover, the topological space $(X, \mathcal{T})$ is quasi-compact.*

Compact topological spaces

1.7. The compact-open topology on function spaces

Let X and Y be topological spaces. We denote the set of all functions from Y to X by X^Y. This is the same thing as the direct product $\prod_Y X$ of X over Y. The space of continuous functions $\mathcal{C}(Y, X)$ sits in X^Y so we can give $\mathcal{C}(Y, X)$ the product topology induced by X^Y. This is the topology of *pointwise convergence* and will not be useful for studying most function spaces.

We will instead be interested in the *compact open topology* which is the topology of *uniform convergence on compact sets.*

1.7.1 Definition Let X and Y be topological spaces. The *compact open topology* on $\mathcal{C}(Y,X)$ is the topology with subbasis given by the sets $\mathcal{V}(K,U) = \{f \in \mathcal{C}(Y,X) | f(K) \subset U\}$ for $K \subset Y$ compact and $U \subset X$ open.

1.7.2 Definition A topology $\mathcal{T}$ on $\mathcal{C}(Y,X)$ is called *admissable* if the evaluation map $e : \mathcal{C}(Y,X) \times Y \to X$, $(f,y) \mapsto f(y)$ is continuous.

1.7.3 Proposition *The compact open topology is coarser than any admissable topology on* $\mathcal{C}(Y,X)$.

Proof. Let $\mathcal{T}$ be an admissable topology on $\mathcal{C}(Y,X)$ so that the evaluation map $e : \mathcal{C}(Y,X) \times Y \to X$ is continuous. Let $K \subset Y$ be compact, $U \subset X$ be open and $f \in T(K,U)$. We have to find $V \in O$ such that $f \in V \subset T(K,U)$. Let $k \in K$. Since e is continuous and U is an open neighborhood of $f(x)$, then there are open sets $W_k \subset Y$ and $V_k \subset C_O(Y,X)$ such that $k \in W_k$, $f(k) \in V_k$ amd $e(V_k \times W_k) \subset U$. Since K is compact, there are $k_1, k_2, ..., k_l \in K$ such that $K \subset \bigcup_{i=1}^{l} W_{k_i}$. Put $V := \bigcap_{i=1}^{l} V_{k_i}$ so that $f \in V$ and V is open in O. Now take $g \in V$ and let $k \in K$. Choose i such that $k \in W_{k_i}$ and observe that $g \in W_{k_i}$ so that

$$g(k) = e(g,k) \in e(V_{k_i} \times W_{k_i} \subset U$$

Hense $g \in T(K,U)$ □

1.7.4 Theorem *If Y is locally compact, then the compact open topology on $\mathcal{C}(Y,X)$ is admissable, and it is the coarsets topology on $\mathcal{C}(Y,X)$ with that property.*

Proof. We have to show that

$$e : \mathcal{C}(Y,X) \times Y \to X(f,y) \mapsto f(y)$$

is continuous. Since sets of the form $T(K,U)$ form a subbasis for the compact open topology, it suffices to show that for an open neighborhood $W \subset X$ of some $e(f,y)$, there is compact $K \subset Y$, open $U \subset X$ and open $V \subset Y$ such that $e(T(K,U) \times V) \subset W$ with $f \in T(K,U)$ and $y \in V$. By assumption, and since f is continuous, there is an open neighborhood $\tilde{W}$ of y such that $f(\tilde{W}) \subset W$. By local compactness, there is an open neighborhood $V \subset Y$ of Y such that $y \in V \subset \bar{V} \subset \tilde{W}$ and $\bar{V}$ is compact. If we put $K := \bar{V}$ and $U = W$, then $e(T(K,U) \times V) \subset W$ since for $f' \in T(K,U)$ and $y' \in V$, we have $e(f', y') = f'(y') \subset W$. □

Let X, Y, Z be topological spaces. As sets, it is always true that $Z^{X \times Y} \cong Z^{Y^X}$ via the maps

$$\Phi : Z^{X \times Y} \to Z^{Y^X} f \mapsto (x \mapsto (y \mapsto f(x,y)))$$

and

$$\Psi : Z^{Y^X} \to Z^{X \times Y} g \mapsto ((x,y) \mapsto g(x)(y))$$

1.7.5 Theorem (The exponential law) *If Y is locally compact, then*

$$\Phi(\mathcal{C}(X \times Y), Z) \subset \mathcal{C}(X, \mathcal{C}(Y,Z))$$

and

$$\Psi(\mathcal{C}(X, \mathcal{C}(Y,Z))) \subset (\mathcal{C}(X \times Y), Z)$$

Proof. For $f \in \mathcal{C}(X \times Y, Z)$ and $x \in X$, we have to show that $\Phi(f)(x) \in \mathcal{C}(Y,Z)$ and $\Phi(f) \in \mathcal{C}(X, \mathcal{C}(Y,Z))$. $\Phi(f)(x)(y) = f \circ i_x(y) = f(x,y)$. Consider $T(K,U)$ for $K \subset Y$ compact and $U \subset X$ open. We need ot prove that the preimage $\Phi(f)^{-1}(T(K,U))$ is open in X. Let $x \in \Phi(f)^{-1}(T(K,U))$ so that $f(x,) \in T(K,U)$. Hence for all $y \in K$, we have $f(x,y) \in U$. By the continuity of f, there are open neighborhoods W_y of x and V_y of y such that $f(W_y \times V_y) \subset U$. Since K us compact, there are open sets $y_1, y_2, \ldots y_k \subset Y$ such that $K \subset V_{y_1} \cup V_{y_2} \cup \ldots \cup V_{y_k}$. Put $W = W_{y_1} \cap W_{y_2} \cap \ldots \cap W_{y_k}$ so that W is a neighborhood of x and $\Phi(f)(W) \subset T(K,U)$.

Now we need to show for $g \in \mathcal{C}(X, \mathcal{C}(Y,Z))$ that $\Psi(g) \in \mathcal{C}(X \times Y, Z)$. Let $g : X \times \mathcal{C}(Y,Z)$ be continuous and assume that $U \subset Z$ be open. We have to show that $\Psi(g)^{-1}(U)$ is open. Take $(x,y) \in \Psi(g)^{-1}(U)$. Since g is continuous, there is an open neighborhood W of y such that $g(x)(W) \subset U$. Since Y is locally compact, there is an open $V \subset Y$ such that $y \in V \subset \bar{V} \subset W$ with $\bar{V}$ compact. Hence $g(x)(V) \subset g(x)(\bar{V}) \subset U$. Thus $g(x) \in T(K,U)$ so there is an open neighborhood $O \subset X$ of x such that $g(O) \subset T(\bar{V}, U)$. Therefore

$$\Psi(g)(O \times V) \subset g(O)(V) \subset g(O)(\bar{V}) \subset U$$

1.7.6 Lemma *The sets $(U^L)^K = T(K, T(L,U))$ with $K \subset X$ and $L \subset Y$ compact and $U \subset Z$ open form a subbasis for the compact open topology on $\mathcal{C}(X, \mathcal{C}(Y,Z))$.*

Proof. Let I be an index set $W_i \subset \mathcal{C}(Y,Z)$ be open and $K \subset X$ be compact.

$$T\left(K, \bigcup_I W_i\right) = \bigcup_{n \in \mathbb{N}^+} \bigcup_{\substack{K_1 \times \ldots \times K_n \subset K^n \\ K_1 \cup \ldots \cup K_n = K \\ K_i = \bar{K}_i \forall i}} \bigcup_{(i_1, \ldots, i_n) \in I^n} \bigcap_{l=1}^{n} T(K_{i_l}, W_{i_l})$$

Suppose J is a finite set. then $T\left(K, \bigcap_{j \in J} W_j\right) = \bigcap_{j \in J} T(K, W_j)$. Sets of the form $T(L,U)$ with $L \subset Y$ compact and $U \subset Z$ open form a subbasis of $\mathcal{C}(Y,Z)$, so if $W \subset \mathcal{C}(Y,Z)$ is open, we have $W = \bigcup_{i \in I} \bigcap_{j \in J_i} T(L_{i_j}, U_{i_j})$ so that

$$T(K,W) = \bigcup_{n \in \mathbb{N}^+} \bigcup_{\substack{K_1 \times \ldots \times K_n \subset K^n \\ K_1 \cup \ldots \cup K_n = K \\ K_i = \bar{K}_i \forall i}} \bigcup_{(i_1, \ldots, i_n) \in J^n} \bigcap_{l=1}^{n} \bigcap_{j \in J_{i_l}} T(K_{i_l}, T(L_{i_l j}, U_{i_l j}))$$

1.7.7 Theorem *Let X, Y, Z be topological spaces with X and Y Hausdorff and Y locally compact. Then the natural isomorphism*

$$\bar{\Phi} : \mathcal{C}(X \times Y, Z) \to \mathcal{C}(X, \mathcal{C}(Y,Z))$$

is a homeomorphism.

Proof. Let $f \in \mathcal{C}(X \times Y, Z)$ and let $W \in \mathcal{C}(X, \mathcal{C}(Y,Z))$ be an open neighborhood of $\bar{\Phi}(f)$. By 1.7.6, there is an open $U \subset Z$ and compact subsets $L \subset Y$ and $K \subset X$ such that $p\bar{h}i(f) \in T(K, T(L,U)) \subset W$. $T(K \times L, U)$ is open in $\mathcal{C}(X \times Y, Z)$ and note that $f \in T(K \times L, U)$ since for $(x,y) \in K \times L$, $\bar{\Phi}(f)(x) \in T(L,U)$ and $f(x,y) = \bar{\Phi}(f)(x)(y) \in U$.

Assume that $g \in T(K \times L, U)$. The $\bar{\Phi}(g)(x)(y) = g(x,y) =\in U$ so $\bar{\Phi}(g)(x) \in T(L,U)$ so $\bar{\Phi}(g) \in T(K, T(L,U))$, hence $\bar{\Phi}$ is continuous. Rest of proo

email 9/27/10

II.2. Sheaf theory

2.1. Presheaves

The category of open sets of a topological space

2.1.1 Before we define presheaves and sheaves on a topological space $(X, \mathcal{T})$, we briefly introduce the category $\mathsf{Ouv}(X)$ of open sets of $(X, \mathcal{T})$. By definition, its object class coincides with the set of open sets $\mathcal{T}$, so $\mathsf{Ouv}(X)$ is in particular a small category. For two open $U, V \subset X$ the morphism set $\mathrm{Mor}_{\mathsf{Ouv}(X)}(U, V)$ is defined to be empty in case $U \not\subset V$ and consists of the canonical (identical) embedding $i_U^V : U \hookrightarrow V$ when $U \subset V$. Obviously, the identity map i_U^U is then a morphism for every open $U \subset X$, and the composition of morphisms in this category is given by

$$i_V^W \circ i_U^V = i_U^W : U \hookrightarrow W \quad \text{for } U, V, W \in \mathcal{T} \text{ with } U \subset V \subset W.$$

This observation entails that $\mathsf{Ouv}(X)$ is a category indeed; it is called the *category of open sets* on the topological space $(X, \mathcal{T})$.

2.1.2 Remarks (a) The topology $\mathcal{T}$ carries a natural partial order given by set-theoretic inclusion, so becomes a poset. The corresponding category structure from Example 1.2.11 is canonically isomorphic to $\mathsf{Ouv}(X)$.

(b) The notation Ouv stems from the French word 'ouvert' for 'open'.

2.1.3 Proposition *Let $(X, \mathcal{T})$ be a topological space. Then the category $\mathsf{Ouv}(X)$ has the following properties.*

(i) *The empty set $\varnothing$ is an initial object in $\mathsf{Ouv}(X)$, the full set X a final object.*

(ii) *Fibered products exist in $\mathsf{Ouv}(X)$. More precisely, if $i_U^W : U \hookrightarrow W$ and $i_V^W : V \hookrightarrow W$ are two morphisms in $\mathsf{Ouv}(X)$, the fibered product $U \times_W V$ is given by the open set $U \cap V$ together with the canonical embeddings $i_{U \cap V}^U : U \cap V \hookrightarrow U$ and $i_{U \cap V}^V : U \cap V \hookrightarrow V$.*

(iii) *Directed colimits exist in $\mathsf{Ouv}(X)$.*

(iv) *Finite limits exist in $\mathsf{Ouv}(X)$.*

Proof. *ad (i)*. The first claim follows from the fact that $\varnothing$ is contained in every element of $\mathcal{T}$ and that every element of $\mathcal{T}$ is contained in X.

ad (ii). Assume to be given $O \in \mathcal{T}$ such that the following diagram commutes:

$$\begin{tikzcd} O \arrow[r] \arrow[d] & V \arrow[d] \\ U \arrow[r] & W \end{tikzcd}$$

Then $O \subset U \cap V$, and the diagram

$$\begin{tikzcd} O \arrow[dr] \arrow[drr, bend left] \arrow[ddr, bend right] & & \\ & U \cap V \arrow[r] \arrow[d] & V \arrow[d] \\ & U \arrow[r] & W \end{tikzcd}$$

commutes with morphisms unique.

ad (iii). Assume that $(I, \leqslant)$ is an upward directed set and $(U_i)_{i \in I}$ a directed system in $\mathsf{Ouv}(X)$. □

The category of presheaves on a topological space

2.1.4 Definition Let C denote a category and $(X, \mathcal{T})$ be a topological space. By a *presheaf* on X with *values* in C one understands a contravariant functor $\mathcal{F} : \mathsf{Ouv}(X) \to \mathsf{C}$. A *morphism* between two presheaves $\mathcal{F} : \mathsf{Ouv}(X) \to \mathsf{C}$ and $\mathcal{G} : \mathsf{Ouv}(X) \to \mathsf{C}$ is a natural transformation $\eta : \mathcal{F} \to \mathcal{G}$. The functor category $\mathsf{Func}\big(\mathsf{Ouv}(X)^{\mathrm{op}}, \mathsf{C}\big)$ is called the *category of presheaves* on X with *values* in C. It is denoted by $\mathsf{PSh}_{\mathsf{C}}(X)$. By definition, its objects are the presheaves on X, its morphisms are given by morphisms of presheaves. When the underlying category is the category of sets we just write $\mathsf{PSh}(X)$ instead of $\mathsf{PSh}_{\mathsf{Ens}}(X)$.

Types of algebraic structures

In general, the category C in which a presheaf $\mathcal{F}$ takes its values comes equipped with some type of additional, usually algebraic structure. Not only is this suggested by examples, also the theory of presheaves becomes richer by that. We want to specify the kind of categories which presheaves are allowed to take values in.

2.1.5 Definition () By a *type of algebraic structure*

The étalé space of a presheaf

2.1.6 Let $\mathcal{F}$ be a presheaf on the topological space $(X, \mathcal{T})$ with values in the category C. We assume that directed colimits exist in C. Given a point $x \in X$, the system $\mathcal{N}_x^\circ$ of open neighborhoods of x is (upward) directed by defining $V \leqslant U$ for $U, V \in \mathcal{N}_x^\circ$ if $U \subset V$. Since $\mathcal{F}$ is a contravariant functor from $\mathsf{Ouv}(X)$ to C, the diagram $(\mathcal{F}(U))_{U\in\mathcal{N}_x^\circ}$ is a directed system in C, hence has a colimit by assumption. We write

$$\mathcal{F}_x := \operatorname*{colim}_{U\in\mathcal{N}_x^\circ} \mathcal{F}(U)$$

and call $\mathcal{F}_x$ the *stalk* of $\mathcal{F}$ at x. For every open neighborhood U of x we denote the image of a section $s \in \mathcal{F}(U)$ in the stalk $\mathcal{F}_x$ by $[s]_x$ or s_x and call it the *germ* of s at x. When denoting a germ at x by s_x we therefore always silently assume that s is an element of some section space $\mathcal{F}(U)$ over an open neighborhood U of x and that that s is a representative of the germ considered. Next we study the disjoint union of the stalks

$$\text{Ét}(\mathcal{F}) := \bigsqcup_{x\in X} \mathcal{F}_x := \bigcup_{x\in X} \mathcal{F}_x \times \{x\} \ .$$

By construction, we have a natural projection map $\pi^+ : \text{Ét}(\mathcal{F}) \to X$ which maps an element $(s_x, x) \in \text{Ét}(\mathcal{F})$ to the footpoint $x \in X$. Given a section $s \in \mathcal{F}(U)$ defined over an open set $U \subset X$ we define a map $s^+ : U \to \text{Ét}(\mathcal{F})$ by

$$s^+(x) = (s_x, x) \quad \text{for all } x \in U \ .$$

Now we make the following observation.

2.1.7 Proposition and Definition *Assume that $(X, \mathcal{T})$ is a topological space and C a category in which directed colimits exist. Let $\mathcal{F}$ be a presheaf on X with values in C. Then the set $\mathcal{B}^+$ consisting of all sets of the form $s^+(U)$ with $U \subset X$ open and $s \in \mathcal{F}(U)$ is the basis of a topology $\mathcal{T}^+$ on $\text{Ét}(\mathcal{F})$. One calls the topological space $\big(\text{Ét}(\mathcal{F}), \mathcal{T}^+\big)$ the* étalé space *of the presheaf $\mathcal{F}$.*

Proof. First observe that $\mathcal{B}$ covers $\text{Ét}(\mathcal{F})$ since every element of $\text{Ét}(\mathcal{F})$ is of the form (s_x, x) for some open $U \subset X$ and section $s \in \mathcal{F}(U)$ and since (s_x, x) is contained in the basis element $s^+(U)$.

It remains to show that the intersection of two elements $s^+(U)$ and $t^+(V)$ of $\mathcal{B}^+$ can be written as the union of elements of $\mathcal{B}^+$. So let $(u_x, x) \in s^+(U) \cap t^+(V)$. The germ u_x is represented by some section $u \in \mathcal{F}(W)$ over some open neighborhood W of x. By assumption and since both s^+ and t^+ are sections of π^+ one obtains

$$(u_x, x) = s^+(x) = t^+(x) \ .$$

Hence there exists an open neighborhood $O \subset U \cap V \cap W$ of x such that

$$u|_O = s|_O = t|_O \ .$$

But that entails

$$(u_x, x) \in u^+(O) = s^+(O) = t^+(O) \subset s^+(U) \cap t^+(V) \ .$$

This proves the claim. □

2.1.8 Corollary *The canoncial projection $\pi^+ : \text{Ét}(\mathcal{F}) \to X$ of a presheaf $\mathcal{F}$ on the topological space $(X, \mathcal{T})$ is an étale map or in other words a local homeomorphism. More precisely, for every open $U \subset X$ and every $s \in \mathcal{F}(U)$ the section $s^+ : U \to \text{Ét}(\mathcal{F})$ is a homeomorphism onto its image.*

Proof. We first prove that π^+ is continuous. So let $U \subset X$ open. Then observe that

$$(\pi^+)^{-1}(U) = \bigcup_{x \in U} \mathcal{F}_x \times \{x\} = \bigcup_{x \in U} \bigcup_{\substack{V \in \mathcal{N}_x^\circ \\ V \subset U}} \bigcup_{s \in \mathcal{F}(V)} \{(s_x, x)\} = \bigcup_{\substack{V \in \mathsf{Ouv}(X) \\ V \subset U}} \bigcup_{s \in \mathcal{F}(V)} s^+(V) \ .$$

But this is open by the preceding proposition, hence π^+ is continuous.

Next we show continuity of s^+. To this end it suffices to verify that $(s^+)^{-1}\big(t^+(V)\big)$ is open in U for every open $V \subset X$ and $t \in \mathcal{F}(V)$. Now compute

$$\begin{aligned} (s^+)^{-1}\big(t^+(V)\big) &= \big\{x \in U \bigm| s^+(x) \in t^+(V)\big\} = \big\{x \in U \bigm| x \in V \ \&\ s^+(x) = t^+(x)\big\} \\ &= \big\{x \in U \cap V \bigm| s_x = t_x\big\} \ . \end{aligned}$$

The right hand side is open since if it contains y, then the sections s and t coincide on an open neighborhood $O \subset U \cap V$ of y which means that $O \subset \big\{x \in U \cap V \bigm| s_x = t_x\big\}$.

Finally observe that s^+ is a section of π^+ and that

$$s^+ \circ \pi^+|_{s^+(U)} = \mathrm{id}_{s^+(U)}$$

which implies that s^+ is a homeomorphism onto its image. The claim now follows from the fact that the sets $s^+(U)$ with $U \in \mathsf{Ouv}(X)$ and $s \in \mathcal{F}(U)$ cover $\acute{\mathrm{E}}\mathrm{t}(\mathcal{F})$. $\square$

II.3. Basic homotopy theory

3.1. Homotopy categories of topological spaces

Notational preliminaries

3.1.1 In this chapter, we always denote by I the compact unit interval $[0,1] \subset \mathbb{R}$ and by ∂I ist boundary in $\mathbb{R}$ that is the set $\{0,1\}$. Morever, for every topological space X and $t \in I$ we denote by $j_{X,t}$ or shortly by j_t, if no confusion can arise, the map $X \to X \times I$, $x \mapsto (x,t)$. The evaluation map $\mathcal{C}(I,X) \to X$, $\gamma \mapsto \gamma(t)$ will be abbreviated by $e_{X,t}$ or shortly by e_t.

Homotopies

3.1.2 Definition Let $f : X \to Y$ and $g : X \to Y$ be continuous maps between two given topological spaces X and Y. A *homotopy* from f to g is a continuous map $H : X \times I \to Y$ such that $H_0 = f$ and $H_1 = g$ where for every $t \in I$ the symbol H_t stands for the composition $H \circ j_t$ or in other words the map $H_t : X \to Y$, $x \mapsto H(x,t)$.

If (X,A) is a topological pair, a continuous map $H : X \times I \to Y$ is called a *homotopy from X to Y relative A* if $H(a,t) = H(a,0)$ for all $a \in A$ and $t \in I$.

Then one says that f is *homotopic to g* if there is a homotopy $H : X \times I \to Y$ such that $H_0 = f$ and $H_1 = g$. This will be denoted by $f \simeq g$ or $H : f \simeq g$ and we will often say that H is a *homotopy between f and g* or a *homotopy from f to g*.

Let $A \subset X$ be a subspace. Then one says that $f : X \to Y$ is *homotopic to $g : X \to Y$ relative A* if there is a homotopy relative A between f and g that means a homotopy $H : X \times I \to Y$ relative A such that $H_0 = f$ and $H_1 = g$. One denotes this by $f \simeq_A g$, $f \simeq g$ rel A, $H : f \simeq_A g$, or $H : f \simeq g$ rel A.

3.1.3 Remarks (a) Observe that homotopy relative $\varnothing$ is usual homotopy.

(b) If f, g are maps from X to Y such that for some subspace $A \subset X$ the relation $f \simeq_A g$ holds true, then $f|_A = g|_A$.

3.1.4 Proposition *Let X and Y be topological spaces and supposed that A is a subspaces of X. Homotopy relative A then is an equivalance relation on the space $\mathcal{C}(X,Y)$ of continuous functions from X to Y.*

Proof. Claim 1. The relation $\simeq_A$ is symmetric. Let $f, g : X \to Y$ be continuous maps. If $H : X \times I \to Y$ is a homotopy relative A from f to g, then $H^- : X \times I \to Y$ defined by $H^-(x,t) := H(x, 1-t)$ is a homotopy relative A from g to f.

Claim 2. The relation $\simeq_A$ is reflexive. Let $f : X \to Y$ be a continuous map. The map $F : X \times I \to Y$ defined by $F(x,t) := f(x)$ is a homotopy relative A from f to f.

Claim 3. The relation $\simeq_A$ is transitive. Let $f, g, h : X \to Y$ be continuous maps such that there exist homotopies $F : f \simeq_A g$ and $G : g \simeq_A h$. Define $H : X \times I \to Y$ by

$$H(x,t) := \begin{cases} F(x, 2t), & \text{if } 0 \leqslant t \leqslant \frac{1}{2}, \\ G(x, 2t-1), & \text{if } \frac{1}{2} \leqslant t \leqslant 1. \end{cases}$$

Since $F(x,1) = g(x) = G(x,0)$ for all $x \in X$, the map H is well-defined and continuous. Moreover, $H_0 = F_0 = f$, $H_1 = G_1 = h$, and $H(a,t) = H(a, \frac{1}{2}) = g(a)$ for all $a \in A$, so that H is a homotopy relative A between f and h. □

3.1.5 The equivalence class of a continuous map $f : X \to Y$ with respect to homotopy relative A will be denoted $[f]_{\simeq_A}$ respectively by $[f]_{\simeq}$ or $[f]$ if $A = \varnothing$. The set of equivalence classes $[f]_{\simeq_A}$ will be denoted $[X,Y]_A$. For ease of notation, one puts $[X,Y] := [X,Y]_{\varnothing}$.

3.1.6 Proposition *Let $f_1, f_2 : X \to Y$ and $g_1, g_2 : Y \to Z$ be continuous. If $F : f_1 \simeq f_2$ and $G : g_1 \simeq g_2$, then $g_1 \circ f_1 \simeq g_2 \circ f_2$. In other words, homotopy is a natural equivalance relation on the morphisms of the category of topological spaces.*

Proof. Construct $H : g_1 \circ f_1 \simeq g_2 \circ f_2$ by $H(x,t) := G(F(x,t),t)$ for all $x \in X, t \in I$. This function is continuous because F and G are. Furthermore,

$$H(x,0) = G(F(x,0),0) = G(f_1(x),0) = g_1(f_1(x))$$

and

$$H(x,1) = G(F(x,1),1) = G(f_2(x),1) = g_2(f_2(x))$$

so we have the desired homotopy. □

Now that we have an equivalance relation on the morphisms of Top which is compatible with composition we can define a corresponding quotient category.

3.1.7 Definition The *homotopy category of topological spaces*, hTop, is the category with objects being topological spaces and morphisms being homotopy classes of continuous maps.

3.1.8 Remark Let $f : X \to Y$ and $g : Y \to Z$ be two continuous maps. Then we denote the composition of $[g]$ and $[f]$ in hTop by $[g]\,[f]$, so we usually omit the symbol $\circ$ for composition in the homotopy category. Note that by definition $[g]\,[f] = [g \circ f]$.

3.1.9 Example Let $C \in \mathbb{R}^n$ be a convex set, $* \in C$ be a point, and consider the maps

$$i : \{*\} \to C,\ * \mapsto * \quad \text{and} \quad p : C \to \{*\},\ x \mapsto * \,.$$

If $C \neq \{*\}$, these maps are not bijective and so are not isomorphisms in Top. However, $[p][i] = [p \circ i] = [\mathrm{id}_{\{*\}}]$, and there is a homotopy

$$H : C \times I \to C,\ (x,t) \mapsto * + t(x - *)$$

such that $H_0 = i \circ p$ and $H_1 = \mathrm{id}_C$. Hence $[i]$ and $[p]$ are isomorphisms in the category hTop.

3.1.10 Definition A continuous function $f : X \to Y$ is called a *homotopy equivalance* if $[f]$ is an isomorphism in hTop that is if there is a continuous map $g : Y \to X$ such that $[f]\,[g] = [\mathrm{id}_Y]$ and $[g]\,[f] = [\mathrm{id}_X]$. Two spaces are called homotopy equivalant if there is a homotopy equivalance $f : X \to Y$. A space X is called *contractible* if X is homotopy equivalent to a one point space.

3.2. Covering spaces

Definitions and first properties

3.2.1 Definition Let $p : \widetilde{X} \to X$ be a continuous map between topological spaces. An open subset $U \subset X$ is called *evenly covered* if the preimage $p^{-1}(U) \subset \widetilde{X}$ is the disjoint union of open subsets $\widetilde{U}_\alpha \subset \widetilde{X}$, $\alpha \in A$, such that for each $\alpha \in A$ the restriction $p|_{\widetilde{U}_\alpha} : \widetilde{U}_\alpha \to U$ is a homeomorphism.

A surjective continuous map $p : \widetilde{X} \to X$ is called a *covering* (*of* X), a *covering map* or a *covering projection* and $\widetilde{X}$ a *covering space of* X if each point of X has an evenly covered open neighborhood. The space X is called the *base space* of the covering. The preimage $p^{-1}(x)$ of a point $x \in X$ is called the *fiber over* x, and a point $\widetilde{x} \in p^{-1}(x)$ is said to *lie over* x.

If $p : \widetilde{X} \to X$ and $q : \widetilde{Y} \to Y$ are two coverings, a *morphism of coverings* from $p : \widetilde{X} \to X$ to $q : \widetilde{Y} \to Y$ is a pair (Φ, f) of continuous maps $\Phi : \widetilde{X} \to \widetilde{Y}$ and $f : X \to Y$ such that the diagram

$$\begin{tikzcd} \widetilde{X} \arrow[r, "\Phi"] \arrow[d, "p"'] & \widetilde{Y} \arrow[d, "q"] \\ X \arrow[r, "f"'] & Y \end{tikzcd} \tag{3.2.1}$$

commutes. We express that (Φ, f) is morphism of coverings from p to q by the notation $(\Phi, f) : p \to q$. In case $p : \widetilde{X} \to X$ and $q : \widehat{X} \to X$ are two coverings of X, a continuous map $\Phi : \widetilde{X} \to \widehat{X}$ is called a *morphism of coverings of* X from p to q if $q \circ \Phi = p$. In other words this means that the diagram

$$\begin{tikzcd} \widetilde{X} \arrow[rr, "\Phi"] \arrow[dr, "p"'] & & \widehat{X} \arrow[dl, "q"] \\ & X & \end{tikzcd}$$

commutes. A homeomorphism $\Phi : \widetilde{X} \to \widetilde{X}$ which satisfies $p \circ \Phi = p$ is called a *deck transformation* of p.

Let (X, x_0) be a pointed topological space. A morphism $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ of pointed topological spaces then is called a *pointed covering* (*of* (X, x_0)) if $p : \widetilde{X} \to X$ is a covering map.

A *morphism of pointed coverings* from a pointed covering $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ to a pointed covering $q : (\widetilde{Y}, \widetilde{y}_0) \to (Y, y_0)$ is a pair (Φ, f) of morphism of pointed spaces $\Phi : (\widetilde{X}, \widetilde{x}_0) \to (\widetilde{Y}, \widetilde{y}_0)$ and $f : (X, x_0) \to (Y, y_0)$ such that Diag. (3.2.1) commutes. Finally, if $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ and $q : (\widehat{X}, \widehat{x}_0) \to (X, x_0)$ are two pointed coverings of (X, x_0), then a morphism of pointed spaces $\Phi : (\widetilde{X}, \widetilde{x}_0) \to (\widehat{X}, \widehat{x}_0)$ which satisfies $q \circ \Phi = p$ is called a *morphism of pointed coverings of* (X, x_0).

3.2.2 Proposition and Definition *Covering spaces as objects together with their morphisms form a category* Cov*, called the* category of covering spaces*. For a topological space* X *the class of coverings of* X *as objects together with the morphisms of coverings of* X *form another category denoted by* $\mathsf{Cov}(X)$*. It is called the* category of covering spaces of X*. The automorphisms of* $\mathsf{Cov}(X)$ *are the deck transformations.*

The category $\mathsf{Cov}(X)$ *can be understood as a subcategory of* Cov *via the functor which is the identical embedding on the object class and which assigns to every morphism* $\Phi : \widetilde{X} \to \widehat{X}$ *of coverings* $p : \widetilde{X} \to X$ *and* $q : \widehat{X} \to X$ *of* X *the morphism* $(\Phi, \mathrm{id}_X) : p \to q$ *in* Cov*.*

Pointed coverings together with their morphisms form a category which is denoted by $\mathsf{Cov}_\bullet$*.*

Finally, given a pointed space (X, x_0)*, the pointed coverings of* (X, x_0) *together with their morphisms form a category* $\mathsf{Cov}_\bullet(X, x_0)$ *which in a canonical way is a subcategory of* $\mathsf{Cov}_\bullet$*.*

Proof. Clearly, if $p : \widetilde{X} \to X$ is a covering, the pair $(\mathrm{id}_{\widetilde{X}}, \mathrm{id}_X)$ is a morphism of covering spaces from p to p. If $q : \widetilde{Y} \to Y$ and $r : \widetilde{Z} \to Z$ are further coverings, and $(\Phi, f) : p \to q$ and $(\Psi, g) : q \to r$ morphisms, then the pair $(\Psi \circ \Phi, g \circ f)$ is a morphism from p to r, since the diagram

$$\begin{tikzcd} \widetilde{X} \arrow[r, "\Phi"] \arrow[d, "p"'] & \widetilde{Y} \arrow[r, "\Psi"] \arrow[d, "q"] & \widetilde{Z} \arrow[d, "r"] \\ X \arrow[r, "f"'] & Y \arrow[r, "g"'] & Z \end{tikzcd}$$

commutes. Since composition of functions is associative, it now follows that covering spaces together with their morphisms form a category indeed.

For a covering $p : \widetilde{X} \to X$, the map $\mathrm{id}_{\widetilde{X}}$ is obviously a morphism of covering spaces of X from p to p. Moreover, if $\Phi : \widetilde{X} \to \widehat{X}$ and $\Psi : \widehat{X} \to \check{X}$ are morphisms of covering spaces of X from $p : \widetilde{X} \to X$ to $q : \widehat{X} \to X$ and from $q : \widehat{X} \to X$ to $r : \check{X} \to X$, respectively, then the composition $\Psi \circ \Phi$ is a morphism of covering spaces of X from p to r, since $r \circ \Psi \circ \Phi = \ q \circ \Phi = p$. So $\mathsf{Cov}(X)$ forms a category indeed.

The remainder of the claim is obvious. □

3.2.3 One of the main goals of this chapter is to show that the category of pointed coverings of a pointed topological space (X, x_0) has, under mild assumptions on X, an initial object. Such an initial object is called a universal cover of (X, x_0). More precisely:

3.2.4 Definition Let (X, x_0) be a pointed topolological space. A pointed covering $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ is called a *universal (pointed) covering of* (X, x_0) if $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ is an initial object in $\mathsf{Cov}_\bullet(X, x_0)$ which in other words means that it satisfies the following universal property:

(UCov) For every covering $q : (\widehat{X}, \widehat{x}_0) \to (X.x_0)$ there exists a unique morphism $\Phi : (\widetilde{X}, \widetilde{x}_0) \to (\widehat{X}, \widehat{x}_0)$ of pointed covering spaces from p to q.

A covering $p : \widetilde{X} \to X$ of a topological space X is said to be a *universal covering of* X and $\widetilde{X}$ a *universal covering space* or a *universal cover of* X if for every $\widetilde{x}_0 \in \widetilde{X}$ the pointed map $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ with $x_0 := p(\widetilde{x}_0)$ is a universal pointed covering of (X, x_0).

3.2.5 Remark We will later see that for reasonable spaces X the condition that $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ is a universal covering for every $\widetilde{x}_0 \in \widetilde{X}$ and $x_0 = p(\widetilde{x}_0)$ holds true already if it is satisfied for one element $\widetilde{x}_0 \in \widetilde{X}$.

The crucial property of covering spaces from a homotopy theoretic point of view is that they possess the homotopy lifting property. Before we come to formulate this let us explain what "lifting" means.

3.2.6 Definition Let $p : \widetilde{X} \to X$ be a covering space, and $f : Y \to X$ a continuous map. A continuous map $\widetilde{f} : Y \to \widetilde{X}$ is then called a *lifting* of f if the diagram

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ Y \arrow[ur, "\widetilde{f}"] \arrow[r, "f"'] & X \end{tikzcd}$$

commutes.

3.2.7 Theorem (Homotopy lifting property of covering spaces) *Let $p : \widetilde{X} \to X$ be a covering space, $f : Y \to X$ a continuous map, and $H : Y \times I \to X$ a homotopy with $H_0 = f$. If $\widetilde{f} : Y \to \widetilde{X}$ is a lifting of f_0, then there exists a unique homotopy $\widetilde{H} : Y \times I \to \widetilde{X}$ lifting H such that $\widetilde{H}_0 = \widetilde{f}$.*

3.2.8 Corollary *Let $\gamma : I \to X$ be a path in the topological space X and $x_0 := \gamma(0)$. If $p : \widetilde{X} \to X$ is a covering of X and $\widetilde{x}_0 \in \widetilde{X}$ a point lying over x_0, then there exists a unique path $\widetilde{\gamma} : I \to \widetilde{X}$ which lifts γ and satisfies $\widetilde{\gamma}(0) = \widetilde{x}_0$.*

Fiber bundles and covering spaces

Covering spaces are closely related to fiber bundles. Before we come to stating that relation, let us recall the definition of a fiber bundle.

3.2.9 Definition A quadruple (E, B, p, F) consisting of topological spaces E, B, and F and a continuous map $p : E \to B$ is called a *fiber bundle* if for every point $x \in B$ there exists an open neighborhood $U \subset B$ of x called *trivializing neighborhood* together with a homeomorphism $\varphi : p^{-1}(U) \to U \times F$ such that the diagram

$$\begin{tikzcd} p^{-1}(U) \arrow[r, "\varphi"] \arrow[d, "p"'] & U \times F \arrow[dl, "\mathrm{pr}_1"] \\ U & \end{tikzcd}$$

commutes, where $\mathrm{pr}_1 : U \times F \to U$ is projection onto the first factor. The homeomorphism is called a *local trivialization of E (over U)*. The space E is called the *total space* of the fiber bundle, B the *base*, F the *typical fiber*, and $p : E \to B$ the *projection*.

The major ingrediant for proving that covering spaces and fiber bundles with discrete typical fibers correspond to each other is the following result.

3.2.10 Lemma *Let $p : \widetilde{X} \to X$ be a continuous map and $U \subset X$ an open subset. Then the following are equivalent:*

(i) *The open set U is evenly covered.*

(ii) *There exists a discrete topological space F and a homeomorphism $\varphi : p^{-1}(U) \to U \times F$ such that $p|_{p^{-1}(U)} = \mathrm{pr}_1 \circ \varphi$, where $\mathrm{pr}_1 : U \times F \to U$ denotes projection onto the first factor.*

Proof. Assume U to be evenly covered, and let $(\widetilde{U}_\alpha)_{\alpha \in A}$ be the family of pairwise disjoint open subsets of $\widetilde{X}$ such that $p^{-1}(U) = \bigcup_{\alpha \in A} \widetilde{U}_\alpha$ and such that $p|_{\widetilde{U}_\alpha} : \widetilde{U}_\alpha \to U$ is a homeomorphism for every $\alpha \in A$. Put $F := A$ and give F the discrete topology. Define $\varphi(\widetilde{x})$ for $\widetilde{x} \in p^{-1}(U)$ as the pair $(x, \alpha) \in U \times F$, where $x = p(\widetilde{x})$ and α is the unique element of F such that $\widetilde{x} \in \widetilde{U}_\alpha$. Then $p|_{p^{-1}(U)} = \mathrm{pr}_1 \circ \varphi$ by construction. Obviously, φ is continuous since $\mathrm{pr}_1 \circ \varphi$ is continuous and $\mathrm{pr}_2 \circ \varphi$ locally constant. Moreover, φ is invertible with inverse given by $U \times F \to p^{-1}(U)$, $(x, \alpha) \mapsto \left(p|_{\widetilde{U}_\alpha}\right)^{-1}(x)$. The inverse map is continuous as well, since each of the $p|_{\widetilde{U}_\alpha}$, $\alpha \in A$, is a homeomorphism. This proves (ii).

Now assume that for the given U there exists a discret topological space F and a homeomorphism $\varphi : p^{-1}(U) \to U \times F$ such that $p|_{p^{-1}(U)} = \mathrm{pr}_1 \circ \varphi$. Put $A := F$ and $\widetilde{U}_\alpha := \varphi^{-1}(U \times \{\alpha\})$ for each $\alpha \in A$. Since F carries the discrete topology, each $\widetilde{U}_\alpha$ is open in $\widetilde{X}$. Moreover, $p^{-1}(U)$ is the disjoint union of the $\widetilde{U}_\alpha$, $\alpha \in A$. The restriction $\varphi|_{\widetilde{U}_\alpha} : \widetilde{U}_\alpha \to U \times \{\alpha\}$ now is a homeomorphism, and pr_1 is both open and continuous. Hence $p|_{\widetilde{U}_\alpha} = \mathrm{pr}_1 \circ \varphi|_{\widetilde{U}_\alpha}$ is a homeomorphism as well, and (i) is proved. □

3.2.11 Proposition *Let X be a connected topological space and $p : \widetilde{X} \to X$ a surjective continuous map. Then the following are equivalent:*

(i) *$p : \widetilde{X} \to X$ is a covering.*

(ii) *There exists a discrete topological space F such that $(\widetilde{X}, X, p, F)$ becomes a fiber bundle.*

(iii) *The map p is a local homeomorphism, and there exists a topological space F such that $(\widetilde{X}, X, p, F)$ becomes a fiber bundle.*

Construction of the universal covering

Not every topological space has a universal cover as the following example shows.

3.2.12 Example

But for a large class of topological spaces, which in particular contains all smooth manifolds, one can construct a universal cover. Before we can formulate this precisely, we need one more concept.

3.2.13 Definition A topological space is called *semi-locally simply-connected* if each point $x \in X$ has a neighborhood N such that the group homomorphism $\pi_1(N, x) \to \pi_1(X, x)$ is trivial.

3.2.14 Example A topological space which is not semi-locally simply connected is the hawaiian earring.

3.2.15 Theorem *Let X be a connected locally path-connected and semi-locally simply connected topological space. Then X has a universal cover $p : \widetilde{X} \to X$ with $\widetilde{X}$ being locally path-connected and simply-connected.*

Proof. First fix a point $x_0 \in X$ and let $\widetilde{X}$ be the set of all homotopy classes (relative endpoints) of paths in X starting at x_0. That means

$$\widetilde{X} := \big\{[\gamma] \in \pi_1(X) \mid \gamma(0) = x_0\big\} .$$

The projection $p : \widetilde{X} \to X$ is defined as the map $[\gamma] \mapsto \gamma(1)$. Next we define for every non-empty path-connected open $U \subset X$ and every $[\gamma] \in \widetilde{X}$ with $\gamma(1) \in U$ the set $U_{[\gamma]}$ by

$$U_{[\gamma]} := \big\{[\gamma * \eta] \in \widetilde{X} \mid \eta \in \mathcal{C}(I, U) \;\&\; \eta(0) = \gamma(1)\big\} \qquad \square$$

and let $\mathcal{B}$ be the set of all such $U_{[\gamma]}$. Now we prove several claims.

Claim 1. Let $U \subset X$ be open and path-connected. Then for all elements $[\gamma], [\gamma'] \in \widetilde{X}$ with $p([\gamma]) \in U$ and $p([\gamma']) \in U$ the intersection $U_{[\gamma]} \cap U_{[\gamma']}$ is either empty or $U_{[\gamma]} = U_{[\gamma']}$.
To verify the first claim assume that $[\varrho] \in U_{[\gamma]} \cap U_{[\gamma']}$. Then there exist paths $\mu, \mu' : I \to U$ such that $\mu(0) = \gamma(1)$, $\mu'(0) = \gamma'(1)$, and $[\varrho] = [\gamma * \mu] = [\gamma' * \mu']$. This implies $\varrho(1) = \mu(1) = \mu'(1)$ and the relation $[\gamma'] = [\varrho * (\mu')^{-1}] = [\gamma * \mu * (\mu')^{-1}]$. Since $\mu * (\mu')^{-1}$ is a path in U with $\mu * (\mu')^{-1}(0) = \mu(0) = \gamma(1)$, one concludes that $[\gamma'] \in U_{[\gamma]}$. Hence if $\eta' : I \to U$ is a path with $\eta'(0) = \gamma'(1)$, then $[\gamma' * \eta'] = [\gamma * \mu * (\mu')^{-1} * \eta'] \in U_{[\gamma]}$ which shows that $U_{[\gamma']} \subset U_{[\gamma]}$. By symmetry, one obtains $U_{[\gamma]} \subset U_{[\gamma']}$, so our first claim is proved.

Claim 2. The set $\mathcal{B}$ of all U as defined above is a basis of a topology on $\widetilde{X}$.
To verify this, let $U_{[\gamma]}, V_{[\varrho]} \in \mathcal{B}$ and assume that $[\mu] \in U_{[\gamma]} \cap V_{[\varrho]}$. Then $\mu(1) \in U \cap V$.

3.3. The fundamental groupoid of a topological space

The fundamental group

3.3.1 Definition Let (X, x_0) be a pointed topological space. The *fundamental group* of (X, x_0) is defined as

$$\pi_1(X, x_0) := \mathcal{C}\big((I, \partial I\}), (X, x_0)\big)/\simeq_{\partial I}$$

which in other words is the set of homotopy classes relative $\partial I = \{0, 1\}$ of closed continuous paths $\gamma : I \to X$ based at x_0.

3.3.2 To turn $\pi_1(X, x_0)$ into a group need to define a binary operation

$$* : \pi_1(X, x_0) \times \pi_1(X, x_0) \to \pi_1(X, x_0) .$$

To this end let $\gamma, \varrho : I \to X$ be closed paths in X based at x_0. Define their *concatenation* $\varrho * \gamma : I \to X$ by

$$\varrho * \gamma(s) := \begin{cases} \gamma(2s) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \varrho(2s - 1) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

Then $\varrho * \gamma$ is again a closed path based at x_0. Its homotopy class depends only on the homotopy classes of γ and ϱ. Namely, if $H : \gamma \simeq_{\partial I} \widetilde{\gamma}$ and $G : \varrho \simeq_{\partial I} \widetilde{\varrho}$ are homotopies, then

$$G * H : I \times I \to X, \quad (s,t) \mapsto \begin{cases} H(2s,t) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ G(2s-1,t) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

is a homotopy relative ∂I from $\varrho * \gamma$ to $\widetilde{\varrho} * \widetilde{\gamma}$. Hence the map

$$* : \pi_1(X,x_0) \times \pi_1(X,x_0) \to \pi_1(X,x_0), \quad [\varrho]_{\partial I} * [\gamma]_{\partial I} := [\varrho * \gamma]_{\partial I}$$

is well-defined.

3.3.3 Theorem *The fundamental group $\pi_1(X,x_0)$ of a based topological space (X,x_0) is a group with binary operation given by concatenation of paths. The homotopy class of the constant path $e_{x_0} : I \to X$, $s \mapsto x_0$ acts as identity, and the inverse of an element $[\gamma]_{\partial I}$ is the homotopy class of the path $\gamma^{-1} : I \to X$, $s \mapsto \gamma(1-s)$.*

Proof. First we show associativity of $*$. Let $\gamma, \varrho, \eta : I \to X$ be closed paths in X based at x_0. Then the paths $\eta * (\varrho * \gamma) : I \to X$ and $(\eta * \varrho) * \gamma : I \to X$ are given by

$$\big(\eta * (\varrho * \gamma)\big)(s) = \begin{cases} \gamma(4s) & \text{for } 0 \leqslant s \leqslant \frac{1}{4}, \\ \varrho(4s-1) & \text{for } \frac{1}{4} \leqslant s \leqslant \frac{1}{2}, \\ \eta(2s-1) & \text{for } \frac{1}{2} \leqslant s \leqslant 1, \end{cases}$$

and

$$\big((\eta * \varrho) * \gamma\big)(s) = \begin{cases} \gamma(2s) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \varrho(4s-2) & \text{for } \frac{1}{2} \leqslant s \leqslant \frac{3}{4}, \\ \eta(4s-3) & \text{for } \frac{3}{4} \leqslant s \leqslant 1. \end{cases}$$

Now define the map $h : I \times I \to I$ by

$$h(s,t) = \begin{cases} s(1+t) & \text{for } 0 \leqslant s \leqslant \frac{1}{4}, \\ s + \frac{t}{4} & \text{for } \frac{1}{4} \leqslant s \leqslant \frac{1}{2}, \\ s + \frac{1}{2}(1-s)t & \text{for } \frac{1}{2} \leqslant s \leqslant 1, \end{cases}$$

and check that it is well-defined and continuous. Moreover, h is a homotopy from $h_0 = \mathrm{id}_I$ to the map

$$h_1 : I \to I, \quad s \mapsto \begin{cases} 2s & \text{for } 0 \leqslant s \leqslant \frac{1}{4}, \\ s + \frac{1}{4} & \text{for } \frac{1}{4} \leqslant s \leqslant \frac{1}{2}, \\ \frac{1}{2}(s+1) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

Hence $H := \big((\eta * \varrho) * \gamma\big) \circ h : I \times I \to X$ is a homotopy from $H_0 = (\eta * \varrho) * \gamma$ to $\big((\eta * \varrho) * \gamma\big) \circ h_1$. But the latter map coincides with $\eta * (\varrho * \gamma)$ since

$$\big((\eta * \varrho) * \gamma\big) \circ h_1(s) = \begin{cases} \gamma(4s) & \text{for } 0 \leqslant s \leqslant \frac{1}{4}, \\ \varrho(4s-1) & \text{for } \frac{1}{4} \leqslant s \leqslant \frac{1}{2}, \\ \eta(2s-1) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

Hence $\eta * (\varrho * \gamma)$ and $(\eta * \varrho) * \gamma$ are homotopic paths which proves associativity of the operation $*$ on $\pi_1(X, x_0)$.

Next consider the concatenations $\gamma * e_{x_0}$ and $e_{x_0} * \gamma$. Then

$$H : I \times I \to X, \quad (s,t) \mapsto \begin{cases} \gamma((1-t)s) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \gamma(s + t(s-1))) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

is a homotopy between $H_0 = \gamma$ and $H_1 = \gamma * e_{x_0}$, and

$$G : I \times I \to X, \quad (s,t) \mapsto \begin{cases} \gamma(s(1+t)) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \gamma(s + t - st) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

a homotopy between $G_0 = \gamma$ and $G_1 = e_{x_0} * \gamma$. Hence e_{x_0} is the identity element of $\pi_1(X, x_0)$.

Finally consider the concatenation

$$\gamma^{-1} * \gamma : I \to X, \quad s \mapsto \begin{cases} \gamma(2s) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \gamma(2(1-s)) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

Then

$$H : I \times I \to X, \quad (s,t) \mapsto \begin{cases} \gamma(2(1-t)s) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ \gamma(2(1-s)(1-t)) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases}$$

is a homotopy between $H_0 = \gamma^{-1} * \gamma$ and $H_1 = e_{x_0}$. Since by definition $\left(\gamma^{-1}\right)^{-1} = \gamma$, the paths $\gamma * \gamma^{-1}$ and e_{x_0} are homotopic as well. So γ^{-1} is the inverse of γ, and $\pi_1(X, x_0)$ is a group as claimed. $\square$

II.4. Differential Topology

4.1. Affine spaces and convex sets

4.1.1 Definition By an *affine space* over a field $\Bbbk$ one understands a set $\mathbb{A}$ together with a free and transitive action

$$\alpha : V \times \mathbb{A} \to \mathbb{A}, \quad (P, v) \mapsto v + P$$

of a $\Bbbk$-vector space V on $\mathbb{A}$. This means that α satisfies the following relations:

(Act1) The zero element acts as identity that is $0 + P = P$ for all $P \in \mathbb{A}$.

(Act2) The map α is compatible with addition in V which means that $v + (w + P) = (v + w) + P$ for all $P \in \mathbb{A}$ and $v, w \in V$.

(ActF) The action is free which means that for all $P \in \mathbb{A}$ and $v \in V$ the relation $v + P = P$ entails $v = 0$.

(ActT) The action is transitive which means that for all elements $P, Q \in \mathbb{A}$ there exists an element $v \in V$ such that $v + P = Q$.

One calls the elements of $\mathbb{A}$ the *points* of the affine space, and V its *vector space of translations*.

4.1.2 Since the action of V on $\mathbb{A}$ is free and transitive, one has a map

$$- : \mathbb{A} \times \mathbb{A} \to V, \quad (P, Q) \mapsto P - Q$$

which associates to each $(P, Q) \in \mathbb{A} \times \mathbb{A}$ the unique vector $v \in V$ such that $P = v + Q$. It is called *subtraction map*.

4.1.3 Proposition *The subtraction map of an affine space* $(\mathbb{A}, V, \alpha)$ *has the following properties, called* Weyl's axioms*:*

(Weyl1) *For every $P \in \mathbb{A}$ the map $V \to \mathbb{A}$, $v \mapsto v + P$ is a bijection.*

(Weyl2) *For all $P, Q, R \in \mathbb{A}$ one has*

$$(P - Q) + (Q - R) = P - R \, .$$

4.1.4 Definition If $(\mathbb{A}, V, \alpha)$ and $(\mathbb{B}, W, \beta)$ are two affine spaces, an *affine map* or *affine homomorphism* between them is a map $f : \mathbb{A} \to \mathbb{B}$ such that the function

$$\mathbb{A} \times \mathbb{A} \to W, \quad (P, Q) \mapsto f(P) - f(Q)$$

factors through a linear map which means that there exists a linear map $F : V \to W$ such that $F(P - Q) = f(P) - f(Q)$ for all $P, Q \in \mathbb{A}$.

4.1.5 Remark The linear map F associated to an affine map f is uniquely determined.

4.1.6 Example Let V be a $\Bbbk$-vector space. Put $\mathbb{A} := V$ and let the action $\alpha : V \times \mathbb{A} \to \mathbb{A}$ coincide with addition. Then $(\mathbb{A}, V, \alpha)$ is an affine space called the affine space *associated* with the vector space V.

4.2. Fiber bundles

Fibered spaces and manifolds

4.2.1 Definition By a *fibered space* we understand a triple (B, E, p) consisting of two topological spaces B, E called *base* and *total space*, respectively, and a continuous surjective map $p : E \top B$ called *projection*. We usually denote a fiber space just by its projection $p : E \to B$ or even only by its total space E if no confusion can arise. For each $b \in B$ the preimage $F_b = p^{-1}(\{b\}$ is called the *fiber* of $p : E \to B$ over b.

If E, B are both smooth manifolds and $p : E \top B$ is a smooth surjective submersion, then we call the triple (B, E, p) a *fibered manifold*.

4.2.2 Remark With our definition of a fibered space we closely follow **?** where a *fibre space* is defined as a triple (B, E, p) such that $p : E \to B$ is a continuous map between topological spaces E and B. Unlike here, surjectivity is not required. Because

4.2.3 Proposition *Let $p : E \to B$ be a fibered space. Then each fiber F_b, $b \in B$ is a nonempty topological space which is Hausdorff, regular, normal, paracompact, or metrizable if E is.*

In case $p : E \to B$ is a fibered manifold, then each fiber is a smooth manifold.

Bundles

4.2.4 We already defined the notion of a fiber bundle in Definition 4.2.5. Here we will extend that notion in two ways, namely first to the smooth case and second to the case where the fibers are vector spaces.

4.2.5 Definition Let F be a topological space. By *fiber bundle* with *typical fiber* F one understands a topological space E, together with another topological space B and a continuous map $\pi : E \to B$ such that for every point $x \in B$ there exists a *local trivialization* over an open neighborhood $U \subset B$ of x that is a homeomorphism $\varphi : \pi^{-1}(U) \to U \times F$ such that the diagram

$$\begin{tikzcd}
\pi^{-1}(U) \arrow[r, "\varphi"] \arrow[d, "p"'] & U \times F \arrow[dl, "\mathrm{pr}_1"] \\
U &
\end{tikzcd}$$

commutes, where $\mathrm{pr}_1 : U \times F \to U$ is projection onto the first factor. The neighborhood U is sometimes called a *trivializing neighborhood* of X. The space E is called the *total space* of fiber bundle, B its *base*, and $\pi : E \to B$ its *projection*.

If all spaces E, B, F are smooth manifolds, the projection $\pi : E \to B$ a smooth map, and around each point $x \in B$ there exists a local trivialization $\varphi : \pi^{-1}(U) \to U \times F$ which is a diffeomorphism, the fiber bundle is called a *smooth fiber bundle*.

4.3. Vector bundles

4.3.1

4.3.2 Definition

Part III.

Commutative Algebra

III.1. The spectrum of a commutative ring

1.1. Introduction

The notion of the Spec of a ring is fundamental in modern algebraic geometry. It is the scheme-theoretic analog of classical affine schemes. The identification occurs when one identifies the maximal ideals of the polynomial ring $k[x_1, \ldots, x_n]$ (for k an algebraically closed field) with the points of the classical variety $\mathbb{A}_k^n = k^n$. In modern algebraic geometry, one adds the "non-closed points" given by the other prime ideals. Just as general varieties were classically defined by gluing affine varieties, a scheme is defined by gluing open affines.

This is not a book on schemes, but it will nonetheless be convenient to introduce the Spec construction, outside of the obvious benefits of including preparatory material for algebraic geometry. First of all, it will provide a convenient notation. Second, and more importantly, it will provide a convenient geometric intuition. For example, an R-module can be thought of as a kind of "vector bundle"—technically, a sheaf—over the space $\operatorname{Spec} R$, with the caveat that the rank might not be locally constant (which is, however, the case when the module is projective).

1.2. The spectrum and the Zariski topology

We shall now associate to every commutative ring R a topological space $\operatorname{Spec} R$ in a functorial manner. That is, there will be a contravariant functor

$$\operatorname{Spec} : \mathsf{CRing} \to \mathsf{Top}$$

where Top is the category of topological spaces. This construction is the basis for scheme-theoretic algebraic geometry and will be used frequently in the sequel.

The motivating observation is the following. If $\Bbbk$ is an algebraically closed field, then the maximal ideals in $\Bbbk[x_1, \ldots, x_n]$ are of the form $(x_1 - a_1, \ldots, x_n - a_n)$ for $(a_1, \ldots, a_n) \in \Bbbk[x_1, \ldots, x_n]$. This is the Nullstellensatz, which we have not proved yet. We can thus identify the maximal ideals in the polynomial ring with the space $\Bbbk^n$. If $I \subset \Bbbk[x_1, \ldots, x_n]$ is an ideal, then the maximal ideals in $\Bbbk[x_1, \ldots, x_n]$ correspond to points where everything in I vanishes. See 1.2.6 for a more detailed explanation. Classical affine algebraic geometry thus studies the set of maximal ideals in an algebra finitely generated over an algebraically closed field.

The Spec of a ring is a generalization of this construction. In general, it is more natural to use all prime ideals instead of just maximal ideals.

Definition and examples

We start by defining Spec as a set. We will next construct the Zariski topology and later the functoriality.

1.2.1 Definition Let R be a commutative ring. The **spectrum** of R, denoted $\operatorname{Spec} R$, is the set of prime ideals of R.

We shall now make $\operatorname{Spec} R$ into a topological space. First, we describe a collection of sets which will become the closed sets. If $I \subset R$ is an ideal, let

$$V(I) = \{\mathfrak{p} : \mathfrak{p} \supset I\} \subset \operatorname{Spec} R.$$

1.2.2 Proposition *There is a topology on* $\operatorname{Spec} R$ *such that the closed subsets are of the form* $V(I)$ *for* $I \subset R$ *an ideal.*

Proof. Indeed, we have to check the familiar axioms for a topology:

1. $\varnothing = V((1))$ because no prime contains 1. So $\varnothing$ is closed.
2. $\operatorname{Spec} R = V((0))$ because any ideal contains zero. So $\operatorname{Spec} R$ is closed.
3. We show the closed sets are stable under intersections. Let $K_\alpha = V(I_\alpha)$ be closed subsets of $\operatorname{Spec} R$ for α ranging over some index set. Let $I = \sum I_\alpha$. Then

   $$V(I) = \bigcap K_\alpha = \bigcap V(I_\alpha),$$

   which follows because I is the smallest ideal containing each I_α, so a prime contains every I_α iff it contains I.
4. The union of two closed sets is closed. Indeed, if $K, K' \subset \operatorname{Spec} R$ are closed, we show $K \cup K'$ is closed. Say $K = V(I), K' = V(I')$. Then we claim:

   $$K \cup K' = V(II').$$

   Here, as usual, II' is the ideal generated by products $ii', i \in I, i' \in I'$. If $\mathfrak{p}$ is **prime** and contains II', it must contain one of I, I'; this implies the displayed equation above and implies the result. □

1.2.3 Definition The topology on $\operatorname{Spec} R$ defined above is called the **Zariski topology**. With it, $\operatorname{Spec} R$ is now a topological space.

1.2.4 Remark What is the Spec of the zero ring?

In order to see the geometry of this construction, let us work several examples.

1.2.5 Example Let $R = \mathbb{Z}$, and consider $\operatorname{Spec} \mathbb{Z}$. Then every prime is generated by one element, since $\mathbb{Z}$ is a PID. We have that $\operatorname{Spec} \mathbb{Z} = \{(0)\} \cup \bigcup_{p \text{ prime}} \{(p)\}$. The picture is that one has all the familiar primes $(2), (3), (5), \dots$, and then a special point (0).

Let us now describe the closed subsets. These are of the form $V(I)$ where $I \subset \mathbb{Z}$ is an ideal, so $I = (n)$ for some $n \in \mathbb{Z}$.

1. If $n = 0$, the closed subset is all of $\operatorname{Spec}\mathbb{Z}$.
2. If $n \neq 0$, then n has finitely many prime divisors. So $V((n))$ consists of the prime ideals corresponding to these prime divisors.

The only closed subsets besides the entire space are the finite subsets that exclude (0).

1.2.6 Example Say $R = \mathbb{C}[x, y]$ is a polynomial ring in two variables. We will not give a complete description of $\operatorname{Spec} R$ here. But we will write down several prime ideals.

1. For every pair of complex numbers $s, t \in \mathbb{C}$, the collection of polynomials $f \in R$ such that $f(s, t) = 0$ is a prime ideal $\mathfrak{m}_{s,t} \subset R$. In fact, it is maximal, as the residue ring is all of $\mathbb{C}$. Indeed, $R/\mathfrak{m}_{s,t} \simeq \mathbb{C}$ under the map $f \to f(s, t)$.

   In fact,

   1.2.7 Theorem *The $\mathfrak{m}_{s,t}$ are all the maximal ideals in R.*

   This will follow from the *Hilbert Nullstellensatz* to be proved later (4.4.5).
2. $(0) \subset R$ is a prime ideal since R is a domain.
3. If $f(x, y) \in R$ is an irreducible polynomial, then (f) is a prime ideal. This is equivalent to unique factorization in R.[1]

To draw $\operatorname{Spec} R$, we start by drawing $\mathbb{C}^2$, which is identified with the collection of maximal ideals $\mathfrak{m}_{s,t}$, $s, t \in \mathbb{C}$. $\operatorname{Spec} R$ has additional (non-closed) points too, as described above, but for now let us consider the topology induced on $\mathbb{C}^2$ as a subspace of $\operatorname{Spec} R$.

The closed subsets of $\operatorname{Spec} R$ are subsets $V(I)$ where I is an ideal, generated by polynomials $\{f_\alpha(x, y)\}$. It is of interest to determine the subset of $\mathbb{C}^2$ that $V(I)$ induces. In other words, we ask:

> What points of $\mathbb{C}^2$ (with (s, t) identified with $\mathfrak{m}_{s,t}$) lie in $V(I)$?

Now, by definition, we know that (s, t) corresponds to a point of $V(I)$ if and only if $I \subset \mathfrak{m}_{s,t}$. This is true iff all the f_α lie in $\mathfrak{m}_{s,t}$, i.e. if $f_\alpha(s, t) = 0$ for all α. So the closed subsets of $\mathbb{C}^2$ (with the induced Zariski topology) are *precisely the subsets that can be defined by polynomial equations.*

This is **much** coarser than the usual topology. For instance, $\{(z_1, z_2) : \mathfrak{Re}(z_1) \geqslant 0\}$ is not Zariski-closed. The Zariski topology is so coarse because one has only algebraic data (namely, polynomials, or elements of R) to define the topology.

1.2.8 Remark Let R_1, R_2 be commutative rings. Give $R_1 \times R_2$ a natural structure of a ring, and describe $\operatorname{Spec}(R_1 \times R_2)$ in terms of $\operatorname{Spec} R_1$ and $\operatorname{Spec} R_2$.

1.2.9 Remark Let X be a compact Hausdorff space, $C(X)$ the ring of real continuous functions $X \to \mathbb{R}$. The maximal ideals in $\operatorname{Spec} C(X)$ are in bijection with the points of X, and the topology induced on X (as a subset of $\operatorname{Spec} C(X)$ with the Zariski topology) is just the usual topology.

[1] To be proved later **??**.

1.2.10 Remark Prove the following result: if X, Y are compact Hausdorff spaces and $C(X), C(Y)$ the associated rings of continuous functions, if $C(X), C(Y)$ are isomorphic as $\mathbb{R}$-algebras, then X is homeomorphic to Y.

The radical ideal-closed subset correspondence

We now return to the case of an arbitrary commutative ring R. If $I \subset R$, we get a closed subset $V(I) \subset \operatorname{Spec} R$. It is called $V(I)$ because one is supposed to think of it as the places where the elements of I "vanish," as the elements of R are something like "functions." This analogy is perhaps best seen in the example of a polynomial ring over an algebraically closed field, e.g. 1.2.6 above.

The map from ideals into closed sets is very far from being injective in general, though by definition it is surjective.

1.2.11 Example If $R = \mathbb{Z}$ and p is prime, then $I = (p), I' = (p^2)$ define the same subset (namely, $\{(p)\}$) of $\operatorname{Spec} R$.

We now ask why the map from ideals to closed subsets fails to be injective. As we shall see, the entire problem disappears if we restrict to *radical* ideals.

1.2.12 Definition If I is an ideal, then the **radical** $\operatorname{Rad}(I)$ or $\sqrt{I}$ is defined as

$$\operatorname{Rad}(I) = \{x \in R : x^n \in I \text{ for some } n\}.$$

An ideal is **radical** if it is equal to its radical. (This is equivalent to the earlier 2.2.5.)

Before proceeding, we must check:

1.2.13 Lemma *If I an ideal, so is* $\operatorname{Rad}(I)$.

Proof. Clearly $\operatorname{Rad}(I)$ is closed under multiplication since I is. Suppose $x, y \in \operatorname{Rad}(I)$; we show $x + y \in \operatorname{Rad}(I)$. Then $x^n, y^n \in I$ for some n (large) and thus for all larger n. The binomial expansion now gives

$$(x+y)^{2n} = x^{2n} + \binom{2n}{1} x^{2n-1} y + \cdots + y^{2n},$$

where every term contains either x, y with power $\geqslant n$, so every term belongs to I. Thus $(x+y)^{2n} \in I$ and, by definition, we see then that $x + y \in \operatorname{Rad}(I)$. $\square$

The map $I \to V(I)$ does in fact depend only on the radical of I. In fact, if I, J have the same radical $\operatorname{Rad}(I) = \operatorname{Rad}(J)$, then $V(I) = V(J)$. Indeed, $V(I) = V(\operatorname{Rad}(I)) = V(\operatorname{Rad}(J)) = V(J)$ by:

1.2.14 Lemma *For any I, $V(I) = V(\operatorname{Rad}(I))$.*

Proof. Indeed, $I \subset \operatorname{Rad}(I)$ and therefore obviously $V(\operatorname{Rad}(I)) \subset V(I)$. We have to show the converse inclusion. Namely, we must prove:

If $\mathfrak{p} \supset I$, then $\mathfrak{p} \supset \mathrm{Rad}(I)$.

So suppose $\mathfrak{p} \supset I$ is prime and $x \in \mathrm{Rad}(I)$; then $x^n \in I \subset \mathfrak{p}$ for some n. But $\mathfrak{p}$ is prime, so whenever a product of things belongs to $\mathfrak{p}$, a factor does. Thus since $x^n = x \cdot x \cdots x$, we must have $x \in \mathfrak{p}$. So

$$\mathrm{Rad}(I) \subset \mathfrak{p},$$

proving the quoted claim, and thus the lemma. ∎

There is a converse to this remark:

1.2.15 Proposition *If $V(I) = V(J)$, then $\mathrm{Rad}(I) = \mathrm{Rad}(J)$.*

So two ideals define the same closed subset iff they have the same radical.

Proof. We write down a formula for $\mathrm{Rad}(I)$ that will imply this at once.

1.2.16 Lemma *For a commutative ring R and an ideal $I \subset R$,*

$$\mathrm{Rad}(I) = \bigcap_{\mathfrak{p} \supset I} \mathfrak{p}.$$

From this, it follows that $V(I)$ determines $\mathrm{Rad}(I)$. This will thus imply the proposition. We now prove the lemma:

Proof.
1. We show $\mathrm{Rad}(I) \subset \bigcap_{\mathfrak{p} \in V(I)} \mathfrak{p}$. In particular, this follows if we show that if a prime contains I, it contains $\mathrm{Rad}(I)$; but we have already discussed this above.
2. If $x \notin \mathrm{Rad}(I)$, we will show that there is a prime ideal $\mathfrak{p} \supset I$ not containing x. This will imply the reverse inclusion and the lemma.

We want to find $\mathfrak{p}$ not containing x, more generally not containing any power of x. In particular, we want $\mathfrak{p} \cap \left\{1, x, x^2 \ldots, \right\} = \varnothing$. This set $S = \{1, x, \ldots\}$ is multiplicatively closed, in that it contains 1 and is closed under finite products. Right now, it does not interset I; we want to find a *prime* containing I that still does not intersect $\{x^n, n \geqslant 0\}$.

More generally, we will prove:

1.2.17 Lemma *Let S be multiplicatively closed set in any ring R and let I be any ideal with $I \cap S = \varnothing$. There is a prime ideal $\mathfrak{p} \supset I$ and does not intersect S (in fact, any ideal maximal with respect to the condition of not intersecting S will do).*

In English, any ideal missing S can be enlarged to a prime ideal missing S. This is actually fancier version of a previous argument. We showed earlier that any ideal not containing the multiplicatively closed subset $\{1\}$ can be contained in a prime ideal not containing 1, in 2.6.8.

Note that the lemma clearly implies the lemma when applied to $S = \{1, x, \ldots\}$.

Proof of the lemma. Let $P = \{J : J \supset I, J \cap S = \varnothing\}$. Then P is a poset with respect to inclusion. Note that $P \neq \varnothing$ because $I \in P$. Also, for any nonempty linearly ordered subset of P, the union is in P (i.e. there is an upper bound). We can invoke Zorn's lemma to get a maximal element of P. This element is an ideal $\mathfrak{p} \supset I$ with $\mathfrak{p} \cap S = \varnothing$. We claim that $\mathfrak{p}$ is prime.

First of all, $1 \notin \mathfrak{p}$ because $1 \in S$. We need only check that if $xy \in \mathfrak{p}$, then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. Suppose otherwise, so $x, y \notin \mathfrak{p}$. Then $(x, \mathfrak{p}) \notin P$ or $\mathfrak{p}$ would not be maximal. Ditto for $(y, \mathfrak{p})$.

In particular, we have that these bigger ideals both intersect S. This means that there are

$$a \in \mathfrak{p}, r \in R \quad \text{such that} \quad a + rx \in S$$

and

$$b \in \mathfrak{p}, r' \in R \quad \text{such that} \quad b + r'y \in S.$$

Now S is multiplicatively closed, so multiply $(a + rx)(b + r'y) \in S$. We find:

$$ab + ar'y + brx + rr'xy \in S. \qquad \square$$

Now $a, b \in \mathfrak{p}$ and $xy \in \mathfrak{p}$, so all the terms above are in $\mathfrak{p}$, and the sum is too. But this contradicts $\mathfrak{p} \cap S = \varnothing$. $\square$

The upshot of the previous lemmata is:

1.2.18 Proposition *There is a bijection between the closed subsets of* $\operatorname{Spec} R$ *and radical ideals* $I \subset R$.

A meta-observation about prime ideals

We saw in the previous subsec (lemma 1.2.17) that an ideal maximal with respect to the property of not intersecting a multiplicatively closed subset is prime. It turns out that this is the case for many such properties of ideals. A general method of seeing this was developed in **?**. In this (optional) subsec, we digress to explain this phenomenon.

If I is an ideal and $a \in R$, we define the notation

$$(I : a) = \{x \in R : xa \in I\}.$$

More generally, if J is an ideal, we define

$$(I : J) = \{x \in R : xJ \subset I\}.$$

Let R be a ring, and $\mathcal{F}$ a collection of ideals of R. We are interested in conditions that will guarantee that the maximal elements of $\mathcal{F}$ are *prime*. Actually, we will do the opposite: the following condition will guarantee that the ideals maximal at *not* being in $\mathcal{F}$ are prime.

1.2.19 Definition The family $\mathcal{F}$ is called an **Oka family** if $R \in \mathcal{F}$ (where R is considered as an ideal) and whenever $I \subset R$ is an ideal and $(I : a), (I, a) \in \mathcal{F}$ (for some $a \in R$), then $I \in \mathcal{F}$.

1.2.20 Example Let us begin with a simple observation. If $(I : a)$ is generated by $a_1, \ldots, a_n$ and (I, a) is generated by $a, b_1, \ldots, b_m$ (where we may take $b_1, \ldots, b_m \in I$, without loss of generality), then I is generated by $aa_1, \ldots, aa_n, b_1, \ldots, b_m$. To see this, note that if $x \in I$, then $x \in (I, a)$ is a linear combination of the $\{a, b_1, \ldots, b_m\}$, but the coefficient of a must lie in $(I : a)$.

As a result, we may deduce that the family of finitely generated ideals is an Oka family.

1.2.21 Example Let us now show that the family of *principal* ideals is an Oka family. Indeed, suppose $I \subset R$ is an ideal, and (I, a) and $(I : a)$ are principal. One can easily check that $(I : a) = (I : (I, a))$. Setting $J = (I, a)$, we find that J is principal and $(I : J)$ is too. However, for *any* principal ideal J, and for any ideal $I \subset J$,

$$I = J(I : J)$$

as one easily checks. Thus we find in our situation that since $J = (I, a)$ and $(I : J)$ are principal, I is principal.

1.2.22 Proposition (?) *If $\mathcal{F}$ is an Oka family of ideals, then any maximal element of the complement of $\mathcal{F}$ is prime.*

Proof. Suppose $I \notin \mathcal{F}$ is maximal with respect to not being in $\mathcal{F}$ but I is not prime. Note that $I \neq R$ by hypothesis. Then there is $a \in R$ such that $(I : a), (I, a)$ both strictly contain I, so they must belong to $\mathcal{F}$. Indeed, we can find $a, b \in R - I$ with $ab \in I$; it follows that $(I, a) \neq I$ and $(I : a)$ contains $b \notin I$.

By the Oka condition, we have $I \in \mathcal{F}$, a contradiction. □

1.2.23 Corollary (Cohen) *If every prime ideal of R is finitely generated, then every ideal of R is finitely generated.*[2]

Proof. Suppose that there existed ideals $I \subset R$ which were not finitely generated. The union of a totally ordered chain $\{I_\alpha\}$ of ideals that are not finitely generated is not finitely generated; indeed, if $I = \bigcup I_\alpha$ were generated by $a_1, \ldots, a_n$, then all the generators would belong to some I_α and would consequently generate it.

By Zorn's lemma, there is an ideal maximal with respect to being not finitely generated. However, by 1.2.22, this ideal is necessarily prime (since the family of finitely generated ideals is an Oka family). This contradicts the hypothesis. □

1.2.24 Corollary *If every prime ideal of R is principal, then every ideal of R is principal.*

Proof. This is proved in the same way. □

1.2.25 Remark Suppose every nonzero prime ideal in R contains a non-zero-divisor. Then R is a domain. (Hint: consider the set S of non-zero-divisors, and argue that any ideal maximal with respect to not intersecting S is prime. Thus, (0) is prime.)

[2]Later we will say that R is *noetherian.*

1.2.26 Remark Let R be a ring. Let κ be an infinite cardinal. By applying 1.2.20 and 1.2.22 we see that any ideal maximal with respect to the property of not being generated by κ elements is prime. This result is not so useful because there exists a ring for which every prime ideal of R can be generated by $\aleph_0$ elements, but some ideal cannot. Namely, let k be a field, let T be a set whose cardinality is greater than $\aleph_0$ and let

$$R = k[\{x_n\}_{n \geqslant 1}, \{z_{t,n}\}_{t \in T, n \geqslant 0}]/(x_n^2, z_{t,n}^2, x_n z_{t,n} - z_{t,n-1})$$

This is a local ring with unique prime ideal $\mathfrak{m} = (x_n)$. But the ideal $(z_{t,n})$ cannot be generated by countably many elements.

Functoriality of Spec

The construction $R \to \operatorname{Spec} R$ is functorial in R in a contravariant sense. That is, if $f : R \to R'$, there is a continuous map $\operatorname{Spec} R' \to \operatorname{Spec} R$. This map sends $\mathfrak{p} \subset R'$ to $f^{-1}(\mathfrak{p}) \subset R$, which is easily seen to be a prime ideal in R. Call this map $F : \operatorname{Spec} R' \to \operatorname{Spec} R$. So far, we have seen that $\operatorname{Spec} R$ induces a contravariant functor from **Rings** $\to$ **Sets**.

1.2.27 Remark A contravariant functor $F : \mathcal{C} \to$ **Sets** (for some category $\mathcal{C}$) is called **representable** if it is naturally isomorphic to a functor of the form $X \to \hom(X, X_0)$ for some $X_0 \in \mathcal{C}$, or equivalently if the induced covariant functor on $\mathcal{C}^{\mathrm{op}}$ is corepresentable.

The functor $R \to \operatorname{Spec} R$ is not representable. (Hint: Indeed, a representable functor must send the initial object into a one-point set.)

Next, we check that the morphisms induced on Spec's from a ring-homomorphism are in fact *continuous* maps of topological spaces.

1.2.28 Proposition Spec *induces a contravariant functor from* **Rings** *to the category* **Top** *of topological spaces.*

Proof. Let $f : R \to R'$. We need to check that this map $\operatorname{Spec} R' \to \operatorname{Spec} R$, which we call F, is continuous. That is, we must check that F^{-1} sends closed subsets of $\operatorname{Spec} R$ to closed subsets of $\operatorname{Spec} R'$.

More precisely, if $I \subset R$ and we take the inverse image $F^{-1}(V(I)) \subset \operatorname{Spec} R'$, it is just the closed set $V(f(I))$. This is best left to the reader, but here is the justification. If $\mathfrak{p} \in \operatorname{Spec} R'$, then $F(\mathfrak{p}) = f^{-1}(\mathfrak{p}) \supset I$ if and only if $\mathfrak{p} \supset f(I)$. So $F(\mathfrak{p}) \in V(I)$ if and only if $\mathfrak{p} \in V(f(I))$.

1.2.29 Example Let R be a commutative ring, $I \subset R$ an ideal, $f : R \to R/I$. There is a map of topological spaces

$$F : \operatorname{Spec}(R/I) \to \operatorname{Spec} R.$$

This map is a closed embedding whose image is $V(I)$. Most of this follows because there is a bijection between ideals of R containing I and ideals of R/I, and this bijection preserves primality.

1.2.30 Remark Show that this map $\operatorname{Spec} R/I \to \operatorname{Spec} R$ is indeed a homeomorphism from $\operatorname{Spec} R/I \to V(I)$.

A basis for the Zariski topology

In the previous section, we were talking about the Zariski topology. If R is a commutative ring, we recall that $\operatorname{Spec} R$ is defined to be the collection of prime ideals in R. This has a topology where the closed sets are the sets of the form

$$V(I) = \{\mathfrak{p} \in \operatorname{Spec} R : \mathfrak{p} \supset I\}.$$

There is another way to describe the Zariski topology in terms of *open* sets.

1.2.31 Definition If $f \in R$, we let

$$U_f = \{\mathfrak{p} : f \notin \mathfrak{p}\}$$

so that U_f is the subset of $\operatorname{Spec} R$ consisting of primes not containing f. This is the complement of $V((f))$, so it is open.

1.2.32 Proposition *The sets U_f form a basis for the Zariski topology.*

Proof. Suppose $U \subset \operatorname{Spec} R$ is open. We claim that U is a union of basic open sets U_f.

Now $U = \operatorname{Spec} R - V(I)$ for some ideal I. Then

$$U = \bigcup_{f \in I} U_f$$

because if an ideal is not in $V(I)$, then it fails to contain some $f \in I$, i.e. is in U_f for that f. Alternatively, we could take complements, whence the above statement becomes

$$V(I) = \bigcap_{f \in I} V((f))$$

which is clear. ∎

The basic open sets have nice properties.

1. $U_1 = \operatorname{Spec} R$ because prime ideals are not allowed to contain the unit element.
2. $U_0 = \varnothing$ because every prime ideal contains 0.
3. $U_{fg} = U_f \cap U_g$ because fg lies in a prime ideal $\mathfrak{p}$ if and only if one of f, g does.

Now let us describe what the Zariski topology has to do with localization. Let R be a ring and $f \in R$. Consider $S = \{1, f, f^2, \dots\}$; this is a multiplicatively closed subset. Last week, we defined $S^{-1}R$.

1.2.33 Definition For S the powers of f, we write R_f or $R[f^{-1}]$ for the localization $S^{-1}R$.

There is a map $\phi : R \to R[f^{-1}]$ and a corresponding map

$$\operatorname{Spec} R[f^{-1}] \to \operatorname{Spec} R$$

sending a prime $\mathfrak{p} \subset R[f^{-1}]$ to $\phi^{-1}(\mathfrak{p})$.

1.2.34 Proposition *This map induces a homeomorphism of* $\operatorname{Spec} R[f^{-1}]$ *onto* $U_f \subset \operatorname{Spec} R$.

So if one takes a commutative ring and inverts an element, one just gets an open subset of Spec. This is why it's called localization: one is restricting to an open subset on the Spec level when one inverts something.

Proof. The reader is encouraged to work this proof out for herself.

1. First, we show that $\operatorname{Spec} R[f^{-1}] \to \operatorname{Spec} R$ lands in U_f. If $\mathfrak{p} \subset R[f^{-1}]$, then we must show that the inverse image $\phi^{-1}(\mathfrak{p})$ can't contain f. If otherwise, that would imply that $\phi(f) \in \mathfrak{p}$; however, $\phi(f)$ is invertible, and then $\mathfrak{p}$ would be (1).

2. Let's show that the map surjects onto U_f. If $\mathfrak{p} \subset R$ is a prime ideal not containing f, i.e. $\mathfrak{p} \in U_f$. We want to construct a corresponding prime in the ring $R[f^{-1}]$ whose inverse image is $\mathfrak{p}$.

   Let $\mathfrak{p}[f^{-1}]$ be the collection of all fractions
   $$\{\frac{x}{f^n}, x \in \mathfrak{p}\} \subset R[f^{-1}],$$
   which is evidently an ideal. Note that whether the numerator is in $\mathfrak{p}$ is **independent** of the representing fraction $\frac{x}{f^n}$ used.[3] In fact, $\mathfrak{p}[f^{-1}]$ is a prime ideal. Indeed, suppose
   $$\frac{a}{f^m}\frac{b}{f^n} \in \mathfrak{p}[f^{-1}].$$
   Then $\frac{ab}{f^{m+n}}$ belongs to this ideal, which means $ab \in \mathfrak{p}$; so one of $a, b \in \mathfrak{p}$ and one of the two fractions $\frac{a}{f^m}, \frac{b}{f^n}$ belongs to $\mathfrak{p}[f^{-1}]$. Also, $1/1 \notin \mathfrak{p}[f^{-1}]$.

   It is clear that the inverse image of $\mathfrak{p}[f^{-1}]$ is $\mathfrak{p}$, because the image of $x \in R$ is $x/1$, and this belongs to $\mathfrak{p}[f^{-1}]$ precisely when $x \in \mathfrak{p}$.

3. The map $\operatorname{Spec} R[f^{-1}] \to \operatorname{Spec} R$ is injective. Suppose $\mathfrak{p}, \mathfrak{p}'$ are prime ideals in the localization and the inverse images are the same. We must show that $\mathfrak{p} = \mathfrak{p}'$.

   Suppose $\frac{x}{f^n} \in \mathfrak{p}$. Then $x/1 \in \mathfrak{p}$, so $x \in \phi^{-1}(\mathfrak{p}) = \phi^{-1}(\mathfrak{p}')$. This means that $x/1 \in \mathfrak{p}'$, so $\frac{x}{f^n} \in \mathfrak{p}'$ too. So a fraction that belongs to $\mathfrak{p}$ belongs to $\mathfrak{p}'$. By symmetry the two ideals must be the same.

4. We now know that the map $\psi : \operatorname{Spec} R[f^{-1}] \to U_f$ is a continuous bijection. It is left to see that it is a homeomorphism. We will show that it is open. In particular, we have to show that a basic open set on the left side is mapped to an open set on the right side. If $y/f^n \in R[f^{-1}]$, we have to show that $U_{y/f^n} \subset \operatorname{Spec} R[f^{-1}]$ has open image under ψ. We'll in fact show what open set it is.

   We claim that
   $$\psi(U_{y/f^n}) = U_{fy} \subset \operatorname{Spec} R.$$

[3] Suppose $\frac{x}{f^n} = \frac{y}{f^k}$ for $y \in \mathfrak{p}$. Then there is N such that $f^N(f^k x - f^n y) = 0 \in \mathfrak{p}$; since $y \in \mathfrak{p}$ and $f \notin \mathfrak{p}$, it follows that $x \in \mathfrak{p}$.

To see this, $\mathfrak{p}$ is contained in U_{f/y^n}. This mean that $\mathfrak{p}$ doesn't contain y/f^n. In particular, $\mathfrak{p}$ doesn't contain the multiple $yf/1$. So $\psi(\mathfrak{p})$ doesn't contain yf. This proves the inclusion $\subset$.

5. To complete the proof of the claim, and the result, we must show that if $\mathfrak{p} \subset \operatorname{Spec} R[f^{-1}]$ and $\psi(\mathfrak{p}) = \phi^{-1}(\mathfrak{p}) \in U_{fy}$, then y/f^n doesn't belong to $\mathfrak{p}$. (This is kosher and dandy because we have a bijection.) But the hypothesis implies that $fy \notin \phi^{-1}(\mathfrak{p})$, so $fy/1 \notin \mathfrak{p}$. Dividing by f^{n+1} implies that
$$y/f^n \notin \mathfrak{p}$$
and $\mathfrak{p} \in U_{f/y^n}$. □

If $\operatorname{Spec} R$ is a space, and f is thought of as a "function" defined on $\operatorname{Spec} R$, the space U_f is to be thought of as the set of points where f "doesn't vanish" or "is invertible." Thinking about rings in terms of their spectra is a very useful idea. We will bring it up when appropriate.

1.2.35 Remark The construction $R \to R[f^{-1}]$ as discussed above is an instance of localization. More generally, we defined $S^{-1}R$ for $S \subset R$ multiplicatively closed. We can thus define maps $\operatorname{Spec} S^{-1}R \to \operatorname{Spec} R$. To understand $S^{-1}R$, it may help to note that
$$\varinjlim_{f \in S} R[f^{-1}]$$
which is a direct limit of rings where one inverts more and more elements.

As an example, consider $S = R - \mathfrak{p}$ for a prime $\mathfrak{p}$, and for simplicity that R is countable. We can write $S = S_0 \cup S_1 \cup \dots$, where each S_k is generated by a finite number of elements $f_0, \dots, f_k$. Then $R_{\mathfrak{p}} = \varinjlim S_k^{-1}R$. So we have
$$S^{-1}R = \varinjlim_k R[f_0^{-1}, f_1^{-1}, \dots, f_k^{-1}] = \varinjlim R[(f_0 \dots f_k)^{-1}].$$

The functions we invert in this construction are precisely those which do not contain $\mathfrak{p}$, or where "the functions don't vanish."

The geometric idea is that to construct $\operatorname{Spec} S^{-1}R = \operatorname{Spec} R_{\mathfrak{p}}$, we keep cutting out from $\operatorname{Spec} R$ vanishing locuses of various functions that do not intersect $\mathfrak{p}$. In the end, you don't restrict to an open set, but to an intersection of them.

1.2.36 Remark Say that R is *semi-local* if it has finitely many maximal ideals. Let $\mathfrak{p}_1, \dots, \mathfrak{p}_n \subset R$ be primes. The complement of the union, $S = R \smallsetminus \bigcup \mathfrak{p}_i$, is closed under multiplication, so we can localize. $R[S^{-1}] = R_S$ is called the *semi-localization* of R at the $\mathfrak{p}_i$.

The result of semi-localization is always semi-local. To see this, recall that the ideals in R_S are in bijection with ideals in R contained in $\bigcup \mathfrak{p}_i$. Now use prime avoidance.

1.2.37 Definition For a finitely generated R-module M, define $\mu_R(M)$ to be the smallest number of elements that can generate M.

This is not the same as the cardinality of a minimal set of generators. For example, 2 and 3 are a minimal set of generators for $\mathbb{Z}$ over itself, but $\mu_{\mathbb{Z}}(\mathbb{Z}) = 1$.

1.2.38 Theorem *Let R be semi-local with maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$. Let $k_i = R/\mathfrak{m}_i$. Then*

$$\mathfrak{m}u_R(M) = \mathfrak{m}ax\{\dim_{k_i} M/\mathfrak{m}_i M\}$$

Proof. **add: proof** □

1.3. Nilpotent elements

We will now prove a few general results about nilpotent results in a ring. Topologically, the nilpotents do very little: quotienting by them will not change the Spec. Nonetheless, they carry geometric importance, and one thinks of these nilpotents as "infinitesimal thickenings" (in a sense to be elucidated below).

The radical of a ring

There is a useful corollary of the analysis in the previous section about the Spec of a ring.

1.3.1 Definition $x \in R$ is called **nilpotent** if a power of x is zero. The set of nilpotent elements in R is called the **radical** of R and is denoted $\mathrm{Rad}(R)$ (which is an abuse of notation).

The set of nilpotents is just the radical $\mathrm{Rad}((0))$ of the zero ideal, so it is an ideal. It can vary greatly. A domain clearly has no nonzero nilpotents. On the other hand, many rings do:

1.3.2 Example For any $n \geqslant 2$, the ring $\mathbb{Z}[X]/(X^n)$ has a nilpotent, namely X. The ideal of nilpotent elements is (X).

It is easy to see that a nilpotent must lie in any prime ideal. The converse is also true by the previous analysis. As a corollary of it, we find in fact:

1.3.3 Corollary *Let R be a commutative ring. Then the set of nilpotent elements of R is precisely $\bigcap_{\mathfrak{p} \in \operatorname{Spec} R} \mathfrak{p}$.*

Proof. Apply 1.2.16 to the zero ideal. □

We now consider a few examples of nilpotent elements.

1.3.4 Example (Nilpotents in polynomial rings) Let us now compute the nilpotent elements in the polynomial $R[x]$. The claim is that a polynomial $\sum_{m=0}^n a_m x^m \in R[x]$ is nilpotent if and only if all the coefficients $a_m \in R$ are nilpotent. That is, $\mathrm{Rad}(R[x]) = (\mathrm{Rad}(R))R[x]$.

If $a_0, \ldots, a_n$ are nilpotent, then because the nilpotent elements form an ideal, $f = a_0 + \cdots + a_n x^n$ is nilpotent. Conversely, if f is nilpotent, then $f^m = 0$ and thus $(a_n x^n)^m = 0$. Thus $a_n x^n$ is nilpotent, and because the nilpotent elements form an ideal, $f - a_n x^n$ is nilpotent. By induction, $a_i x^i$ is nilpotent for all i, so that all a_i are nilpotent.

Before the next example, we need to define a new notion. We now define a power series ring intuitively in the same way they are used in calculus. In fact, we will use power series rings much the same way we used them in calculus; they will serve as keeping track of fine local data that the Zariski topology might "miss" due to its coarseness.

1.3.5 Definition Let R be a ring. The **power series ring** $R[[x]]$ is just the set of all expressions of the form $\sum_{i=0}^{\infty} c_i x^i$. The arithmetic for the power series ring will be done term by term formally (since we have no topology, we can't consider questions of convergence, though a natural topology can be defined making $R[[x]]$ the *completion* of another ring, as we shall see later).

1.3.6 Example (Nilpotence in power series rings) Let R be a ring such that $\mathrm{Rad}(R)$ is a finitely generated ideal. (This is satisfied, e.g., if R is *noetherian*, cf. III.2.) Let us consider the question of how $\mathrm{Rad}(R)$ and $\mathrm{Rad}(R[[x]])$ are related. The claim is that

$$\mathrm{Rad}(R[[x]]) = (\mathrm{Rad}(R))R[[x]].$$

If $f \in R[[x]]$ is nilpotent, say with $f^n = 0$, then certainly $a_0^n = 0$, so that a_0 is nilpotent. Because the nilpotent elements form an ideal, we have that $f - a_0$ is also nilpotent, and hence by induction every coefficient of f must be nilpotent in R. For the converse, let $I = \mathrm{Rad}(R)$. There exists an $N > 0$ such that the ideal power $I^N = 0$ by finite generation. Thus if $f \in IR[[x]]$, then $f^N \in I^N R[[x]] = 0$.

1.3.7 Remark Prove that $x \in R$ is nilpotent if and only if the localization R_x is the zero ring.

1.3.8 Remark Construct an example where $\mathrm{Rad}(R)R[[x]] \neq \mathrm{Rad}(R[[x]])$. (Hint: consider $R = \mathbb{C}[X_1, X_2, X_3, \dots]/(X_1, X_2^2, X_3^3, \dots)$.)

Lifting idempotents

If R is a ring, and $I \subset R$ a nilpotent ideal, then we want to think of R/I as somehow close to R. For instance, the inclusion $\operatorname{Spec} R/I \hookrightarrow \operatorname{Spec} R$ is a homeomorphism, and one pictures that $\operatorname{Spec} R$ has some "fuzz" added (with the extra nilpotents in I) that is killed in $\operatorname{Spec} R/I$.

One manifestation of the "closeness" of R and R/I is the following result, which states that the idempotent elements[4] of the two are in natural bijection. For convenience, we state it in additional generality (that is, for noncommutative rings).

1.3.9 Lemma (Lifting idempotents) *Suppose $I \subset R$ is a nilpotent two-sided ideal, for R any[5] ring. Let $\overline{e} \in R/I$ be an idempotent. Then there is an idempotent $e \in R$ which reduces to $\overline{e}$.*

Note that if J is a two-sided ideal in a noncommutative ring, then so are the powers of J.

[4] Recall that an element $e \in R$ is idempotent if $e^2 = e$.
[5] Not necessarily commutative.

Proof. Let us first assume that $I^2 = 0$. We can find $e_1 \in R$ which reduces to e, but e_1 is not necessarily idempotent. By replacing R with $\mathbb{Z}[e_1]$ and I with $\mathbb{Z}[e_1] \cap I$, we may assume that R is in fact commutative. However,

$$e_1^2 \in e_1 + I.$$

Suppose we want to modify e_1 by i such that $e = e_1 + i$ is idempotent and $i \in I$; then e will do as in the lemma. We would then necessarily have

$$e_1 + i = (e_1 + i)^2 = e_1^2 + 2e_1 i \quad \text{as } I^2 = 0.$$

In particular, we must satisfy

$$i(1 - 2e_1) = e_1^2 - e_1 \in I.$$

We claim that $1-2e_1 \in R$ is invertible, so that we can solve for $i \in I$. However, R is commutative. It thus suffices to check that $1 - 2e_1$ lies in no maximal ideal of R. But the image of e_1 in $R/\mathfrak{m}$ for any maximal ideal $\mathfrak{m} \subset R$ is either zero or one. So $1 - 2e_1$ has image either 1 or -1 in $R/\mathfrak{m}$. Thus it is invertible.

This establishes the result when I has zero square. In general, suppose $I^n = 0$. We have the sequence of noncommutative rings:

$$R \twoheadrightarrow R/I^{n-1} \twoheadrightarrow R/I^{n-2} \cdots \twoheadrightarrow R/I.$$

The kernel at each step is an ideal whose square is zero. Thus, we can use the lifting idempotents partial result proved above each step of the way and left $\overline{e} \in R/I$ to some $e \in R$. ∎

While the above proof has the virtue of applying to noncommutative rings, there is a more conceptual argument for commutative rings. The idea is that idempotents in A measure disconnections of $\operatorname{Spec} A$.[6] Since the topological space underlying $\operatorname{Spec} A$ is unchanged when one quotients by nilpotents, idempotents are unaffected. We prove:

1.3.10 Proposition *If $X = \operatorname{Spec} A$, then there is a one-to-one correspondence between* $\operatorname{Idem}(A)$ *and the open and closed subsets of X.*

Proof. Suppose I is the radical of (e) for an an idempotent $e \in R$. We show that $V(I)$ is open and closed. Since V is unaffected by passing to the radical, we will assume without loss of generality that

$$I = (e).$$

I claim that $\operatorname{Spec} R - V(I)$ is just $V(1 - e) = V((1 - e))$. This is a closed set, so proving this claim will imply that $V(I)$ is open. Indeed, $V(e) = V((e))$ cannot intersect $V(1 - e)$ because if

$$\mathfrak{p} \in V(e) \cap V(1 - e),$$

then $e, 1 - e \in \mathfrak{p}$, so $1 \in \mathfrak{p}$. This is a contradiction since $\mathfrak{p}$ is necessarily prime.

[6] More generally, in any *ringed space* (a space with a sheaf of rings), the idempotents in the ring of global sections correspond to the disconnections of the topological space.

Conversely, suppose that $\mathfrak{p} \in \operatorname{Spec} R$ belongs to neither $V(e)$ nor $V(1-e)$. Then $e \notin \mathfrak{p}$ and $1 - e \notin \mathfrak{p}$. So the product

$$e(1-e) = e - e^2 = 0$$

cannot lie in $\mathfrak{p}$. But necessarily $0 \in \mathfrak{p}$, contradiction. So $V(e) \cup V(1-e) = \operatorname{Spec} R$. This implies the claim.

Next, we show that if $V(I)$ is open, then I is the radical of (e) for an idempotent e. For this it is sufficient to prove:

1.3.11 Lemma *Let $I \subset R$ be such that $V(I) \subset \operatorname{Spec} R$ is open. Then I is principal, generated by (e) for some idempotent $e \in R$.*

Proof. Suppose that $\operatorname{Spec} R - V(I) = V(J)$ for some ideal $J \subset R$. Then the intersection $V(I) \cap V(J) = V(I+J)$ is all of R, so $I + J$ cannot be a proper ideal (or it would be contained in a prime ideal). In particular, $I + J = R$. So we can write

$$1 = x + y, \quad x \in I, y \in J.$$

Now $V(I) \cup V(J) = V(IJ) = \operatorname{Spec} R$. This implies that every element of IJ is nilpotent by the next lemma. ☐

1.3.12 Lemma *Suppose $V(X) = \operatorname{Spec} R$ for $X \subset R$ an ideal. Then every element of X is nilpotent.*

Proof. Indeed, suppose $x \in X$ were non-nilpotent. Then the ring R_x is not the zero ring, so it has a prime ideal. The map $\operatorname{Spec} R_x \to \operatorname{Spec} R$ is, as discussed in class, a homeomorphism of $\operatorname{Spec} R_x$ onto $D(x)$. So $D(x) \subset \operatorname{Spec} R$ (the collection of primes not containing x) is nonempty. In particular, there is $\mathfrak{p} \in \operatorname{Spec} R$ with $x \notin \mathfrak{p}$, so $\mathfrak{p} \notin V(X)$. So $V(X) \neq \operatorname{Spec} R$, contradiction. ☐

Return to the proof of the main result. We have shown that IJ is nilpotent. In particular, in the expression $x + y = 1$ we had earlier, we have that xy is nilpotent. Say $(xy)^k = 0$. Then expand

$$1 = (x+y)^{2k} = \sum_{i=0}^{2k} \binom{2k}{i} x^i y^{2k-i} = \sum{}' + \sum{}''$$

where $\sum'$ is the sum from $i = 0$ to $i = k$ and $\sum''$ is the sum from $k+1$ to $2k$. Then $\sum' \sum'' = 0$ because in every term occurring in the expansion, a multiple of $x^k y^k$ occurs. Also, $\sum' \in I$ and $\sum'' \in J$ because $x \in I, y \in J$.

All in all, we find that it is possible to write

$$1 = x' + y', \quad x' \in I, y' \in J, \ x'y' = 0.$$

(We take $x' = \sum', y' = \sum''$.) Then $x'(1 - x') = 0$ so $x' \in I$ is idempotent. Similarly $y' = 1 - x'$ is. We have that

$$V(I) \subset V(x'), \quad V(J) \subset V(y')$$

and $V(x'), V(y')$ are complementary by the earlier arguments, so necessarily

$$V(I) = V(x'), \quad V(J) = V(y').$$

Since an ideal generated by an idempotent is automatically radical, it follows that:

$$I = (x'), \quad , J = (y'). \qquad \square$$

There are some useful applications of this in representation theory, because one can look for idempotents in endomorphism rings; these indicate whether a module can be decomposed as a direct sum into smaller parts. Except, of course, that endomorphism rings aren't necessarily commutative and this proof breaks down.

Thus we get:

1.3.13 Proposition *Let A be a ring and I a nilpotent ideal. Then* $\mathrm{Idem}(A) \to \mathrm{Idem}(A/I)$ *is bijective.*

Proof. Indeed, the topological spaces of Spec A and Spec A/I are the same. The result then follows from **??**. $\square$

Units

Finally, we make a few remarks on *units* modulo nilideals. It is a useful and frequently used trick that adding a nilpotent does not affect the collection of units. This trick is essentially an algebraic version of the familiar "geometric series;" convergence questions do not appear thanks to nilpotence.

1.3.14 Example Suppose u is a unit in a ring R and $v \in R$ is nilpotent; we show that $a + v$ is a unit.

Suppose $ua = 1$ and $v^m = 0$ for some $m > 1$. Then $(u + v) \cdot a(1 - av + (av)^2 - \cdots \pm (av)^{m-1}) = (1 - (-av))(1 + (-av) + (-av)^2 + \cdots + (-av)^{m-1}) = 1 - (-av)^m = 1 - 0 = 1$, so $u + v$ is a unit.

So let R be a ring, $I \subset R$ a nilpotent ideal *of square zero.* Let R^* denote the group of units in R, as usual, and let $(R/I)^*$ denote the group of units in R/I. We have an exact sequence of abelian groups:

$$0 \to I \to R^* \to (R/I)^* \to 0$$

where the second map is reduction and the first map sends $i \to 1+i$. The hypothesis that $I^2 = 0$ shows that the first map is a homomorphism. We should check that the last map is surjective. But if any $a \in R$ maps to a unit in R/I, it clearly can lie in no prime ideal of R, so is a unit itself.

1.4. Vista: sheaves on $\operatorname{Spec} R$

Presheaves

Let X be a topological space.

1.4.1 Definition A **presheaf of sets** $\mathcal{F}$ on X assigns to every open subset $U \subset X$ a set $\mathcal{F}(U)$, and to every inclusion $U \subset V$ a **restriction map** $\mathrm{res}_U^V : \mathcal{F}(V) \to \mathcal{F}(U)$. The restriction map is required to satisfy:

1. $\mathrm{res}_U^U = \mathrm{id}_{\mathcal{F}(U)}$ for all open sets U.
2. $\mathrm{res}_U^W = \mathrm{res}_U^V \circ \mathrm{res}_V^W$ if $U \subset V \subset W$.

If the sets $\mathcal{F}(U)$ are all groups (resp. rings), and the restriction maps are morphisms of groups (resp. rings), then we say that $\mathcal{F}$ is a sheaf of groups (resp. rings). Often the restriction of an element $a \in U$ to a subset W is denoted $a|_W$.

A **morphism** of presheaves $\mathcal{F} \to \mathcal{G}$ is a collection of maps $\mathcal{F}(U) \to \mathcal{G}(U)$ for each open set U, that commute with the restriction maps in the obvious way. Thus the collection of presheaves on a topological space forms a category.

One should think of the restriction maps as kind of like restricting the domain of a function. The standard example of presheaves is given in this way, in fact.

1.4.2 Example Let X be a topological space, and $\mathcal{F}$ the presheaf assigning to each $U \subset X$ the set of continuous functions $U \to \mathbb{R}$. The restriction maps come from restricting the domain of a function.

Now, in classical algebraic geometry, there are likely to be more continuous functions in the Zariski topology than one really wants. One wants to focus on functions that are given by polynomial equations.

1.4.3 Example Let X be the topological space $\mathbb{C}^n$ with the topology where the closed sets are those defined by the zero loci of polynomials (that is, the topology induced on $\mathbb{C}^n$ from the Zariski topology of $\operatorname{Spec} \mathbb{C}[x_1, \ldots, x_n]$ via the canonical imbedding $\mathbb{C}^n \hookrightarrow \operatorname{Spec} \mathbb{C}[x_1, \ldots, x_n]$). Then there is a presheaf assigning to each open set U the collection of rational functions defined everywhere on U, with the restriction maps being the obvious ones.

1.4.4 Remark The notion of presheaf thus defined relied very little on the topology of X. In fact, we could phrase it in purely categorical terms. Let $\mathcal{C}$ be the category consisting of open subsets $U \subset X$ and inclusions of open subsets $U \subset U'$. This is a rather simple category (the hom-sets are either empty or consist of one element). Then a *presheaf* is just a contravariant functor from $\mathcal{C}$ to **Sets** (or **Grp**, etc.). A morphism of presheaves is a natural transformation of functors.

In fact, given any category $\mathcal{C}$, we can define the *category of presheaves* on it to be the category of functors $\mathbf{Fun}(\mathcal{C}^{op}, \mathbf{Set})$. This category is complete and cocomplete (we can calculate limits and colimits "pointwise"), and the Yoneda embedding realizes $\mathcal{C}$ as a full subcategory of it. So if

$X \in \mathcal{C}$, we get a presheaf $Y \mapsto \hom_{\mathcal{C}}(Y, X)$. In general, however, such representable presheaves are rather special; for instance, what do they look like for the category of open sets in a topological space?

Sheaves

1.4.5 Definition Let $\mathcal{F}$ be a presheaf of sets on a topological space X. We call $\mathbb{F}$ a **sheaf** if $\mathcal{F}$ further satisfies the following two "sheaf conditions."

1. (Separatedness) If U is an open set of X covered by a family of open subsets $\{U_i\}$ and there are two elements $a, b \in \mathcal{F}(U)$ such that $a|_{U_i} = b|_{U_i}$ for all U_i, then $a = b$.
2. (Gluability) If U is an open set of X covered by U_i and there are elements $a_i \in \mathcal{F}(U_i)$ such that $a_i|_{U_i \cap U_j} = a_j|_{U_i \cap U_j}$ for all i and j, then there exists an element $a \in \mathcal{F}(U)$ that restricts to the a_i. Notice that by the first axiom, this element is unique.

A *morphism* of sheaves is just a morphism of presheaves, so the sheaves on a topological space X form a full subcategory of presheaves on X.

The above two conditions can be phrased more compactly as follows. Whenever $\{U_i\}_{i \in I}$ is an open cover of $U \subset X$, we require that the following sequence be an equalizer of sets:

$$\mathcal{F}(U) \to \prod_{i \in I} \mathcal{F}(U_i) \rightrightarrows \prod_{i,j \in I} \mathcal{F}(U_i \cap U_j)$$

where the two arrows correspond to the two allowable restriction maps. Similarly, we say that a presheaf of abelian groups (resp. rings) is a **sheaf** if it is a sheaf of sets.

1.4.6 Example The example of functions gives an example of a sheaf, because functions are determined by their restrictions to an open cover! Namely, if X is a topological space, and we consider the presheaf

$$U \mapsto \{\text{continuous functions } U \to \mathbb{R}\},$$

then this is clearly a presheaf, because we can piece together continuous functions in a unique manner.

1.4.7 Example Here is a refinement of the above example. Let X be a smooth manifold. For each U, let $\mathcal{F}(U)$ denote the group of smooth functions $U \to \mathbb{R}$. This is easily checked to be a sheaf.

We could, of course, replace "smooth" by "C^r" or by "holomorphic" in the case of a complex manifold.

1.4.8 Remark As remarked above, the notion of presheaf can be defined on any category, and does not really require a topological space. The definition of a sheaf requires a bit more topologically, because the idea that a family $\{U_i\}$ *covers* an open set U was used inescapably in the definition. The idea of covering required the internal structure of the open sets and was not a purely categorical idea. However, Grothendieck developed a way to axiomatize this, and introduced the idea of a *Grothendieck topology* on a category (which is basically a notion of when a family of maps *covers* something). On a category with a Grothendieck topology (also known as a *site*), one can define the notion of a sheaf in a similar manner as above. See **?**.

There is a process that allows one to take any presheaf and associate a sheaf to it. In some sense, this associated sheaf should also be the best "approximation" of our presheaf with a sheaf. This motivates the following universal property:

1.4.9 Definition Let $\mathcal{F}$ be a presheaf. Then $\mathcal{F}'$ is said to be the sheafification of $\mathcal{F}$ if for any sheaf $\mathcal{G}$ and a morphism $\mathcal{F} \to \mathcal{G}$, there is a unique factorization of this morphism as $\mathcal{F} \to \mathcal{F}' \to \mathcal{G}$.

1.4.10 Theorem *We can construct the sheafification of a presheaf $\mathcal{F}$ as follows: $\mathcal{F}'(U) = \{s : U \to \coprod_{x \in U} \mathcal{F}_x |$for all $x \in U, s(x) \in \mathcal{F}_x$ and there is a neighborhood $V \subset U$ and $t \in \mathcal{F}(V)$ such that for all $y \in V, s(y)$ is the image of t in the local ring $\mathcal{F}_y\}$.*

add: proof

In the theory of schemes, when one wishes to replace polynomial rings over $\mathbb{C}$ (or an algebraically closed field) with arbitrary commutative rings, one must drop the idea that a sheaf is necessarily given by functions. A *scheme* is defined as a space with a certain type of sheaf of rings on it. We shall not define a scheme formally, but show how on the building blocks of schemes—objects of the form Spec A—a sheaf of rings can be defined.

Sheaves on Spec A

add: we need to describe how giving sections over basic open sets gives a presheaf in general.

1.4.11 Proposition *Let A be a ring and let $X = \operatorname{Spec}(A)$. Then the assignment of the ring A_f to the basic open set X_f defines a presheaf of rings on X.*

Proof.

Part (i). If $X_g \subset X_f$ are basic open sets, then there exist $n \geqslant 1$ and $u \in A$ such that $g^n = uf$.

Proof of part (i). Let $S = \{g^n : n \geqslant 0\}$ and suppose $S \cap (f) = \varnothing$. Then the extension $(f)^e$ into $S^{-1}A$ is a proper ideal, so there exists a maximal ideal $S^{-1}\mathfrak{p}$ of $S^{-1}A$, where $\mathfrak{p} \cap S = \varnothing$. Since $(f)^e \in S^{-1}\mathfrak{p}$, we see that $f/1 \in S^{-1}\mathfrak{p}$, so $f \in \mathfrak{p}$. But $S \cap \mathfrak{p} = \varnothing$ implies that $g \notin \mathfrak{p}$. This is a contradiction, since then $\mathfrak{p} \in X_g \backslash X_f$.

Part (ii). If $X_g \subset X_f$, then there exists a unique map $\rho : A_f \to A_g$, called the restriction map, which makes the following diagram commute.

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] & \\ A_f \arrow[rr] & & A_g \end{tikzcd}$$

Proof of part (ii). Let $n \geqslant 1$ and $u \in A$ be such that $g^n = uf$ by part (i). Note that in A_g,

$$(f/1)(u/g^n) = (fu/g^n) = 1/1 = 1$$

which means that f maps to a unit in A_g. Hence every f^m maps to a unit in A_g, so the universal property of A_f yields the desired unique map $\rho : A_f \to A_g$.

Part (iii). If $X_g = X_f$, then the corresponding restriction $\rho : A_f \to A_g$ is an isomorphism.

Proof of part (iii). The reverse inclusion yields a $\rho' : A_g \to A_f$ such that the diagram

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] & \\ A_f \arrow[rr, bend left, "\rho"] & & A_g \arrow[ll, bend left, "\rho'"] \end{tikzcd}$$

commutes. But since the localization map is epic, this implies that $\rho\rho' = \rho'\rho = \mathbf{1}$.

Part (iv). If $X_h \subset X_g \subset X_f$, then the diagram

$$\begin{tikzcd} A_f \arrow[rr] \arrow[dr] & & A_h \\ & A_g \arrow[ur] & \end{tikzcd}$$

of restriction maps commutes.

Proof of part (iv). Consider the following tetrahedron.

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] \arrow[dd] & \\ A_f \arrow[rr, dotted] \arrow[dr] & & A_h \\ & A_g \arrow[ur] & \end{tikzcd}$$

Except for the base, the commutativity of each face of the tetrahedron follows from the universal property of part (ii). But its easy to see that commutativity of the those faces implies commutativity of the base, which is what we want to show.

Part (v). If $X_{\tilde{g}} = X_g \subset X_f = X_{\tilde{f}}$, then the diagram

$$\begin{tikzcd} A_f \arrow[r] \arrow[d] & A_g \arrow[d] \\ A_{\tilde{f}} \arrow[r] & A_{\tilde{g}} \end{tikzcd}$$

of restriction maps commutes. (Note that the vertical maps here are isomorphisms.)

Proof of part (v). By part (iv), the two triangles of

$$\begin{tikzcd} A_f \arrow[r] \arrow[d] \arrow[dr] & A_g \arrow[d] \\ A_{\tilde{f}} \arrow[r] & A_{\tilde{g}} \end{tikzcd}$$

commute. Therefore the square commutes.

Part (vi). Fix a prime ideal $\mathfrak{p}$ in A. Consider the direct system consisting of rings A_f for every $f \notin \mathfrak{p}$ and restriction maps $\rho_{fg} : A_f \to A_g$ whenever $X_g \subset X_f$. Then $\varinjlim A_f \cong A_{\mathfrak{p}}$.

proof of part (vi). First, note that since $f \notin \mathfrak{p}$ and $\mathfrak{p}$ is prime, we know that $f^m \notin \mathfrak{p}$ for all $m \geqslant 0$. Therefore the image of f^m under the localization $A \to A_{\mathfrak{p}}$ is a unit, which means the universal property of A_f yields a unique map $\alpha_f : A_f \to A_{\mathfrak{p}}$ such that the following diagram commutes.

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] & \\ A_f \arrow[rr, "\alpha_f"] & & A_{\mathfrak{p}} \end{tikzcd}$$

Then consider the following tetrahedron.

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] \arrow[dd] & \\ A_f \arrow[rr, dotted] \arrow[dr] & & A_h \\ & A_{\mathfrak{p}} \arrow[ur] & \end{tikzcd}$$

All faces except the bottom commute by construction, so the bottom face commutes as well. This implies that the α_f commute with the restriction maps, as necessary. Now, to see that $\varinjlim A_f \cong A_{\mathfrak{p}}$, we show that $A_{\mathfrak{p}}$ satisfies the universal property of $\varinjlim A_f$.

Suppose B is a ring and there exist maps $\beta_f : A_f \to B$ which commute with the restrictions. Define $\beta : A \to B$ as the composition $A \to A_f \to B$. The fact that β is independent of choice of f follows from the commutativity of the following diagram.

$$\begin{tikzcd} & A \arrow[dl] \arrow[dr] & \\ A_f \arrow[rr, "\rho_{fg}"] \arrow[dr, "\beta_f"'] & & A_g \arrow[dl, "\beta_g"] \\ & B & \end{tikzcd}$$

Now, for every $f \notin \mathfrak{p}$, we know that $\beta(f)$ must be a unit since $\beta(f) = \beta_f(f/1)$ and $f/1$ is a unit in A_f. Therefore the universal property of $A_{\mathfrak{p}}$ yields a unique map $A_{\mathfrak{p}} \to B$, which clearly commutes with all the arrows necessary to make $\varinjlim A_f \cong A_{\mathfrak{p}}$. □

1.4.12 Proposition *Let A be a ring and let $X = \operatorname{Spec}(A)$. The presheaf of rings $\mathcal{O}_X$ defined on X is a sheaf.*

Proof. The proof proceeds in two parts. Let $(U_i)_{i \in I}$ be a covering of X by basic open sets.

Part 1. If $s \in A$ is such that $s_i := \rho_{X,U_i}(s) = 0$ for all $i \in I$, then $s = 0$.

Proof of part 1. Suppose $U_i = X_{f_i}$. Note that s_i is the fraction $s/1$ in the ring A_{f_i}, so $s_i = 0$ implies that there exists some integer m_i such that $sf_i^{m_i} = 0$. Define $g_i = f_i^{m_i}$, and note that we still have an open cover by sets X_{g_i} since $X_{f_i} = X_{g_i}$ (a prime ideal contains an element if and only if it contains every power of that element). Also $sg_i = 0$, so the fraction $s/1$ is still 0 in the ring A_{g_i}. (Essentially, all we're observing here is that we are free to change representation of the basic open sets in our cover to make notation more convenient).

Since X is quasi-compact, choose a finite subcover $X = X_{g_1} \cup \cdots \cup X_{g_n}$. This means that $g_1, \ldots, g_n$ must generate the unit ideal, so there exists some linear combination $\sum x_i g_i = 1$ with $x_i \in A$. But then

$$s = s \cdot 1 = s\left(\sum x_i g_i\right) = \sum x_i(sg_i) = 0.$$

Part 2. Let $s_i \in \mathcal{O}_X(U_i)$ be such that for every $i, j \in I$,

$$\rho_{U_i, U_i \cap U_j}(s_i) = \rho_{U_j, U_i \cap U_j}(s_j).$$

(That is, the collection $(s_i)_{i \in I}$ agrees on overlaps). Then there exists a unique $s \in A$ such that $\rho_{X, U_i}(s) = s_i$ for every $i \in I$.

Proof of part 2. Let $U_i = X_{f_i}$, so that $s_i = a_i/(f_i^{m_i})$ for some integers m_i. As in part 1, we can clean up notation by defining $g_i = f_i^{m_i}$, so that $s_i = a_i/g_i$. Choose a finite subcover $X = X_{g_1} \cup \cdots \cup X_{g_n}$. Then the condition that the cover agrees on overlaps means that

$$\frac{a_i g_j}{g_i g_j} = \frac{a_j g_i}{g_i g_j}$$

for all i, j in the finite subcover. This is equivalent to the existence of some k_{ij} such that

$$(a_i g_j - a_j g_i)(g_i g_j)^{k_{ij}} = 0.$$

Let k be the maximum of all the k_{ij}, so that $(a_i g_j - a_j g_i)(g_i g_j)^k = 0$ for all i, j in the finite subcover. Define $b_i = a_i g_i^k$ and $h_i = g_i^{k+1}$. We make the following observations:

$$b_i h_j - b_j h_i = 0, X_{g_i} = X_{h_i}, \text{ and } s_i = a_i/g_i = b_i/h_i$$

The first observation implies that the X_{h_i} cover X, so the h_i generate the unit ideal. Then there exists some linear combination $\sum x_i h_i = 1$. Define $s = \sum x_i b_i$. I claim that this is the global section that restricts to s_i on the open cover.

The first step is to show that it restricts to s_i on our chosen finite subcover. In other words, we want to show that $s/1 = s_i = b_i/h_i$ in A_{h_i}, which is equivalent to the condition that there exist some l_i such that $(sh_i b_i)h_i^{l_i} = 0$. But in fact, even $l_i = 0$ works:

$$sh_i - b_i = \left(\sum x_j b_j\right) h_i - b_i \left(\sum x_j h_j\right) = \sum x_j \left(h_i b_j - b_i h_j\right) = 0.$$

This shows that s restricts to s_i on each set in our finite subcover. Now we need to show that in fact, it restricts to s_i for all of the sets in our cover. Choose any $j \in I$. Then $U_1, \ldots, U_n, U_j$ still cover X, so the above process yields an s' such that s' restricts to s_i for all $i \in \{1, \ldots, n, j\}$. But then $s - s'$ satisfies the assumptions of part 1 using the cover $\{U_1, \ldots, U_n, U_j\}$, so this means $s = s'$. Hence the restriction of s to U_j is also s_j. $\square$

III.2. Noetherian rings and modules

The finiteness condition of a noetherian ring is necessary for much of commutative algebra; many of the results we prove after this will apply only (or mostly) to the noetherian case. In algebraic geometry, the noetherian condition guarantees that the topological space associated to the ring (the Spec) has all its sets quasi-compact; this condition can be phrased as saying that the space itself is noetherian in a certain sense.

We shall start by proving the basic properties of noetherian rings. These are fairly standard and straightforward; they could have been placed after **??**, in fact. More subtle is the structure theory for finitely generated modules over a noetherian ring. While there is nothing as concrete as there is for PIDs (there, one has a very explicit descrition for the isomorphism classes), one can still construct a so-called "primary decomposition." This will be the primary focus after the basic properties of noetherian rings and modules have been established. Finally, we finish with an important subclass of noetherian rings, the *artinian* ones.

2.1. Basics

The noetherian condition

2.1.1 Definition Let R be a commutative ring and M an R-module. We say that M is **noetherian** if every submodule of M is finitely generated.

There is a convenient reformulation of the finiteness hypothesis above in terms of the *ascending chain condition.*

2.1.2 Proposition *M is a module over R. The following are equivalent:*

1. *M is noetherian.*
2. *Every chain of submodules $M_0 \subset M_1 \subset \cdots \subset M$, eventually stabilizes at some M_N. (Ascending chain condition.)*
3. *Every nonempty collection of submodules of M has a maximal element.*

Proof. Say M is noetherian and we have such a chain

$$M_0 \subset M_1 \subset \dots .$$

Write

$$M' = \bigcup M_i \subset M,$$

which is finitely generated since M is noetherian. Let it be generated by $x_1, \dots, x_n$. Each of these finitely many elements is in the union, so they are all contained in some M_N. This means that

$$M' \subset M_N, \quad \text{so} \quad M_N = M'$$

and the chain stabilizes.

For the converse, assume the ACC. Let $M' \subset M$ be any submodule. Define a chain of submodules $M_0 \subset M_1 \subset \cdots \subset M'$ inductively as follows. First, just take $M_0 = \{0\}$. Take M_{n+1} to be $M_n + Rx$ for some $x \in M' - M_n$, if such an x exists; if not take $M_{n+1} = M_n$. So M_0 is zero, M_1 is generated by some nonzero element of M', M_2 is M_1 together with some element of M' not in M_1, and so on, until (if ever) the chain stabilizes.

However, by construction, we have an ascending chain, so it stabilizes at some finite place by the ascending chain condition. Thus, at some point, it is impossible to choose something in M' that does not belong to M_N. In particular, M' is generated by N elements, since M_N is and $M' = M_N$. This proves the reverse implication. Thus the equivalence of 1 and 2 is clear. The equivalence of 2 and 3 is left to the reader. $\square$

Working with noetherian modules over non-noetherian rings can be a little funny, though, so normally this definition is combined with:

2.1.3 Definition The ring R is **noetherian** if R is noetherian as an R-module. Equivalently phrased, R is noetherian if all of its ideals are finitely generated.

We start with the basic examples:

2.1.4 Example

1. Any field is noetherian. There are two ideals: (1) and (0).
2. Any PID is noetherian: any ideal is generated by one element. So $\mathbb{Z}$ is noetherian.

The first basic result we want to prove is that over a noetherian ring, the noetherian modules are precisely the finitely generated ones. This will follow from 2.1.7 in the next subsec. So the defining property of noetherian rings is that a submodule of a finitely generated module is finitely generated. (Compare 2.1.10.)

2.1.5 Remark The ring $\mathbb{C}[X_1, X_2, \dots]$ of polynomials in infinitely many variables is not noetherian. Note that the ring itself is finitely generated (by the element 1), but there are ideals that are not finitely generated.

2.1.6 Remark Let R be a ring such that every *prime* ideal is finitely generated. Then R is noetherian. See 1.2.23, or prove it as an exercise.

Stability properties

The class of noetherian rings is fairly robust. If one starts with a noetherian ring, most of the elementary operations one can do to it lead to noetherian rings.

2.1.7 Proposition *If*

$$0 \to M' \to M \to M'' \to 0$$

is an exact sequence of modules, then M is noetherian if and only if M', M'' are.

One direction states that noetherianness is preserved under subobjects and quotients. The other direction states that noetherianness is preserved under extensions.

Proof. If M is noetherian, then every submodule of M' is a submodule of M, so is finitely generated. So M' is noetherian too. Now we show that M'' is noetherian. Let $N \subset M''$ and let $\widetilde{N} \subset M$ the inverse image. Then $\widetilde{N}$ is finitely generated, so N—as the homomorphic image of $\widetilde{N}$—is finitely generated So M'' is noetherian.

Suppose M', M'' noetherian. We prove M noetherian. We verify the ascending chain condition. Consider

$$M_1 \subset M_2 \subset \cdots \subset M.$$

Let M_i'' denote the image of M_i in M'' and let M_i' be the intersection of M_i with M'. Here we think of M' as a submodule of M. These are ascending chains of submodules of M', M'', respectively, so they stabilize by noetherianness. So for some N, we have that $n \geqslant N$ implies

$$M_n' = M_{n+1}', \quad M_n'' = M_{n+1}''.$$

We claim that this implies, for such n,

$$M_n = M_{n+1}.$$

Indeed, say $x \in M_{n+1} \subset M$. Then x maps into something in $M_{n+1}'' = M_n''$. So there is something in M_n, call it y, such that x, y go to the same thing in M''. In particular,

$$x - y \in M_{n+1}$$

goes to zero in M'', so $x - y \in M'$. Thus $x - y \in M_{n+1}' = M_n'$. In particular,

$$x = (x - y) + y \in M_n' + M_n = M_n.$$

So $x \in M_n$, and

$$M_n = M_{n+1}.$$

This proves the result. ∎

The class of noetherian modules is thus "robust." We can get from that the following.

2.1.8 Proposition *If $\phi : A \to B$ is a surjection of commutative rings and A is noetherian, then B is noetherian too.*

Proof. Indeed, B is noetherian as an A-module; indeed, it is the quotient of a noetherian A-module (namely, A). However, it is easy to see that the A-submodules of B are just the B-modules in B, so B is noetherian as a B-module too. So B is noetherian. □

We know show that noetherianness of a ring is preserved by localization:

2.1.9 Proposition *Let R be a commutative ring, $S \subset R$ a multiplicatively closed subset. If R is noetherian, then $S^{-1}R$ is noetherian.*

I.e., the class of noetherian rings is closed under localization.

Proof. Say $\phi : R \to S^{-1}R$ is the canonical map. Let $I \subset S^{-1}R$ be an ideal. Then $\phi^{-1}(I) \subset R$ is an ideal, so finitely generated. It follows that

$$\phi^{-1}(I)(S^{-1}R) \subset S^{-1}R$$

is finitely generated as an ideal in $S^{-1}R$; the generators are the images of the generators of $\phi^{-1}(I)$.

Now we claim that

$$\phi^{-1}(I)(S^{-1}R) = I.$$

The inclusion $\subset$ is trivial. For the latter inclusion, if $x/s \in I$, then $x \in \phi^{-1}(I)$, so

$$x = (1/s)x \in (S^{-1}R)\phi^{-1}(I).$$

This proves the claim and implies that I is finitely generated. □

Let R be a noetherian ring. We now characterize the noetherian R-modules.

2.1.10 Proposition *An R-module M is noetherian if and only if M is finitely generated.*

Proof. The only if direction is obvious. A module is noetherian if and only if every submodule is finitely generated.

For the if direction, if M is finitely generated, then there is a surjection of R-modules

$$R^n \to M$$ □

where R is noetherian. But R^n is noetherian by 2.1.7 because it is a direct sum of copies of R. So M is a quotient of a noetherian module and is noetherian.

The basis theorem

Let us now prove something a little less formal. This is, in fact, the biggest of the "stability" properties of a noetherian ring: we are going to see that finitely generated algebras over noetherian rings are still noetherian.

2.1.11 Theorem (Hilbert basis theorem) *If R is a noetherian ring, then the polynomial ring $R[X]$ is noetherian.*

Proof. Let $I \subset R[X]$ be an ideal. We prove that it is finitely generated. For each $m \in \mathbb{Z}_{\geqslant 0}$, let $I(n)$ be the collection of elements $a \in R$ consisting of the coefficients of x^n of elements of I of degree $\leqslant n$. This is an ideal, as is easily seen.

In fact, we claim that

$$I(1) \subset I(2) \subset \dots$$

which follows because if $a \in I(1)$, there is an element $aX + \dots$ in I. Thus $X(aX + \dots) = aX^2 + \dots \in I$, so $a \in I(2)$. And so on.

Since R is noetherian, this chain stabilizes at some $I(N)$. Also, because R is noetherian, each $I(n)$ is generated by finitely many elements $a_{n,1}, \dots, a_{n,m_n} \in I(n)$. All of these come from polynomials $P_{n,i} \in I$ such that $P_{n,i} = a_{n,i}X^n + \dots$.

The claim is that the $P_{n,i}$ for $n \leqslant N$ and $i \leqslant m_n$ generate I. This is a finite set of polynomials, so if we prove the claim, we will have proved the basis theorem. Let J be the ideal generated by $\{P_{n,i}, n \leqslant N, i \leqslant m_n\}$. We know $J \subset I$. We must prove $I \subset J$.

We will show that any element $P(X) \in I$ of degree n belongs to J by induction on n. The degree is the largest nonzero coefficient. In particular, the zero polynomial does not have a degree, but the zero polynomial is obviously in J.

There are two cases. In the first case, $n \geqslant N$. Then we write

$$P(X) = aX^n + \dots.$$

By definition, $a \in I(n) = I(N)$ since the chain of ideals $I(n)$ stabilized. Thus we can write a in terms of the generators: $a = \sum a_{N,i}\lambda_i$ for some $\lambda_i \in R$. Define the polynomial

$$Q = \sum \lambda_i P_{N,i} x^{n-N} \in J.$$

Then Q has degree n and the leading term is just a. In particular,

$$P - Q$$

is in I and has degree less than n. By the inductive hypothesis, this belongs to J, and since $Q \in J$, it follows that $P \in J$.

Now consider the case of $n < N$. Again, we write $P(X) = aX^n + \dots$. Then $a \in I(n)$. We can write

$$a = \sum a_{n,i}\lambda_i, \quad \lambda_i \in R.$$

But the $a_{n,i} \in I(n)$. The polynomial

$$Q = \sum \lambda_i P_{n,i}$$

belongs to J since $n < N$. In the same way, $P - Q \in I$ has a lower degree. Induction as before implies that $P \in J$. $\square$

2.1.12 Example Let k be a field. Then $k[x_1, \ldots, x_n]$ is noetherian for any n, by the Hilbert basis theorem and induction on n.

2.1.13 Corollary *If R is a noetherian ring and R' a finitely generated R-algebra, then R' is noetherian too.*

Proof. Each polynomial ring $R[X_1, \ldots, X_n]$ is noetherian by theorem 2.1.11 and an easy induction on n. Consequently, any quotient of a polynomial ring (i.e. any finitely generated R-algebra, such as R') is noetherian. $\square$

2.1.14 Example Any finitely generated commutative ring R is noetherian. Indeed, then there is a surjection

$$\mathbb{Z}[x_1, \ldots, x_n] \twoheadrightarrow R$$

where the x_i get mapped onto generators in R. The former is noetherian by the basis theorem, and R is as a quotient noetherian.

2.1.15 Corollary *Any ring R can be obtained as a filtered direct limit of noetherian rings.*

Proof. Indeed, R is the filtered direct limit of its finitely generated subrings. $\square$

This observation is sometimes useful in commutative algebra and algebraic geometry, in order to reduce questions about arbitrary commutative rings to noetherian rings. Noetherian rings have strong finiteness hypotheses that let you get numerical invariants that may be useful. For instance, we can do things like inducting on the dimension for noetherian local rings.

2.1.16 Example Take $R = \mathbb{C}[x_1, \ldots, x_n]$. For any algebraic variety V defined by polynomial equations, we know that V is the vanishing locus of some ideal $I \subset R$. Using the Hilbert basis theorem, we have shown that I is finitely generated. This implies that V can be described by *finitely* many polynomial equations.

Noetherian induction

The finiteness condition on a noetherian ring allows for "induction" arguments to be made; we shall see examples of this in the future.

2.1.17 Proposition (Noetherian Induction Principle) *Let R be a noetherian ring, let $\mathcal{P}$ be a property, and let $\mathcal{F}$ be a family of ideals R. Suppose the inductive step: if all ideals in $\mathcal{F}$ strictly larger than $I \in \mathcal{F}$ satisfy $\mathcal{P}$, then I satisfies $\mathcal{P}$. Then all ideals in $\mathcal{F}$ satisfy $\mathcal{P}$.*

Proof. Assume $\mathcal{F}_{\text{crim}} = \{J \in \mathcal{F} | J \text{ does not satisfy } \mathcal{P}\} \neq \varnothing$. Since R is noetherian, $\mathcal{F}_{\text{crim}}$ has a maximal member I. By maximality, all ideals in $\mathcal{F}$ strictly containing I satisfy $\mathcal{P}$, so I also does by the inductive step. $\square$

2.2. Associated primes

We shall now begin the structure theory for noetherian modules. The first step will be to associate to each module a collection of primes, called the *associated primes*, which lie in a bigger collection of primes (the *support*) where the localizations are nonzero.

The support

Let R be a noetherian ring. An R-module M is supposed to be thought of as something like a vector bundle, somehow spread out over the topological space $\operatorname{Spec} R$. If $\mathfrak{p} \in \operatorname{Spec} R$, then let $\Bbbk(\mathfrak{p}) = \text{fr. field } R/\mathfrak{p}$, which is the residue field of $R_{\mathfrak{p}}$. If M is any R-module, we can consider $M \otimes_R \Bbbk(\mathfrak{p})$ for each $\mathfrak{p}$; it is a vector space over $\Bbbk(\mathfrak{p})$. If M is finitely generated, then $M \otimes_R \Bbbk(\mathfrak{p})$ is a finite-dimensional vector space.

2.2.1 Definition Let M be a finitely generated R-module. Then $\operatorname{supp} M$, the **support** of M, is defined to be the set of primes $\mathfrak{p} \in \operatorname{Spec} R$ such that $M \otimes_R \Bbbk(\mathfrak{p}) \neq 0$.

One is supposed to think of a module M as something like a vector bundle over the topological space $\operatorname{Spec} R$. At each $\mathfrak{p} \in \operatorname{Spec} R$, we associate the vector space $M \otimes_R \Bbbk(\mathfrak{p})$; this is the “fiber.” Of course, the intuition of M's being a vector bundle is somewhat limited, since the fibers do not generally have the same dimension. Nonetheless, we can talk about the support, i.e. the set of spaces where the “fiber” is not zero.

Note that $\mathfrak{p} \in \operatorname{supp} M$ if and only if $M_{\mathfrak{p}} \neq 0$. This is because

$$(M \otimes_R R_{\mathfrak{p}})/(\mathfrak{p}R_{\mathfrak{p}}(M \otimes_R R_{\mathfrak{p}})) = M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} \Bbbk(\mathfrak{p})$$

and we can use Nakayama's lemma over the local ring $R_{\mathfrak{p}}$. (We are using the fact that M is finitely generated.)

A vector bundle whose support is empty is zero. Thus the following easy proposition is intuitive:

2.2.2 Proposition *$M = 0$ if and only if $\operatorname{supp} M = \varnothing$.*

Proof. Indeed, $M = 0$ if and only if $M_{\mathfrak{p}} = 0$ for all primes $\mathfrak{p} \in \operatorname{Spec} R$. This is equivalent to $\operatorname{supp} M = \varnothing$. ∎

2.2.3 Remark Let $0 \to M' \to M \to M'' \to 0$ be exact. Then

$$\operatorname{supp} M = \operatorname{supp} M' \cup \operatorname{supp} M''.$$

We will see soon that that $\operatorname{supp} M$ is closed in $\operatorname{Spec} R$. One imagines that M lives on this closed subset $\operatorname{supp} M$, in some sense.

Associated primes

Throughout this section, R is a noetherian ring. The *associated primes* of a module M will refer to primes that arise as the annihilators of elements in M. As we shall see, the support of a module is determined by the associated primes. Namely, the associated primes contain the "generic points" (that is, the minimal primes) of the support. In some cases, however, they may contain more.

add: We are currently using the notation $\mathrm{Ann}(x)$ **for the annihilator of** $x \in M$**. This has not been defined before. Should we add this in a previous chapter?**

2.2.4 Definition Let M be a finitely generated R-module. The prime ideal $\mathfrak{p}$ is said to be **associated** to M if there exists an element $x \in M$ such that $\mathfrak{p}$ is the annihilator of x. The set of associated primes is $\mathrm{Ass}(M)$.

Note that the annihilator of an element $x \in M$ is not necessarily prime, but it is possible that the annihilator might be prime, in which case it is associated.

2.2.5 Remark Show that $\mathfrak{p} \in \mathrm{Ass}(M)$ if and only if there is an injection $R/\mathfrak{p} \hookrightarrow M$.

2.2.6 Remark Let $\mathfrak{p} \in \operatorname{Spec} R$. Then $\mathrm{Ass}(R/\mathfrak{p}) = \{\mathfrak{p}\}$.

2.2.7 Example Take $R = k[x, y, z]$, where k is an integral domain, and let $I = (x^2 - yz, x(z - 1))$. Any prime associated to I must contain I, so let's consider $\mathfrak{p} = (x^2 - yz, z - 1) = (x^2 - y, z - 1)$, which is $I : x$. It is prime because $R/\mathfrak{p} = k[x]$, which is a domain. To see that $(I : x) \subset \mathfrak{p}$, assume $tx \in I \subset \mathfrak{p}$; since $x \notin \mathfrak{p}$, $t \in p$, as desired.

There are two more associated primes, but we will not find them here.

We shall start by proving that $\mathrm{Ass}(M) \neq \varnothing$ for nonzero modules.

2.2.8 Proposition *If $M \neq 0$, then M has an associated prime.*

Proof. Consider the collection of ideals in R that arise as the annihilator of a nonzero element in M. Let $I \subset R$ be a maximal element among this collection. The existence of I is guaranteed thanks to the noetherianness of R. Then $I = \mathrm{Ann}(x)$ for some $x \in M$, so $1 \notin I$ because the annihilator of a nonzero element is not the full ring.

I claim that I is prime, and hence $I \in \mathrm{Ass}(M)$. Indeed, suppose $ab \in I$ where $a, b \in R$. This means that

$$(ab)x = 0. \qquad \square$$

Consider the annihilator $\mathrm{Ann}(bx)$ of bx. This contains the annihilator of x, so I; it also contains a.

There are two cases. If $bx = 0$, then $b \in I$ and we are done. Suppose to the contrary $bx \neq 0$. In this case, $\mathrm{Ann}(bx)$ contains $(a) + I$, which contains I. By maximality, it must happen that $\mathrm{Ann}(bx) = I$ and $a \in I$.

In either case, we find that one of a, b belongs to I, so that I is prime.

2.2.9 Example (A module with no associated prime) Without the noetherian hypothesis, 2.2.8 is *false*. Consider $R = \mathbb{C}[x_1, x_2, \dots]$, the polynomial ring over $\mathbb{C}$ in infinitely many variables, and the ideal $I = (x_1, x_2^2, x_3^3, \dots) \subset R$. The claim is that

$$\mathrm{Ass}(R/I) = \varnothing.$$

To see this, suppose a prime $\mathfrak{p}$ was the annihilator of some $\overline{f} \in R/I$. Then $\overline{f}$ lifts to $f \in R$; it follows that $\mathfrak{p}$ is precisely the set of $g \in R$ such that $fg \in I$. Now f contains only finitely many of the variables x_i, say $x_1, \dots, x_n$. It is then clear that $x_{n+1}^{n+1} f \in I$ (so $x_{n+1}^{n+1} \in \mathfrak{p}$), but $x_{n+1} f \notin I$ (so $x_{n+1} \notin \mathfrak{p}$). It follows that $\mathfrak{p}$ is not a prime, a contradiction.

We shall now show that the associated primes are finite in number.

2.2.10 Proposition *If M is finitely generated, then* $\mathrm{Ass}(M)$ *is finite.*

The idea is going to be to use the fact that M is finitely generated to build M out of finitely many pieces, and use that to bound the number of associated primes to each piece. For this, we need:

2.2.11 Lemma *Suppose we have an exact sequence of finitely generated R-modules*

$$0 \to M' \to M \to M'' \to 0.$$

Then

$$\mathrm{Ass}(M') \subset \mathrm{Ass}(M) \subset \mathrm{Ass}(M') \cup \mathrm{Ass}(M'')$$

Proof. The first claim is obvious. If $\mathfrak{p}$ is the annihilator of in $x \in M'$, it is an annihilator of something in M (namely the image of x), because $M' \to M$ is injective. So $\mathrm{Ass}(M') \subset \mathrm{Ass}(M)$.

The harder direction is the other inclusion. Suppose $\mathfrak{p} \in \mathrm{Ass}(M)$. Then there is $x \in M$ such that $\mathfrak{p} = \mathrm{Ann}(x)$. Consider the submodule $Rx \subset M$. If $Rx \cap M' \neq 0$, then we can choose $y \in Rx \cap M' - \{0\}$. I claim that $\mathrm{Ann}(y) = \mathfrak{p}$ and so $\mathfrak{p} \in \mathrm{Ass}(M')$. To see this, $y = ax$ for some $a \in R$. The annihilator of y is the set of elements $b \in R$ such that

$$abx = 0$$

or, equivalently, the set of $b \in R$ such that $ab \in \mathfrak{p} = \mathrm{Ann}(x)$. But $y = ax \neq 0$, so $a \notin \mathfrak{p}$. As a result, the condition $b \in \mathrm{Ann}(y)$ is the same as $b \in \mathfrak{p}$. In other words,

$$\mathrm{Ann}(y) = \mathfrak{p}$$

which proves the claim.

Suppose now that $Rx \cap M' = 0$. Let $\phi : M \twoheadrightarrow M''$ be the surjection. I claim that $\mathfrak{p} = \mathrm{Ann}(\phi(x))$ and consequently that $\mathfrak{p} \in \mathrm{Ass}(M'')$. The proof is as follows. Clearly $\mathfrak{p}$ annihilates $\phi(x)$ as it annihilates x. Suppose $a \in \mathrm{Ann}(\phi(x))$. This means that $\phi(ax) = 0$, so $ax \in \ker \phi = M'$; but $\ker \phi \cap Rx = 0$. So $ax = 0$ and consequently $a \in \mathfrak{p}$. It follows $\mathrm{Ann}(\phi(x)) = \mathfrak{p}$. $\square$

The next step in the proof of 2.2.10 is that any finitely generated module admits a filtration each of whose quotients are of a particularly nice form. This result is quite useful and will be referred to in the future.

2.2.12 Proposition (Dévissage) *For any finitely generated R-module M, there exists a finite filtration*

$$0 = M_0 \subset M_1 \subset \cdots \subset M_k = M$$

such that the successive quotients M_{i+1}/M_i are isomorphic to various $R/\mathfrak{p}_i$ with the $\mathfrak{p}_i \subset R$ prime.

Proof. Let $M' \subset M$ be maximal among submodules for which such a filtration (ending with M') exists. We would like to show that $M' = M$. Now M' is well-defined since 0 has such a filtration and M is noetherian.

There is a filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_l = M' \subset M$$

where the successive quotients, *except* possibly the last M/M', are of the form $R/\mathfrak{p}_i$ for $\mathfrak{p}_i \in \operatorname{Spec} R$. If $M' = M$, we are done. Otherwise, consider the quotient $M/M' \neq 0$. There is an associated prime of M/M'. So there is a prime $\mathfrak{p}$ which is the annihilator of $x \in M/M'$. This means that there is an injection

$$R/\mathfrak{p} \hookrightarrow M/M'.$$

Now, take M_{l+1} as the inverse image in M of $R/\mathfrak{p} \subset M/M'$. Then, we can consider the finite filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_{l+1},$$

all of whose successive quotients are of the form $R/\mathfrak{p}_i$; this is because $M_{l+1}/M_l = M_{l+1}/M'$ is of this form by construction. We have thus extended this filtration one step further, a contradiction since M' was assumed to be maximal. □

Now we are in a position to meet the goal, and prove that $\operatorname{Ass}(M)$ is always a finite set.

Proof of 2.2.10. Suppose M is finitely generated Take our filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_k = M.$$

By induction, we show that $\operatorname{Ass}(M_i)$ is finite for each i. It is obviously true for $i = 0$. Assume now that $\operatorname{Ass}(M_i)$ is finite; we prove the same for $\operatorname{Ass}(M_{i+1})$. We have an exact sequence

$$0 \to M_i \to M_{i+1} \to R/\mathfrak{p}_i \to 0$$

which implies that, by 2.2.11,

$$\operatorname{Ass}(M_{i+1}) \subset \operatorname{Ass}(M_i) \cup \operatorname{Ass}(R/\mathfrak{p}_i) = \operatorname{Ass}(M_i) \cup \{\mathfrak{p}_i\},$$

so $\operatorname{Ass}(M_{i+1})$ is also finite. By induction, it is now clear that $\operatorname{Ass}(M_i)$ is finite for every i.

This proves the proposition; it also shows that the number of associated primes is at most the length of the filtration. □

Finally, we characterize the zero divisors on M in terms of the associated primes. The last characterization of the result will be useful in the future. It implies, for instance, that if R is local and $\mathfrak{m}$ the maximal ideal, then if every element of $\mathfrak{m}$ is a zero divisor on a finitely generated module M, then $\mathfrak{m} \in \mathrm{Ass}(M)$.

2.2.13 Proposition *If M is a finitely generated module over a noetherian ring R, then the zero divisors on M are the union $\bigcup_{\mathfrak{p} \in \mathrm{Ass}(M)} \mathfrak{p}$.*

More strongly, if $I \subset R$ is any ideal consisting of zero divisors on M, then I is contained in an associated prime.

Proof. Any associated prime is an annihilator of some element of M, so it consists of zero divisors. Conversely, if $a \in R$ annihilates $x \in M$, then a belongs to every associated prime of the nonzero module $Ra \subset M$. (There is at least one by proposition 2.2.10.)

For the last statement, we use prime avoidance (theorem 2.6.20): if I consists of zero divisors, then I is contained in the union $\bigcup_{\mathfrak{p} \in \mathrm{Ass}(M)} \mathfrak{p}$ by the first part of the proof. This is a finite union by **??**, so prime avoidance implies I is contained one of these primes. $\square$

2.2.14 Remark For every module M over any (not necessarily noetherian) ring R, the set of M-zero divisors $\mathcal{Z}(M)$ is a union of prime ideals. In general, there is an easy characterization of sets Z which are a union of primes: it is exactly when $R \setminus Z$ is a *saturated multiplicative set.* This is Kaplansky's Theorem 2.

2.2.15 Definition A multiplicative set $S \neq \varnothing$ is a *saturated multiplicative set* if for all $a, b \in R$, $a, b \in S$ if and only if $ab \in S$. ("multiplicative set" just means the "if" part)

To see that $\mathcal{Z}(M)$ is a union of primes, just verify that its complement is a saturated multiplicative set.

Localization and $\mathrm{Ass}(M)$

It turns out to be extremely convenient that the construction $M \to \mathrm{Ass}(M)$ behaves about as nicely with respect to localization as we could possibly want. This lets us, in fact, reduce arguments to the case of a local ring, which is a significant simplification.

So, as usual, let R be noetherian, and M a finitely generated R-module. Let further $S \subset R$ be a multiplicative subset. Then $S^{-1}M$ is a finitely generated module over the noetherian ring $S^{-1}M$. So it makes sense to consider both $\mathrm{Ass}(M) \subset \operatorname{Spec} R$ and $\mathrm{Ass}(S^{-1}M) \subset \operatorname{Spec} S^{-1}R$. But we also know that $\operatorname{Spec} S^{-1}R \subset \operatorname{Spec} R$ is just the set of primes of R that do not intersect S. Thus, we can directly compare $\mathrm{Ass}(M)$ and $\mathrm{Ass}(S^{-1}M)$, and one might conjecture (correctly, as it happens) that $\mathrm{Ass}(S^{-1}M) = \mathrm{Ass}(M) \cap \operatorname{Spec} S^{-1}R$.

2.2.16 Proposition *Let R noetherian, M finitely generated and $S \subset R$ multiplicatively closed. Then*

$$\mathrm{Ass}(S^{-1}M) = \left\{ S^{-1}\mathfrak{p} : \mathfrak{p} \in \mathrm{Ass}(M), \mathfrak{p} \cap S = \varnothing \right\}.$$

Proof. We first prove the easy direction, namely that $\mathrm{Ass}(S^{-1}M)$ *contains* primes in $\operatorname{Spec} S^{-1}R \cap \mathrm{Ass}(M)$.

Suppose $\mathfrak{p} \in \mathrm{Ass}(M)$ and $\mathfrak{p} \cap S = \varnothing$. Then $\mathfrak{p} = \mathrm{Ann}(x)$ for some $x \in M$. Then the annihilator of $x/1 \in S^{-1}M$ is just $S^{-1}\mathfrak{p}$, as one can directly check. Thus $S^{-1}\mathfrak{p} \in \mathrm{Ass}(S^{-1}M)$. So we get the easy inclusion.

Let us now do the harder inclusion. Call the localization map $R \to S^{-1}R$ as ϕ. Let $\mathfrak{q} \in \mathrm{Ass}(S^{-1}M)$. By definition, this means that $\mathfrak{q} = \mathrm{Ann}(x/s)$ for some $x \in M$, $s \in S$. We want to see that $\phi^{-1}(\mathfrak{q}) \in \mathrm{Ass}(M) \subset \operatorname{Spec} R$. By definition $\phi^{-1}(\mathfrak{q})$ is the set of elements $a \in R$ such that

$$\frac{ax}{s} = 0 \in S^{-1}M.$$

In other words, by definition of the localization, this is

$$\phi^{-1}(\mathfrak{q}) = \bigcup_{t \in S} \{a \in R : atx = 0 \in M\} = \bigcup \mathrm{Ann}(tx) \subset R.$$

We know, however, that among elements of the form $\mathrm{Ann}(tx)$, there is a *maximal* element $I = \mathrm{Ann}(t_0 x)$ for some $t_0 \in S$, since R is noetherian. The claim is that $I = \phi^{-1}(\mathfrak{q})$, so $\phi^{-1}(\mathfrak{q}) \in \mathrm{Ass}(M)$.

Indeed, any other annihilator $I' = \mathrm{Ann}(tx)$ (for $t \in S$) must be contained in $\mathrm{Ann}(t_0 t x)$. However, $I \subset \mathrm{Ann}(t_0 x)$ and I is maximal, so $I = \mathrm{Ann}(t_0 t x)$ and $I' \subset I$. In other words, I contains all the other annihilators $\mathrm{Ann}(tx)$ for $t \in S$. In particular, the big union above, i.e. $\phi^{-1}(\mathfrak{q})$, is just $I = \mathrm{Ann}(t_0 x)$. In particular, $\phi^{-1}(\mathfrak{q}) = \mathrm{Ann}(t_0 x)$ is in $\mathrm{Ass}(M)$. This means that every associated prime of $S^{-1}M$ comes from an associated prime of M, which completes the proof. ∎

2.2.17 Remark Show that, if M is a finitely generated module over a noetherian ring, that the map

$$M \to \bigoplus_{\mathfrak{p} \in \mathrm{Ass}(M)} M_{\mathfrak{p}}$$

is injective. Is this true if M is not finitely generated?

Associated primes determine the support

The next claim is that the support and the associated primes are related.

2.2.18 Proposition *The support is the closure of the associated primes:*

$$\operatorname{supp} M = \bigcup_{\mathfrak{q} \in \mathrm{Ass}(M)} \overline{\{\mathfrak{q}\}}$$

By definition of the Zariski topology, this means that a prime $\mathfrak{p} \in \operatorname{Spec} R$ belongs to $\operatorname{supp} M$ if and only if it contains an associated prime.

Proof. First, we show that $\mathrm{supp}(M)$ contains the set of primes $\mathfrak{p} \in \operatorname{Spec} R$ containing an associated prime; this will imply that $\mathrm{supp}(M) \supset \bigcup_{\mathfrak{q} \in \mathrm{Ass}(M)} \overline{\{\mathfrak{q}\}}$. So let $\mathfrak{q}$ be an associated prime and $\mathfrak{p} \supset \mathfrak{q}$. We need to show that

$$\mathfrak{p} \in \mathrm{supp}\, M, \text{ i.e. } M_{\mathfrak{p}} \neq 0.$$

But, since $\mathfrak{q} \in \mathrm{Ass}(M)$, there is an injective map

$$R/\mathfrak{q} \hookrightarrow M,$$

so localization gives an injective map

$$(R/\mathfrak{q})_{\mathfrak{p}} \hookrightarrow M_{\mathfrak{p}}.$$

Here, however, the first object $(R/\mathfrak{q})_{\mathfrak{p}}$ is nonzero since nothing nonzero in $R/\mathfrak{q}$ can be annihilated by something outside $\mathfrak{p}$. So $M_{\mathfrak{p}} \neq 0$, and $\mathfrak{p} \in \mathrm{supp}\, M$.

Let us now prove the converse inclusion. Suppose that $\mathfrak{p} \in \mathrm{supp}\, M$. We have to show that $\mathfrak{p}$ contains an associated prime. By assumption, $M_{\mathfrak{p}} \neq 0$, and $M_{\mathfrak{p}}$ is a finitely generated module over the noetherian ring $R_{\mathfrak{p}}$. So $M_{\mathfrak{p}}$ has an associated prime. It follows by 2.2.16 that $\mathrm{Ass}(M) \cap \operatorname{Spec} R_{\mathfrak{p}}$ is nonempty. Since the primes of $R_{\mathfrak{p}}$ correspond to the primes contained in $\mathfrak{p}$, it follows that there is a prime contained in $\mathfrak{p}$ that lies in $\mathrm{Ass}(M)$. This is precisely what we wanted to prove. □

2.2.19 Corollary *For M finitely generated,* $\mathrm{supp}\, M$ *is closed. Further, every minimal element of* $\mathrm{supp}\, M$ *lies in* $\mathrm{Ass}(M)$.

Proof. Indeed, the above result says that

$$\mathrm{supp}\, M = \bigcup_{\mathfrak{q} \in \mathrm{Ass}(M)} \overline{\{\mathfrak{q}\}}.$$

Since $\mathrm{Ass}(M)$ is finite, it follows that $\mathrm{supp}\, M$ is closed. The above equality also shows that any minimal element of $\mathrm{supp}\, M$ must be an associated prime. □

2.2.20 Example 2.2.19 is *false* for modules that are not finitely generated. Consider for instance the abelian group $\bigoplus_p \mathbb{Z}/p$. The support of this as a $\mathbb{Z}$-module is precisely the set of all closed points (i.e., maximal ideals) of $\operatorname{Spec} \mathbb{Z}$, and is consequently is not closed.

2.2.21 Corollary *The ring R has finitely many minimal prime ideals.*

Proof. Clearly, $\mathrm{supp}\, R = \operatorname{Spec} R$. Thus every prime ideal of R contains an associated prime of R by 2.2.18. □

So $\operatorname{Spec} R$ is the finite union of the irreducible closed pieces $\overline{\mathfrak{q}}$ if R is noetherian. **add: I am not sure if "irreducibility" has already been defined. Check this.**

We have just seen that $\mathrm{supp}\, M$ is a closed subset of $\operatorname{Spec} R$ and is a union of finitely many irreducible subsets. More precisely,

$$\mathrm{supp}\, M = \bigcup_{\mathfrak{q} \in \mathrm{Ass}(M)} \overline{\{\mathfrak{q}\}}$$

though there might be some redundancy in this expression. Some associated prime might be contained in others.

2.2.22 Definition A prime $\mathfrak{p} \in \mathrm{Ass}(M)$ is an **isolated** associated prime of M if it is minimal (with respect to the ordering on $\mathrm{Ass}(M)$); it is **embedded** otherwise.

So the embedded primes are not needed to describe the support of M.

add: Examples of embedded primes

2.2.23 Remark It follows that in a noetherian ring, every minimal prime consists of zero divisors. Although we shall not use this in the future, the same is true in non-noetherian rings as well. Here is an argument.

Let R be a ring and $\mathfrak{p} \subset R$ a minimal prime. Then $R_{\mathfrak{p}}$ has precisely one prime ideal. We now use:

2.2.24 Lemma *If a ring R has precisely one prime ideal $\mathfrak{p}$, then any $x \in \mathfrak{p}$ is nilpotent.*

Proof. Indeed, it suffices to see that $R_x = 0$ (1.3.7 in **??**) if $x \in \mathfrak{p}$. But $\operatorname{Spec} R_x$ consists of the primes of R not containing x. However, there are no such primes. Thus $\operatorname{Spec} R_x = \emptyset$, so $R_x = 0$. □

It follows that every element in $\mathfrak{p}$ is a zero divisor in $R_{\mathfrak{p}}$. As a result, if $x \in \mathfrak{p}$, there is $\frac{s}{t} \in R_{\mathfrak{p}}$ such that $xs/t = 0$ but $\frac{s}{t} \neq 0$. In particular, there is $t' \notin \mathfrak{p}$ with

$$xst' = 0, \quad st' \neq 0,$$

so that x is a zero divisor.

Primary modules

A primary modules are ones that has only one associated prime. It is equivalent to say that any homothety is either injective or nilpotent. As we will see in the next section, any module has a "primary decomposition:" in fact, it embeds as a submodule of a sum of primary modules.

2.2.25 Definition Let $\mathfrak{p} \subset R$ be prime, M a finitely generated R-module. Then M is **$\mathfrak{p}$-primary** if

$$\mathrm{Ass}(M) = \{\mathfrak{p}\}.$$

A module is **primary** if it is $\mathfrak{p}$-primary for some prime $\mathfrak{p}$, i.e., has precisely one associated prime.

2.2.26 Proposition *Let M be a finitely generated R-module. Then M is $\mathfrak{p}$-primary if and only if, for every $m \in M - \{0\}$, the annihilator $\mathrm{Ann}(m)$ has radical $\mathfrak{p}$.*

Proof. We first need a small observation.

2.2.27 Lemma *If M is $\mathfrak{p}$-primary, then any nonzero submodule $M' \subset M$ is $\mathfrak{p}$-primary.*

Proof. Indeed, we know that $\mathrm{Ass}(M') \subset \mathrm{Ass}(M)$ by 2.2.11. Since $M' \neq 0$, we also know that M' has an associated prime (2.2.8). Thus $\mathrm{Ass}(M') = \{\mathfrak{p}\}$, so M' is $\mathfrak{p}$-primary. □

Let us now return to the proof of the main result, 2.2.26. Assume first that M is $\mathfrak{p}$-primary. Let $x \in M$, $x \neq 0$. Let $I = \mathrm{Ann}(x)$; we are to show that $\mathrm{Rad}(I) = \mathfrak{p}$. By definition, there is an injection

$$R/I \hookrightarrow M$$

sending $1 \to x$. As a result, R/I is $\mathfrak{p}$-primary by the above lemma. We want to know that $\mathfrak{p} = \mathrm{Rad}(I)$. We saw that the support $\operatorname{supp} R/I = \{\mathfrak{q} : \mathfrak{q} \supset I\}$ is the union of the closures of the associated primes. In this case,

$$\operatorname{supp}(R/I) = \{\mathfrak{q} : \mathfrak{q} \supset \mathfrak{p}\}.$$

But we know that $\mathrm{Rad}(I) = \bigcap_{\mathfrak{q} \supset I} \mathfrak{q}$, which by the above is just $\mathfrak{p}$. This proves that $\mathrm{Rad}(I) = \mathfrak{p}$. We have shown that if R/I is primary, then I has radical $\mathfrak{p}$.

The converse is easy. Suppose the condition holds and $\mathfrak{q} \in \mathrm{Ass}(M)$, so $\mathfrak{q} = \mathrm{Ann}(x)$ for $x \neq 0$. But then $\mathrm{Rad}(\mathfrak{q}) = \mathfrak{p}$, so

$$\mathfrak{q} = \mathfrak{p}$$

and $\mathrm{Ass}(M) = \{\mathfrak{p}\}$. ∎

We have another characterization.

2.2.28 Proposition *Let $M \neq 0$ be a finitely generated R-module. Then M is primary if and only if for each $a \in R$, then the homothety $M \xrightarrow{a} M$ is either injective or nilpotent.*

Proof. Suppose first that M is $\mathfrak{p}$-primary. Then multiplication by anything in $\mathfrak{p}$ is nilpotent because the annihilator of everything nonzero has radical $\mathfrak{p}$ by 2.2.26. But if $a \notin \mathfrak{p}$, then $\mathrm{Ann}(x)$ for $x \in M - \{0\}$ has radical $\mathfrak{p}$ and cannot contain a.

Let us now do the other direction. Assume that every element of a acts either injectively or nilpotently on M. Let $I \subset R$ be the collection of elements $a \in R$ such that $a^n M = 0$ for n large. Then I is an ideal, since it is closed under addition by the binomial formula: if $a, b \in I$ and a^n, b^n act by zero, then $(a+b)^{2n}$ acts by zero as well.

I claim that I is actually prime. If $a, b \notin I$, then a, b act by multiplication injectively on M. So $a : M \to M, b : M \to M$ are injective. However, a composition of injections is injective, so ab acts injectively and $ab \notin I$. So I is prime.

We need now to check that if $x \in M$ is nonzero, then $\mathrm{Ann}(x)$ has radical I. Indeed, if $a \in R$ annihilates x, then the homothety $M \xrightarrow{a} M$ cannot be injective, so it must be nilpotent (i.e. in I). Conversely, if $a \in I$, then a power of a is nilpotent, so a power of a must kill x. It follows that $\mathrm{Ann}(x) = I$. Now, by 2.2.26, we see that M is I-primary. ∎

We now have this notion of a primary module. The idea is that all the torsion is somehow concentrated in some prime.

2.2.29 Example If R is a noetherian ring and $\mathfrak{p} \in \operatorname{Spec} R$, then $R/\mathfrak{p}$ is $\mathfrak{p}$-primary. More generally, if $I \subset R$ is an ideal, then R/I is ideal if and only if $\mathrm{Rad}(I)$ is prime. This follows from 2.2.28.

2.2.30 Remark If $0 \to M' \to M \to M'' \to 0$ is an exact sequence with M', M, M'' nonzero and finitely generated, then M is $\mathfrak{p}$-primary if and only if M', M'' are.

2.2.31 Remark Let M be a finitely generated R-module. Let $\mathfrak{p} \in \operatorname{Spec} R$. Show that the sum of two $\mathfrak{p}$-primary submodules is $\mathfrak{p}$-primary. Deduce that there is a $\mathfrak{p}$-primary submodule of M which contains every $\mathfrak{p}$-primary submodule.

2.2.32 Remark (Bourbaki) Let M be a finitely generated R-module. Let $T \subset \mathrm{Ass}(M)$ be a subset of the associated primes. Prove that there is a submodule $N \subset M$ such that

$$\mathrm{Ass}(N) = T, \quad \mathrm{Ass}(M/N) = \mathrm{Ass}(M) - T.$$

2.3. Primary decomposition

This is the structure theorem for modules over a noetherian ring, in some sense. Throuoghout, we fix a noetherian ring R.

Irreducible and coprimary modules

2.3.1 Definition Let M be a finitely generated R-module. A submodule $N \subset M$ is **$\mathfrak{p}$-coprimary** if M/N is $\mathfrak{p}$-primary.

Similarly, we can say that $N \subset M$ is **coprimary** if it is $\mathfrak{p}$-coprimary for some $\mathfrak{p} \in \operatorname{Spec} R$.

We shall now show we can represent any submodule of M as an intersection of coprimary submodules. In order to do this, we will define a submodule of M to be *irreducible* if it cannot be written as a nontrivial intersection of submodules of M. It will follow by general nonsense that any submodule is an intersection of irreducible submodueles. We will then see that any irreducible submodule is coprimary.

2.3.2 Definition The submomdule $N \subsetneq M$ is **irreducible** if whenever $N = N_1 \cap N_2$ for $N_1, N_2 \subset M$ submodules, then either one of N_1, N_2 equals N. In other words, it is not the intersection of larger submodules.

2.3.3 Proposition *An irreducible submodule $N \subset M$ is coprimary.*

Proof. Say $a \in R$. We would like to show that the homothety

$$M/N \xrightarrow{a} M/N$$

is either injective or nilpotent. Consider the following submodules of M/N:

$$K(n) = \{x \in M/N : a^n x = 0\}.$$

Then clearly $K(0) \subset K(1) \subset \ldots$; this chain stabilizes as the quotient module is noetherian. In particular, $K(n) = K(2n)$ for large n.

It follows that if $x \in M/N$ is divisible by a^n (n large) and nonzero, then $a^n x$ is also nonzero. Indeed, say $x = a^n y \neq 0$; then $y \notin K(n)$, so $a^n x = a^{2n} y \neq 0$ or we would have $y \in K(2n) = K(n)$. In M/N, the submodules

$$a^n(M/N) \cap \ker(a^n)$$

are equal to zero for large n. But our assumption was that N is irreducible. So one of these submodules of M/N is zero. That is, either $a^n(M/N) = 0$ or $\ker a^n = 0$. We get either injectivity or nilpotence on M/N. This proves the result. ☐

Irreducible and primary decompositions

We shall now show that in a finitely generated module over a noetherian ring, we can write 0 as an intersection of coprimary modules. This decomposition, which is called a *primary decomposition*, will be deduced from purely general reasoning.

2.3.4 Definition An **irreducible decomposition** of the module M is a representation $N_1 \cap N_2 \cdots \cap N_k = 0$, where the $N_i \subset M$ are irreducible submodules.

2.3.5 Proposition *If M is finitely generated, then M has an irreducible decomposition. There exist finitely many irreducible submodules $N_1, \ldots, N_k$ with*

$$N_1 \cap \cdots \cap N_k = 0.$$

In other words,

$$M \to \bigoplus M/N_i$$

is injective. So a finitely generated module over a noetherian ring can be imbedded in a direct sum of primary modules, since by 2.3.3 the M/N_i are primary.

Proof. This is now purely formal.

Among the submodules of M, some may be expressible as intersections of finitely many irreducibles, while some may not be. Our goal is to show that 0 is such an intersection. Let $M' \subset M$ be a maximal submodule of M such that M' *cannot* be written as such an intersection. If no such M' exists, then we are done, because then 0 can be written as an intersection of finitely many irreducible submodules.

Now M' is not irreducible, or it would be the intersection of one irreducible submodule. It follows M' can be written as $M' = M_1' \cap M_2'$ for two strictly larger submodules of M. But by maximality, M_1', M_2' admit decompositions as intersections of irreducibles. So M' admits such a decomposition as well, a contradiction. ☐

2.3.6 Corollary *For any finitely generated M, there exist coprimary submodules $N_1, \ldots, N_k \subset M$ such that $N_1 \cap \cdots \cap N_k = 0$.*

Proof. Indeed, every irreducible submodule is coprimary. ☐

For any M, we have an **irreducible decomposition**

$$0 = \bigcap N_i$$

for the N_i a finite set of irreducible (and thus coprimary) submodules. This decomposition here is highly non-unique and non-canonical. Let's try to pare it down to something which is a lot more canonical.

The first claim is that we can collect together modules which are coprimary for some prime.

2.3.7 Lemma *Let $N_1, N_2 \subset M$ be $\mathfrak{p}$-coprimary submodules. Then $N_1 \cap N_2$ is also $\mathfrak{p}$-coprimary.*

Proof. We have to show that $M/N_1 \cap N_2$ is $\mathfrak{p}$-primary. Indeed, we have an injection

$$M/N_1 \cap N_2 \rightarrowtail M/N_1 \oplus M/N_2$$

which implies that $\mathrm{Ass}(M/N_1 \cap N_2) \subset \mathrm{Ass}(M/N_1) \cup \mathrm{Ass}(M/N_2) = \{\mathfrak{p}\}$. So we are done. $\square$

In particular, if we do not want irreducibility but only primariness in the decomposition

$$0 = \bigcap N_i,$$

we can assume that each N_i is $\mathfrak{p}_i$ coprimary for some prime $\mathfrak{p}_i$ with the $\mathfrak{p}_i$ *distinct*.

2.3.8 Definition Such a decomposition of zero, where the different modules N_i are $\mathfrak{p}_i$-coprimary for different $\mathfrak{p}_i$, is called a **primary decomposition**.

Uniqueness questions

In general, primary decomposition is *not* unique. Nonetheless, we shall see that a limited amount of uniqueness does hold. For instance, the primes that occur are determined.

Let M be a finitely generated module over a noetherian ring R, and suppose $N_1 \cap \cdots \cap N_k = 0$ is a primary decomposition. Let us assume that the decomposition is *minimal*: that is, if we dropped one of the N_i, the intersection would no longer be zero. This implies that

$$N_i \not\supset \bigcap_{j \neq i} N_j$$

or we could omit one of the N_i. Then the decomposition is called a **reduced primary decomposition**.

Again, what this tells us is that $M \rightarrowtail \bigoplus M/N_i$. What we have shown is that M can be imbedded in a sum of pieces, each of which is $\mathfrak{p}$-primary for some prime, and the different primes are distinct.

This is **not** unique. However,

2.3.9 Proposition *The primes $\mathfrak{p}_i$ that appear in a reduced primary decomposition of zero are uniquely determined. They are the associated primes of M.*

Proof. All the associated primes of M have to be there, because we have the injection

$$M \rightarrowtail \bigoplus M/N_i$$

so the associated primes of M are among those of M/N_i (i.e. the $\mathfrak{p}_i$).

The hard direction is to see that each $\mathfrak{p}_i$ is an associated prime. I.e. if M/N_i is $\mathfrak{p}_i$-primary, then $\mathfrak{p}_i \in \mathrm{Ass}(M)$; we don't need to use primary modules except for primes in the associated primes. Here we need to use the fact that our decomposition has no redundancy. Without loss of generality, it suffices to show that $\mathfrak{p}_1$, for instance, belongs to $\mathrm{Ass}(M)$. We will use the fact that

$$N_1 \not\supset N_2 \cap \dots.$$

So this tells us that $N_2 \cap N_3 \cap \dots$ is not equal to zero, or we would have a containment. We have a map

$$N_2 \cap \dots \cap N_k \to M/N_1;$$

it is injective, since the kernel is $N_1 \cap N_2 \cap \dots \cap N_k = 0$ as this is a decomposition. However, M/N_1 is $\mathfrak{p}_1$-primary, so $N_2 \cap \dots \cap N_k$ is $\mathfrak{p}_1$-primary. In particular, $\mathfrak{p}_1$ is an associated prime of $N_2 \cap \dots \cap N_k$, hence of M. □

The primes are determined. The factors are not. However, in some cases they are.

2.3.10 Proposition *Let $\mathfrak{p}_i$ be a minimal associated prime of M, i.e. not containing any smaller associated prime. Then the submodule N_i corresponding to $\mathfrak{p}_i$ in the reduced primary decomposition is uniquely determined: it is the kernel of*

$$M \to M_{\mathfrak{p}_i}.$$

Proof. We have that $\bigcap N_j = \{0\} \subset M$. When we localize at $\mathfrak{p}_i$, we find that

$$(\bigcap N_j)_{\mathfrak{p}_i} = \bigcap (N_j)_{\mathfrak{p}_i} = 0$$

as localization is an exact functor. If $j \neq i$, then M/N_j is $\mathfrak{p}_j$ primary, and has only $\mathfrak{p}_j$ as an associated prime. It follows that $(M/N_j)_{\mathfrak{p}_i}$ has no associated primes, since the only associated prime could be $\mathfrak{p}_j$, and that's not contained in $\mathfrak{p}_j$. In particular, $(N_j)_{\mathfrak{p}_i} = M_{\mathfrak{p}_i}$.

Thus, when we localize the primary decomposition at $\mathfrak{p}_i$, we get a trivial primary decomposition: most of the factors are the full $M_{\mathfrak{p}_i}$. It follows that $(N_i)_{\mathfrak{p}_i} = 0$. When we draw a commutative diagram

$$\begin{tikzcd} N_i \arrow[r] \arrow[d] & (N_i)_{\mathfrak{p}_i} = 0 \arrow[d] \\ M \arrow[r] & M_{\mathfrak{p}_i}. \end{tikzcd}$$

we find that N_i goes to zero in the localization.

Now if $x \in \ker(M \to M_{\mathfrak{p}_i}$, then $sx = 0$ for some $s \notin \mathfrak{p}_i$. When we take the map $M \to M/N_i$, sx maps to zero; but s acts injectively on M/N_i, so x maps to zero in M/N_i, i.e. is zero in N_i. □

This has been abstract, so:

2.3.11 Example Let $R = \mathbb{Z}$. Let $M = \mathbb{Z} \oplus \mathbb{Z}/p$. Then zero can be written as

$$\mathbb{Z} \cap \mathbb{Z}/p$$

as submodules of M. But $\mathbb{Z}$ is $\mathfrak{p}$-coprimary, while $\mathbb{Z}/p$ is (0)-coprimary.

This is not unique. We could have considered

$$\{(n, n), n \in \mathbb{Z}\} \subset M.$$

However, the zero-coprimary part has to be the p-torsion. This is because (0) is the minimal ideal.

The decomposition is always unique, in general, if we have no inclusions among the prime ideals. For $\mathbb{Z}$-modules, this means that primary decomposition is unique for torsion modules. Any torsion group is a direct sum of the p-power torsion over all primes p.

2.3.12 Remark Suppose R is a noetherian ring and $R_{\mathfrak{p}}$ is a domain for each prime ideal $\mathfrak{p} \subset R$. Then R is a finite direct product $\prod R_i$ for each R_i a domain.

To see this, consider the minimal primes $\mathfrak{p}_i \in \operatorname{Spec} R$. There are finitely many of them, and argue that since every localization is a domain, $\operatorname{Spec} R$ is disconnected into the pieces $V(\mathfrak{p}_i)$. It follows that there is a decomposition $R = \prod R_i$ where $\operatorname{Spec} R_i$ has $\mathfrak{p}_i$ as the unique minimal prime. Each R_i satisfies the same condition as R, so we may reduce to the case of R having a unique minimal prime ideal. In this case, however, R is reduced, so its unique minimal prime ideal must be zero.

2.4. Artinian rings and modules

The notion of an *artinian ring* appears to be dual to that of a noetherian ring, since the chain condition is simply reversed in the definition. However, the artinian condition is much stronger than the noetherian one. In fact, artinianness actually implies noetherianness, and much more. Artinian modules over non-artinian rings are frequently of interest as well; for instance, if R is a noetherian ring and $\mathfrak{m}$ is a maximal ideal, then for any finitely generated R-module M, the module $M/\mathfrak{m}M$ is artinian.

Definitions

2.4.1 Definition A commutative ring R is **Artinian** every descending chain of ideals $I_0 \supset I_1 \supset I_2 \supset \dots$ stabilizes.

2.4.2 Definition The same definition makes sense for modules. We can define an R-module M to be **Artinian** if every descending chain of submodules stabilizes.

In fact, as we shall see when we study dimension theory, we actually often do want to study artinian modules over non-artinian rings, so this definition is useful.

2.4.3 Remark A module is artinian if and only if every nonempty collection of submodules has a minimal element.

2.4.4 Remark A ring which is a finite-dimensional algebra over a field is artinian.

2.4.5 Proposition *If $0 \to M' \to M \to M'' \to 0$ is an exact sequence, then M is Artinian if and only if M', M'' are.*

This is proved in the same way as for noetherianness.

2.4.6 Corollary *Let R be artinian. Then every finitely generated R-module is artinian.*

Proof. Standard. ☐

The main result

This definition is obviously dual to the notion of noetherianness, but it is much more restrictive. The main result is:

2.4.7 Theorem *A commutative ring R is artinian if and only if:*

1. *R is noetherian.*
2. *Every prime ideal of R is maximal.*[1]

So artinian rings are very simple—small in some sense. They all look kind of like fields.

We shall prove this result in a series of small pieces. We begin with a piece of the forward implication in 2.4.7:

2.4.8 Lemma *Let R be artinian. Every prime $\mathfrak{p} \subset R$ is maximal.*

Proof. Indeed, if $\mathfrak{p} \subset R$ is a prime ideal, $R/\mathfrak{p}$ is artinian, as it is a quotient of an artinian ring. We want to show that $R/\mathfrak{p}$ is a field, which is the same thing as saying that $\mathfrak{p}$ is maximal. (In particular, we are essentially proving that an artinian *domain* is a field.)

Let $x \in R/\mathfrak{p}$ be nonzero. We have a descending chain

$$R/\mathfrak{p} \supset (x) \supset (x^2) \dots$$

which necessarily stabilizes. Then we have $(x^n) = (x^{n+1})$ for some n. In particular, we have $x^n = yx^{n+1}$ for some $y \in R/\mathfrak{p}$. But x is a non-zero-divisor, and we find $1 = xy$. So x is invertible. Thus $R/\mathfrak{p}$ is a field. ☐

Next, we claim there are only a few primes in an artinian ring:

2.4.9 Lemma *If R is artinian, there are only finitely many maximal ideals.*

[1] This is much different from the Dedekind ring condition—there, zero is not maximal. An artinian domain is necessarily a field, in fact.

Proof. Assume otherwise. Then we have an infinite sequence

$$\mathfrak{m}_1, \mathfrak{m}_2, \ldots$$

of distinct maximal ideals. Then we have the descending chain

$$R \supset \mathfrak{m}_1 \supset \mathfrak{m}_1 \cap \mathfrak{m}_2 \supset \ldots.$$

This necessarily stabilizes. So for some n, we have that $\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \subset \mathfrak{m}_{n+1}$. However, this means that $\mathfrak{m}_{n+1}$ contains one of the $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ since these are prime ideals (a familiar argument). Maximality and distinctness of the $\mathfrak{m}_i$ give a contradiction. □

In particular, we see that $\operatorname{Spec} R$ for an artinian ring is just a finite set. In fact, since each point is closed, as each prime is maximal, the set has the *discrete topology.* As a result, $\operatorname{Spec} R$ for an artinian ring is *Hausdorff.* (There are very few other cases.)

This means that R factors as a product of rings. Whenever $\operatorname{Spec} R$ can be written as a disjoint union of components, there is a factoring of R into a product $\prod R_i$. So $R = \prod R_i$ where each R_i has only one maximal ideal. Each R_i, as a homomorphic image of R, is artinian. We find, as a result,

add: mention that disconnections of $\operatorname{Spec} R$ are the same thing as idempotents.

2.4.10 Proposition *Any artinian ring is a finite product of local artinian rings.*

Now, let us continue our analysis. We may as well assume that we are working with *local* artinian rings R in the future. In particular, R has a unique prime $\mathfrak{m}$, which must be the radical of R as the radical is the intersection of all primes.

We shall now see that the unique prime ideal $\mathfrak{m} \subset R$ is nilpotent by:

2.4.11 Lemma *If R is artinian (not necessarily local), then* $\operatorname{Rad}(R)$ *is nilpotent.*

It is, of course, always true that any *element* of the radical $\operatorname{Rad}(R)$ is nilpotent, but it is not true for a general ring R that $\operatorname{Rad}(R)$ is nilpotent as an *ideal.*

Proof. Call $J = \operatorname{Rad}(R)$. Consider the decreasing filtration

$$R \supset J \supset J^2 \supset J^3 \supset \ldots.$$

We want to show that this stabilizes at zero. A priori, we know that it stabilizes *somewhere.* For some n, we have

$$J^n = J^{n'}, \quad n' \geqslant n.$$

Call the eventual stabilization of these ideals I. Consider ideals I' such that

$$II' \neq 0.$$

There are now two cases:

1. No such I' exists. Then $I = 0$, and we are done: the powers of J^n stabilize at zero.

2. Otherwise, there is a *minimal* such I' (minimal for satisfying $II' \neq 0$) as R is artinian. Necessarily I' is nonzero, and furthermore there is $x \in I'$ with $xI \neq 0$.

   It follows by minimality that
   $$I' = (x),$$
   so I' is principal. Then $xI \neq 0$; observe that xI is also $(xI)I$ as $I^2 = I$ from the definition of I. Since $(xI)I \neq 0$, it follows again by minimality that
   $$xI = (x).$$
   Hence, there is $y \in I$ such that $xy = x$; but now, by construction $I \subset J = \mathrm{Rad}(R)$, implying that y is nilpotent. It follows that
   $$x = xy = xy^2 = \cdots = 0$$
   as y is nilpotent. However, $x \neq 0$ as $xI \neq 0$. This is a contradiction, which implies that the second case cannot occur.

We have now proved the lemma. □

Finally, we may prove:

2.4.12 Lemma *A local artinian ring R is noetherian.*

Proof. We have the filtration $R \supset \mathfrak{m} \supset \mathfrak{m}^2 \supset \dots$. This eventually stabilizes at zero by 2.4.11. I claim that R is noetherian as an R-module. To prove this, it suffices to show that $\mathfrak{m}^k/\mathfrak{m}^{k+1}$ is noetherian as an R-module. But of course, this is annihilated by $\mathfrak{m}$, so it is really a vector space over the field $R/\mathfrak{m}$. But $\mathfrak{m}^k/\mathfrak{m}^{k+1}$ is a subquotient of an artinian module, so is artinian itself. We have to show that it is noetherian. It suffices to show now that if k is a field, and V a k-vector space, then TFAE:

1. V is artinian.
2. V is noetherian.
3. V is finite-dimensional.

This is evident by linear algebra. □

Now, finally, we have shown that an artinian ring is noetherian. We have to discuss the converse. Let us assume now that R is noetherian and has only maximal prime ideals. We show that R is artinian. Let us consider $\operatorname{Spec} R$; there are only finitely many minimal primes by the theory of associated primes: every prime ideal is minimal in this case. Once again, we learn that $\operatorname{Spec} R$ is finite and has the discrete topology. This means that R is a product of factors $\prod R_i$ where each R_i is a local noetherian ring with a unique prime ideal. We might as well now prove:

2.4.13 Lemma *Let $(R, \mathfrak{m})$ be a local noetherian ring with one prime ideal. Then R is artinian.*

Proof. We know that $\mathfrak{m} = \mathrm{rad}(R)$. So $\mathfrak{m}$ consists of nilpotent elements, so by finite generatedness it is nilpotent. Then we have a finite filtration

$$R \supset \mathfrak{m} \supset \cdots \supset \mathfrak{m}^k = 0.$$

Each of the quotients are finite-dimensional vector spaces, so artinian; this implies that R itself is artinian. □

2.4.14 Remark Note that artinian implies noetherian! This statement is true for rings (even non-commutative rings), but not for modules. Take the same example $M = \varinjlim \mathbb{Z}/p^n\mathbb{Z}$ over $\mathbb{Z}$. However, there is a module-theoretic statement which is related.

2.4.15 Corollary *For a finitely generated module M over any commutative ring R, the following are equivalent.*

1. *M is an artinian module.*
2. *M has finite length (i.e. is noetherian and artinian).*
3. *$R/\operatorname{Ann} M$ is an artinian ring.*

Proof. **add: proof** □

2.4.16 Remark If R is an artinian ring, and S is a finite R-algebra (finite as an R-module), then S is artinian.

2.4.17 Remark Let M be an artinian module over a commutative ring R, $f : M \to M$ an *injective* homomorphism. Show that f is surjective, hence an isomorphism.

Vista: zero-dimensional non-noetherian rings

2.4.18 Definition (von Neumann) An element $a \in R$ is called *von Neumann regular* if there is some $x \in R$ such that $a = axa$.

2.4.19 Definition (McCoy) A element $a \in R$ is *π-regular* if some power of a is von Neumann regular.

2.4.20 Definition A element $a \in R$ is *strongly π-regular* (in the commutative case) if the chain $aR \supset a^2R \supset a^3R \supset \cdots$ stabilizes.

A ring R is von Neumann regular (resp. (strongly) π-regular) if every element of R is.

2.4.21 Theorem (5.2) *For a commutative ring R, the following are equivalent.*

1. $\dim R = 0$.
2. *R is rad-nil (i.e. the Jacobson radical $J(R)$ is the nilradical) and $R/\operatorname{Rad} R$ is von Neumann regular.*
3. *R is strongly π-regular.*

4. *R is π-regular.*

And any one of these implies

5. *Any non-zero-divisor is a unit.*

Proof. $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$ and $4 \Rightarrow 5$. We will not do $1 \Rightarrow 2 \Rightarrow 3$ here.

($3 \Rightarrow 4$) Given $a \in R$, there is some n such that $a^n R = a^{n+1} R = a^{2n} R$, which implies that $a^n = a^n x a^n$ for some x.

($4 \Rightarrow 1$) Is $\mathfrak{p}$ maximal? Let $a \notin \mathfrak{p}$. Since a is π-regular, we have $a^n = a^{2n} x$, so $a^n(1 - a^n x) = 0$, so $1 - a^n x \in \mathfrak{p}$. It follows that a has an inverse mod $\mathfrak{p}$.

($4 \Rightarrow 5$) Using $1 - a^n x = 0$, we get an inverse for a. $\square$

2.4.22 Example Any local rad-nil ring is zero dimensional, since 2 holds. In particular, for a ring S and maximal ideal $\mathfrak{m}$, $R = S/\mathfrak{m}^n$ is zero dimensional because it is a rad-nil local ring.

2.4.23 Example (Split-Null Extension) For a ring A and A-module M, let $R = A \oplus M$ with the multiplication $(a, m)(a', m') = (aa', am' + a'm)$ (i.e. take the multiplication on M to be zero). In R, M is an ideal of square zero. (A is called a *retract* of R because it sits in R and can be recovered by quotienting by some complement.) If A is a field, then R is a rad-nil local ring, with maximal ideal M.

III.3. Graded and filtered rings

In algebraic geometry, working in classical affine space $\mathbb{A}^n_{\mathbb{C}}$ of points in $\mathbb{C}^n$ turns out to be insufficient for various reasons. Instead, it is often more convenient to consider varieties in *projective space* $\mathbb{P}^n_{\mathbb{C}}$, which is the set of lines through the origin in $\mathbb{C}^{n+1}$. In other words, it is the set of all $n+1$-tuples $[z_0, \dots, z_n] \in \mathbb{C}^{n+1} - \{0\}$ modulo the relation that

$$[z_0, \dots, z_n] = [\lambda z_0, \dots, \lambda z_n], \quad \lambda \in \mathbb{C}^*. \tag{3.0.1}$$

Varieties in projective space have many convenient properties that affine varieties do not: for instance, intersections work out much more nicely when intersections at the extra "points at infinity" are included. Moreover, when endowed with the complex topology, (complex) projective varieties are *compact*, unlike all but degenerate affine varieties (i.e. finite sets).

It is when defining the notion of a "variety" in projective space that one encounters gradedness. Now a variety in $\mathbb{P}^n$ must be cut out by polynomials $F_1, \dots, F_k \in \mathbb{C}[x_0, \dots, x_n]$; that is, a point represented by $[z_0, \dots, z_n]$ lies in the associated variety if and only if $F_i(z_0, \dots, z_n) = 0$ for each i. For this to make sense, or to be independent of the choice of $z_0, \dots, z_n$ up to rescaling as in (3.0.1), it is necessary to assume that each F_i is *homogeneous.*

Algebraically, $\mathbb{A}^n_{\mathbb{C}}$ is the set of maximal ideals in the polynomial ring $\mathbb{C}^n$. Projective space is defined somewhat more geometrically (as a set of lines) but it turns out that there is an algebraic interpretation here too. The points of projective space are in bijection with the *homogeneous maximal ideals* of the polynomial ring $\mathbb{C}[x_0, \dots, x_n]$. We shall define more generally the Proj of a *graded* ring in this chapter. Although we shall not repeatedly refer to this concept in the sequel, it will be useful for readers interested in algebraic geometry.

We shall also introduce the notion of a *filtration.* A filtration allows one to endow a given module with a topology, and one can in fact complete with respect to this topology. This construction will be studied in **??**.

3.1. Graded rings and modules

Much of the material in the present section is motivated by algebraic geometry; see **?**, volume II for the construction of $\operatorname{Proj} R$ as a scheme.

Basic definitions

3.1.1 Definition A **graded ring** R is a ring together with a decomposition (as abelian groups)

$$R = R_0 \oplus R_1 \oplus \dots$$

such that $R_m R_n \subset R_{m+n}$ for all $m, n \in \mathbb{Z}_{\geqslant 0}$, and where R_0 is a subring (i.e. $1 \in R_0$). A **$\mathbb{Z}$-graded ring** is one where the decomposition is into $\bigoplus_{n \in \mathbb{Z}} R_n$. In either case, the elements of the subgroup R_n are called **homogeneous of degree** n.

The basic example to keep in mind is, of course, the polynomial ring $R[x_1, \dots, x_n]$ for R any ring. The graded piece of degree n consists of the homogeneous polynomials of degree n.

Consider a graded ring R.

3.1.2 Definition A **graded** R-module is an ordinary R-module M together with a decomposition

$$M = \bigoplus_{k \in \mathbb{Z}} M_k$$

as abelian groups, such that $R_m M_n \subset M_{m+n}$ for all $m \in \mathbb{Z}_{\geqslant 0}, n \in \mathbb{Z}$. Elements in one of these pieces are called **homogeneous.** Any $m \in M$ is thus uniquely a finite sum $\sum m_{n_i}$ where each $m_{n_i} \in M_{n_i}$ is homogeneous of degree n_i.

Clearly there is a *category* of graded R-modules, where the morphisms are the morphisms of R-modules that preserve the grading (i.e. take homogeneous elements to homogeneous elements of the same degree).

Since we shall focus on positively graded rings, we shall simply call them graded rings; when we do have to consider rings with possibly negative gradings, we shall highlight this explicitly. Note, however, that we allow modules with negative gradings freely.

In fact, we shall note an important construction that will generally shift the graded pieces such that some of them might be negative:

3.1.3 Definition Given a graded module M, we define the **twist** $M(n)$ as the same R-module but with the grading

$$M(n)_k = M_{n+k}.$$

This is a functor on the category of graded R-modules.

In algebraic geometry, the process of twisting allows one to construct canonical line bundles on projective space. Namely, a twist of R itself will lead to a line bundle on projective space that in general is not trivial. See **?**, II.5.

Here are examples:

3.1.4 Example (An easy example) If R is a graded ring, then R is a graded module over itself.

3.1.5 Example (Another easy example) If S is any ring, then S can be considered as a graded ring with $S_0 = S$ and $S_i = 0$ for $i > 0$. Then a *graded* S-module is just a $\mathbb{Z}$-indexed collection of (ordinary) S-modules.

3.1.6 Example (The blowup algebra) This example is a bit more interesting, and will be used in the sequel. Let S be any ring, and let $J \subset S$ be an ideal. We can make $R = S \oplus J \oplus J^2 \oplus \dots$ (the so-called *blowup algebra*) into a graded ring, by defining the multiplication the normal way except that something in the ith component times something in the jth component goes into the $i + j$th component.

Given any S-module M, there is a graded R-module $M \oplus JM \oplus J^2M \oplus \dots$, where multiplication is defined in the obvious way. We thus get a functor from S-modules to graded R-modules.

3.1.7 Definition Fix a graded ring R. Let M be a graded R-module and $N \subset M$ an R-submodule. Then N is called a **graded submodule** if the homogeneous components of anything in N are in N. If $M = R$, then a graded ideal is also called a **homogeneous ideal**.

In particular, a graded submodule is automatically a graded module in its own right.

3.1.8 Lemma

1. *The sum of two graded submodules (in particular, homogeneous ideals) is graded.*
2. *The intersection of two graded submodules is graded.*

Proof. Immediate. ∎

One can grade the quotients of a graded module by a graded submodule. If $N \subset M$ is a graded submodule, then M/N can be made into a graded module, via the isomorphism of abelian groups

$$M/N \simeq \bigoplus_{k \in \mathbb{Z}} M_k/N_k.$$

In particular, if $\mathfrak{a} \subset R$ is a homogeneous ideal, then $R/\mathfrak{a}$ is a graded ring in a natural way.

3.1.9 Remark (exercise) Let R be a graded ring. Does the category of graded R-modules admit limits and colimits?

Homogeneous ideals

Recall that a homogeneous ideal in a graded ring R is simply a graded submodule of R. We now prove a useful result that enables us tell when an ideal is homogeneous.

3.1.10 Proposition *Let R be a graded ring, $I \subset R$ an ideal. Then I is a homogeneous ideal if and only if it can be generated by homogeneous elements.*

Proof. If I is a homogeneous ideal, then by definition

$$I = \bigoplus_i I \cap R_i,$$

so I is generated by the sets $\{I \cap R_i\}_{i \in \mathbb{Z}_{\geqslant 0}}$ of homogeneous elements.

Conversely, let us suppose that I is generated by homogeneous elements $\{h_\alpha\}$. Let $x \in I$ be arbitrary; we can uniquely decompose x as a sum of homogeneous elements, $x = \sum x_i$, where each $x_i \in R_i$. We need to show that each $x_i \in I$ in fact.

To do this, note that $x = \sum q_\alpha h_\alpha$ where the q_α belong to R. If we take ith homogeneous components, we find that

$$x_i = \sum (q_\alpha)_{i - \deg h_\alpha} h_\alpha,$$

where $(q_\alpha)_{i-\deg h_\alpha}$ refers to the homogeneous component of q_α concentrated in the degree $i - \deg h_\alpha$. From this it is easy to see that each x_i is a linear combination of the h_α and consequently lies in I. ∎

3.1.11 Example If $\mathfrak{a}, \mathfrak{b} \subset R$ are homogeneous ideals, then so is $\mathfrak{ab}$. This is clear from proposition 3.1.10.

3.1.12 Example Let k be a field. The ideal $(x^2 + y)$ in $k[x, y]$ is *not* homogeneous. However, we find from proposition 3.1.10 that the ideal $(x^2 + y^2, y^3)$ is.

Since we shall need to use them to define $\operatorname{Proj} R$ in the future, we now prove a result about homogeneous *prime* ideals specifically. Namely, "primeness" can be checked just on homogeneous elements for a homogeneous ideal.

3.1.13 Lemma *Let $\mathfrak{p} \subset R$ be a homogeneous ideal. In order that $\mathfrak{p}$ be prime, it is necessary and sufficient that whenever x, y are* homogeneous *elements such that $xy \in \mathfrak{p}$, then at least one of $x, y \in \mathfrak{p}$.*

Proof. Necessity is immediate. For sufficiency, suppose $a, b \in R$ and $ab \in \mathfrak{p}$. We must prove that one of these is in $\mathfrak{p}$. Write

$$a = a_{k_1} + a_1 + \cdots + a_{k_2}, \ b = b_{m_1} + \cdots + b_{m_2}$$

as a decomposition into homogeneous components (i.e. a_i is the ith component of a), where a_{k_2}, b_{m_2} are nonzero and of the highest degree.

Let $k = k_2 - k_1, m = m_2 - m_1$. So there are k homogeneous terms in the expression for a, m in the expression for b. We will prove that one of $a, b \in \mathfrak{p}$ by induction on $m + n$. When $m + n = 0$, then it is just the condition of the lemma. Suppose it true for smaller values of $m + n$. Then ab has highest homogeneous component $a_{k_2} b_{m_2}$, which must be in $\mathfrak{p}$ by homogeneity. Thus one of a_{k_2}, b_{m_2} belongs to $\mathfrak{p}$. Say for definiteness it is a_k. Then we have that

$$(a - a_{k_2}) b \equiv ab \equiv 0 \bmod \mathfrak{p}$$

so that $(a - a_{k_2}) b \in \mathfrak{p}$. But the resolutions of $a - a_{k_2}, b$ have a smaller $m + n$-value: $a - a_{k_2}$ can be expressed with $k - 1$ terms. By the inductive hypothesis, it follows that one of these is in $\mathfrak{p}$, and since $a_k \in \mathfrak{p}$, we find that one of $a, b \in \mathfrak{p}$. ∎

Finiteness conditions

There are various finiteness conditions (e.g. noetherianness) that one often wants to impose in algebraic geometry. Since projective varieties (and schemes) are obtained from graded rings, we briefly discuss these finiteness conditions for them.

3.1.14 Definition For a graded ring R, write $R_+ = R_1 \oplus R_2 \oplus \dots$. Clearly $R_+ \subset R$ is a homogeneous ideal. It is called the **irrelevant ideal.**

When we define the Proj of a ring, prime ideals containing the irrelevant ideal will be no good. The intuition is that when one is working with $\mathbb{P}^n_{\mathbb{C}}$, the irrelevant ideal in the corresponding ring $\mathbb{C}[x_0, \dots, x_n]$ corresponds to *all* homogeneous polynomials of positive degree. Clearly these have no zeros except for the origin, which is not included in projective space: thus the common zero locus of the irrelevant ideal should be $\varnothing \subset \mathbb{P}^n_{\mathbb{C}}$.

3.1.15 Proposition *Suppose $R = R_0 \oplus R_1 \oplus \dots$ is a graded ring. Then if a subset $S \subset R_+$ generates the irrelevant ideal R_+ as R-ideal, it generates R as R_0-algebra.*

The converse is clear as well. Indeed, if $S \subset R_+$ generates R as an R_0-algebra, clearly it generates R_+ as an R-ideal.

Proof. Let $T \subset R$ be the R_0-algebra generated by S. We shall show inductively that $R_n \subset T$. This is true for $n = 0$. Suppose $n > 0$ and the assertion true for smaller n. Then, we have

$$\begin{aligned} R_n &= RS \cap R_n \text{ by assumption} \\ &= (R_0 \oplus R_1 \oplus \dots \oplus R_{n-1})(S) \cap R_n \text{ because } S \subset R_+ \\ &\subset (R_0[S])(S) \cap R_n \text{ by inductive hypothesis} \\ &\subset R_0(S). \end{aligned}$$

□

3.1.16 Theorem *The graded ring R is noetherian if and only if R_0 is noetherian and R is finitely generated as R_0-algebra.*

Proof. One direction is clear by Hilbert's basis theorem. For the other, suppose R noetherian. Then R_0 is noetherian because any sequence $I_1 \subset I_2 \subset \dots$ of ideals of R_0 leads to a sequence of ideals $I_1 R \subset I_2 R \subset \dots$, and since these stabilize, the original $I_1 \subset I_2 \subset \dots$ must stabilize too. (Alternatively, $R_0 = R/R_+$, and taking quotients preserves noetherianness.) Moreover, since R_+ is a finitely generated R-ideal by noetherianness, it follows that R is a finitely generated R_0-algebra too: we can, by proposition 3.1.15, take as R_0-algebra generators for R a set of generators for the *ideal* R_+. □

The basic finiteness condition one often needs is that R should be finitely generated as an R_0-algebra. We may also want to have that R is generated by R_1, quite frequently—in algebraic geometry, this implies a bunch of useful things about certain sheaves being invertible. (See **?**, volume II.2.) As one example, having R generated as R_0-algebra by R_1 is equivalent to having R a *graded* quotient of a polynomial algebra over R_0 (with the usual grading). Geometrically, this equates to having $\operatorname{Proj} R$ contained as a closed subset of some projective space over R_0.

However, sometimes we have the first condition and not the second, though if we massage things we can often assure generation by R_1. Then the next idea comes in handy.

3.1.17 Definition Let R be a graded ring and $d \in \mathbb{N}$. We set $R^{(d)} = \bigoplus_{k \in \mathbb{Z}_{\geqslant 0}} R_{kd}$; this is a graded ring and R_0-algebra. If M is a graded R-module and $l \in \{0, 1, \ldots, d-1\}$, we write $M^{(d,l)} = \bigoplus_{k \equiv l \bmod d} M_k$. Then $M^{(d,l)}$ is a graded $R^{(d)}$-module.

We in fact have a functor $\cdot^{(d,l)}$ from graded R-modules to graded $R^{(d)}$-modules.

One of the implications of the next few results is that, by replacing R with $R^{(d)}$, we can make the condition "generated by terms of degree 1" happen. But first, we show that basic finiteness is preserved if we filter out some of the terms.

3.1.18 Proposition *Let R be a graded ring and a finitely generated R_0-algebra. Let M be a finitely generated R-module.*

1. *Each M_i is finitely generated over R_0, and the M_i become zero when $i \ll 0$.*
2. *$M^{(d,l)}$ is a finitely generated $R^{(d)}$ module for each d, l. In particular, M itself is a finitely generated $R^{(d)}$-module.*
3. *$R^{(d)}$ is a finitely generated R_0-algebra.*

Proof. Choose homogeneous generators $m_1, \ldots, m_k \in M$. For instance, we can choose the homogeneous components of a finite set of generators for M. Then every nonzero element of M has degree at least $\min(\deg m_i)$. This proves the last part of (1). Moreover, let $r_1, \ldots, r_p$ be algebra generators of R over R_0. We can assume that these are homogeneous with positive degrees $d_1, \ldots, d_p > 0$. Then the R_0-module M_i is generated by the elements

$$r_1^{a_1} \ldots r_p^{a_p} m_s$$

where $\sum a_j d_j + \deg m_s = i$. Since the $d_j > 0$ and there are only finitely many m_s's, there are only finitely many such elements. This proves the rest of (1).

To prove (2), note first that it is sufficient to show that M is finitely generated over $R^{(d)}$, because the $M^{(d,l)}$ are $R^{(d)}$-homomorphic images (i.e. quotient by the $M^{(d',l)}$ for $d' \neq d$). Now M is generated as R_0-module by the $r_1^{a_1} \ldots r_p^{a_p} m_s$ for $a_1, \ldots, a_p \geqslant 0$ and $s = 1, \ldots, k$. In particular, by the euclidean algorithm in elementary number theory, it follows that the $r_1^{a_1} \ldots r_p^{a_p} m_s$ for $a_1, \ldots, a_p \in [0, d-1]$ and $s = 1, \ldots, k$ generate M over $R^{(d)}$, as each power $r_i^d \in R^{(d)}$. In particular, R is finitely generated over $R^{(d)}$.

When we apply (2) to the finitely generated R-module R_+, it follows that $R_+^{(d)}$ is a finitely generated $R^{(d)}$-module. This implies that $R^{(d)}$ is a finitely generated R_0-algebra by proposition 3.1.15. ∎

In particular, by proposition 4.1.12 (later in the book!) R is *integral* over $R^{(d)}$: this means that each element of R satisfies a monic polynomial equation with $R^{(d)}$-coefficients. This can easily be seen directly. The dth power of a homogeneous element lies in $R^{(d)}$.

3.1.19 Remark Part (3), the preservation of the basic finiteness condition, could also be proved as follows, at least in the noetherian case (with $S = R^{(d)}$). We shall assume familiarity with the material in **??** for this brief digression.

3.1.20 Lemma *Suppose $R_0 \subset S \subset R$ is an inclusion of rings with R_0 noetherian. Suppose R is a finitely generated R_0-algebra and R/S is an integral extension. Then S is a finitely generated R_0-algebra.*

In the case of interest, we can take $S = R^{(d)}$. The point of the lemma is that finite generation can be deduced for *subrings* under nice conditions.

Proof. We shall start by finding a subalgebra $S' \subset S$ such that R is integral over S', but S' is a finitely generated R_0-algebra. The procedure will be a general observation of the flavor of "noetherian descent" to be developed in **??**. Then, since R is integral over S' and finitely generated as an *algebra*, it will be finitely generated as a S'-module. S, which is a sub-S'-module, will equally be finitely generated as a S'-module, hence as an R_0-algebra. So the point is to make S finitely generated as a module over a "good" ring.

Indeed, let $r_1, \dots, r_m$ be generators of R/R_0. Each satisfies an integral equation $r_k^{n_k} + P_k(r_k) = 0$, where $P_k \in S[X]$ has degree less than n_k. Let $S' \subset S \subset R$ be the subring generated over R_0 by the coefficients of all these polynomials P_k.

Then R is, by definition, integral over S'. Since R is a finitely generated S'-algebra, it follows by proposition 4.1.12 that it is a finitely generated S'-module. Then S, as a S'-submodule is a finitely generated S'-module by noetherianness. Therefore, S is a finitely generated R_0-algebra.□

This result implies, incidentally, the following useful corollary:

3.1.21 Corollary *Let R be a noetherian ring. If a finite group G acts on a finitely generated R-algebra S, the ring of invariants S^G is finitely generated.*

Proof. Apply lemma 3.1.20 to R, S^G, S. One needs to check that S is integral over S^G. But each $s \in S$ satisfies the equation

$$\prod_{\sigma \in G} (X - \sigma(s)),$$

which has coefficients in S^G. □

This ends the digression.

We next return to our main goals, and let R be a graded ring, finitely generated as an R_0-algebra, as before; let M be a finitely generated R-module. We show that we can have $R^{(d)}$ generated by terms of degree d (i.e. "degree 1" if we rescale) for d chosen large.

3.1.22 Lemma *Hypotheses as above, there is a pair (d, n_0) such that*

$$R_d M_n = M_{n+d}$$

for $n \geqslant n_0$.

Proof. Indeed, select R-module generators $m_1, \ldots, m_k \in M$ and R_0-algebra generators $r_1, \ldots, r_p \in R$ as in the proof of proposition 3.1.18; use the same notation for their degrees, i.e. $d_j = \deg r_j$. Let d be the least common multiple of the d_j. Consider the family of elements

$$s_i = r_i^{d/d_i} \in R_d.$$

Then suppose $m \in M_n$ for $n > d + \sup \deg m_i$. We have that m is a sum of products of powers of the $\{r_j\}$ and the $\{m_i\}$, each term of which we can assume is of degree n. In this case, since in each term, at least one of the $\{r_j\}$ must occur to power $\geqslant \frac{d}{d_j}$, we can write each term in the sum as some s_j times something in M_{n-d}.

In particular, $M_n = R_d M_{n-d}$. ∎

3.1.23 Proposition *Suppose R is a graded ring and finitely generated R_0-algebra. Then there is $d \in \mathbb{N}$ such that $R^{(d)}$ is generated over R_0 by R_d.*

What this proposition states geometrically is that if we apply the functor $R \mapsto R^{(d)}$ for large d (which, geometrically, is actually harmless), one can arrange things so that $\operatorname{Proj} R$ (not defined yet!) is contained as a closed subscheme of ordinary projective space.

Proof. Consider R as a finitely generated, graded R-module. Suppose d' is as in the proposition 3.1.23 (replacing d, which we reserve for something else), and choose n_0 accordingly. So we have $R_{d'} R_m = R_{m+d'}$ whenever $m \geqslant n_0$. Let d be a multiple of d' which is greater than n_0.

Then, iterating, we have $R_d R_n = R_{d+n}$ if $n \geqslant d$ since d is a multiple of d'. In particular, it follows that $R_{nd} = (R_d)^n$ for each $n \in \mathbb{N}$, which implies the statement of the proposition. ∎

As we will see below, taking $R^{(d)}$ does not affect the Proj, so this is extremely useful.

3.1.24 Example Let k be a field. Then $R = k[x^2] \subset k[x]$ (with the grading induced from $k[x]$) is a finitely generated graded k-algebra, which is not generated by its elements in degree one (there are none!). However, $R^{(2)} = k[x^2]$ is generated by x^2.

We next show that taking the $R^{(d)}$ *always* preserves noetherianness.

3.1.25 Proposition *If R is noetherian, then so is $R^{(d)}$ for any $d > 0$.*

Proof. If R is noetherian, then R_0 is noetherian and R is a finitely generated R_0-algebra by theorem 3.1.16. proposition 3.1.18 now implies that $R^{(d)}$ is also a finitely generated R_0-algebra, so it is noetherian. ∎

The converse is also true, since R is a finitely generated $R^{(d)}$-module.

Localization of graded rings

Next, we include a few topics that we shall invoke later on. First, we discuss the interaction of homogeneity and localization. Under favorable circumstances, we can give $\mathbb{Z}$-gradings to localizations of graded rings.

3.1.26 Definition If $S \subset R$ is a multiplicative subset of a graded (or $\mathbb{Z}$-graded) ring R consisting of homogeneous elements, then $S^{-1}R$ is a $\mathbb{Z}$-graded ring: we let the homogeneous elements of degree n be of the form r/s where $r \in R_{n+\deg s}$. We write $R_{(S)}$ for the subring of elements of degree zero; there is thus a map $R_0 \to R_{(S)}$.

If S consists of the powers of a homogeneous element f, we write $R_{(f)}$ for R_S. If $\mathfrak{p}$ is a homogeneous ideal and S the set of homogeneous elements of R not in $\mathfrak{p}$, we write $R_{(\mathfrak{p})}$ for $R_{(S)}$.

Of course, $R_{(S)}$ has a trivial grading, and is best thought of as a plain, unadorned ring. We shall show that $R_{(f)}$ is a special case of something familiar.

3.1.27 Proposition *Suppose f is of degree d. Then, as plain rings, there is a canonical isomorphism $R_{(f)} \simeq R^{(d)}/(f-1)$.*

Proof. The homomorphism $R^{(d)} \to R_{(f)}$ is defined to map $g \in R_{kd}$ to $g/f^d \in R_{(f)}$. This is then extended by additivity to non-homogeneous elements. It is clear that this is multiplicative, and that the ideal $(f-1)$ is annihilated by the homomorphism. Moreover, this is surjective.

We shall now define an inverse map. Let $x/f^n \in R_{(f)}$; then x must be a homogeneous element of degree divisible by d. We map this to the residue class of x in $R^{(d)}/(f-1)$. This is well-defined; if $x/f^n = y/f^m$, then there is N with

$$f^N(xf^m - yf^n) = 0,$$

so upon reduction (note that f gets reduced to 1!), we find that the residue classes of x, y are the same, so the images are the same.

Clearly this defines an inverse to our map. $\square$

3.1.28 Corollary *Suppose R is a graded noetherian ring. Then each of the $R_{(f)}$ is noetherian.*

Proof. This follows from the previous result and the fact that $R^{(d)}$ is noetherian (3.1.25).

More generally, we can define the localization procedure for graded modules.

3.1.29 Definition Let M be a graded R-module and $S \subset R$ a multiplicative subset consisting of homogeneous elements. Then we define $M_{(S)}$ as the submodule of the graded module $S^{-1}M$ consisting of elements of degree zero. When S consists of the powers of a homogeneous element $f \in R$, we write $M_{(f)}$ instead of $M_{(S)}$. We similarly define $M_{(\mathfrak{p})}$ for a homogeneous prime ideal $\mathfrak{p}$.

Then clearly $M_{(S)}$ is a $R_{(S)}$-module. This is evidently a functor from graded R-modules to $R_{(S)}$-modules.

We next observe that there is a generalization of 3.1.27.

3.1.30 Proposition *Suppose M is a graded R-module, $f \in R$ homogeneous of degree d. Then there is an isomorphism*

$$M_{(f)} \simeq M^{(d)}/(f-1)M^{(d)}$$

of $R^{(d)}$-modules.

Proof. This is proved in the same way as 3.1.27. Alternatively, both are right-exact functors that commute with arbitrary direct sums and coincide on R, so must be naturally isomorphic by a well-known bit of abstract nonsense.[1] ☐

In particular:

3.1.31 Corollary *Suppose M is a graded R-module, $f \in R$ homogeneous of degree 1. Then we have*

$$M_{(f)} \simeq M/(f-1)M \simeq M \otimes_R R/(f-1).$$

The Proj of a ring

Let $R = R_0 \oplus R_1 \oplus \dots$ be a **graded ring**.

3.1.32 Definition Let $\operatorname{Proj} R$ denote the set of *homogeneous prime ideals* of R that do not contain the **irrelevant ideal** R_+.[2]

We can put a topology on $\operatorname{Proj} R$ by setting, for a homogeneous ideal $\mathfrak{b}$,

$$V(\mathfrak{b}) = \{\mathfrak{p} \in \operatorname{Proj} R : \mathfrak{p} \supset \mathfrak{b}\}$$

. These sets satisfy

1. $V(\sum \mathfrak{b}_{\mathfrak{i}}) = \bigcap V(\mathfrak{b}_{\mathfrak{i}})$.
2. $V(\mathfrak{a}\mathfrak{b}) = V(\mathfrak{a}) \cup V(\mathfrak{b})$.
3. $V(\operatorname{Rad} \mathfrak{a}) = V(\mathfrak{a})$.

Note incidentally that we would not get any more closed sets if we allowed all ideals $\mathfrak{b}$, since to any $\mathfrak{b}$ we can consider its "homogenization." We could even allow all sets.

In particular, the V's do in fact yield a topology on $\operatorname{Proj} R$ (setting the open sets to be complements of the V's). As with the affine case, we can define basic open sets. For f homogeneous of positive degree, define $D'(f)$ to be the collection of homogeneous ideals (not containing R_+) that do not contain f; clearly these are open sets.

Let $\mathfrak{a}$ be a homogeneous ideal. Then we claim that:

[1] Citation needed.
[2] Recall that an ideal $\mathfrak{a} \subset R$ for R graded is *homogeneous* if the homogeneous components of $\mathfrak{a}$ belong to $\mathfrak{a}$.

3.1.33 Lemma $V(\mathfrak{a}) = V(\mathfrak{a} \cap R_+)$.

Proof. Indeed, suppose $\mathfrak{p}$ is a homogeneous prime not containing S_+ such that all homogeneous elements of positive degree in $\mathfrak{a}$ (i.e., anything in $\mathfrak{a} \cap R_+$) belongs to $\mathfrak{p}$. We will show that $\mathfrak{a} \subset \mathfrak{p}$.

Choose $a \in \mathfrak{a} \cap R_0$. It is sufficient to show that any such a belongs to $\mathfrak{p}$ since we are working with homogeneous ideals. Let f be a homogeneous element of positive degree that is not in $\mathfrak{p}$. Then $af \in \mathfrak{a} \cap R_+$, so $af \in \mathfrak{p}$. But $f \notin \mathfrak{p}$, so $a \in \mathfrak{p}$. ☐

Thus, when constructing these closed sets $V(\mathfrak{a})$, it suffices to work with ideals contained in the irrelevant ideal. In fact, we could take $\mathfrak{a}$ in any prescribed power of the irrelevant ideal, since taking radicals does not affect V.

3.1.34 Proposition *We have $D'(f) \cap D'(g) = D'(fg)$. Also, the $D'(f)$ form a basis for the topology on* $\operatorname{Proj} R$*.*

Proof. The first part is evident, by the definition of a prime ideal. We prove the second. Note that $V(\mathfrak{a})$ is the intersection of the $V((f))$ for the homogeneous $f \in \mathfrak{a} \cap R_+$. Thus $\operatorname{Proj} R - V(\mathfrak{a})$ is the union of these $D'(f)$. So every open set is a union of sets of the form $D'(f)$. ☐

We shall now show that the topology is actually rather familiar from the affine case, which is not surprising, since the definition is similar.

3.1.35 Proposition *$D'(f)$ is homeomorphic to* $\operatorname{Spec} R_{(f)}$ *under the map*

$$\mathfrak{p} \to \mathfrak{p}R_f \cap R_{(f)}$$

sending homogeneous prime ideals of R not containing f into primes of $R_{(f)}$.

Proof. Indeed, let $\mathfrak{p}$ be a homogeneous prime ideal of R not containing f. Consider $\phi(\mathfrak{p}) = \mathfrak{p}R_f \cap R_{(f)}$ as above. This is a prime ideal, since $\mathfrak{p}R_f$ is a prime ideal in R_f by basic properties of localization, and $R_{(f)} \subset R_f$ is a subring. (It cannot contain the identity, because that would imply that a power of f lay in $\mathfrak{p}$.)

So we have defined a map $\phi : D'(f) \to \operatorname{Spec} R_{(f)}$. We can define its inverse ψ as follows. Given $\mathfrak{q} \subset R_{(f)}$ prime, we define a prime ideal $\mathfrak{p} = \psi(\mathfrak{q})$ of R by saying that a *homogeneous* element $x \in R$ belongs to $\mathfrak{p}$ if and only if $x^{\deg f}/f^{\deg x} \in \mathfrak{q}$. It is easy to see that this is indeed an ideal, and that it is prime by 3.1.13.

Furthermore, it is clear that $\phi \circ \psi$ and $\psi \circ \phi$ are the identity. This is because $x \in \mathfrak{p}$ for $\mathfrak{p} \in D'(f)$ if and only if $f^n x \in \mathfrak{p}$ for some n.

We next need to check that these are continuous, hence homeomorphisms. If $\mathfrak{a} \subset R$ is a homogeneous ideal, then $V(\mathfrak{a}) \cap D'(f)$ is mapped to $V(\mathfrak{a}R_f \cap R_{(f)}) \subset \operatorname{Spec} R_{(f)}$, and vice versa. ☐

3.2. Filtered rings

In practice, one often has something weaker than a grading. Instead of a way of saying that an element is of degree d, one simply has a way of saying that an element is "of degree at most d." This leads to the definition of a *filtered* ring (and a filtered module). We shall use this definition in placing topologies on rings and modules and, later, completing them.

Definition

3.2.1 Definition A **filtration** on a ring R is a sequence of ideals $R = I_0 \supset I_1 \supset \dots$ such that $I_m I_n \subset I_{m+n}$ for each $m, n \in \mathbb{Z}_{\geqslant 0}$. A ring with a filtration is called a **filtered ring**.

A filtered ring is supposed to be a generalization of a graded ring. If $R = \bigoplus R_k$ is graded, then we can make R into a filtered ring in a canonical way by taking the ideal $I_m = \bigoplus_{k \geqslant m} R_k$ (notice that we are using the fact that R has only pieces in nonnegative gradings!).

We can make filtered rings into a category: a morphism of filtered rings $\phi : R \to S$ is a ring-homomorphism preserving the filtration.

3.2.2 Example (The *I*-adic filtration) Given an ideal $I \subset R$, we can take powers of I to generate a filtration. This filtration $R \supset I \supset I^2 \supset \dots$ is called the **I-adic filtration,** and is especially important when R is local and I the maximal ideal.

If one chooses the polynomial ring $k[x_1, \dots, x_n]$ over a field with n variables and takes the $(x_1, \dots, x_n)$-adic filtration, one gets the same as the filtration induced by the usual grading.

3.2.3 Example As a specialization of the previous example, consider the power series ring $R = k[[x]]$ over a field k with one indeterminate x. This is a local ring (with maximal ideal (x)), and it has a filtration with $R_i = (x^i)$. Note that this ring, unlike the polynomial ring, is *not* a graded ring in any obvious way.

When we defined graded rings, the first thing we did thereafter was to define the notion of a graded module over a graded ring. We do the analogous thing for filtered modules.

3.2.4 Definition Let R be a filtered ring with a filtration $I_0 \supset I_1 \supset \dots$. A **filtration** on an R-module M is a decreasing sequence of submodules

$$M = M_0 \supset M_1 \supset M_2 \supset \dots$$

such that $I_m M_n \subset M_{n+m}$ for each m, n. A module together with a filtration is called a **filtered module.**

As usual, there is a category of filtered modules over a fixed filtered ring R, with morphisms the module-homomorphisms that preserve the filtrations.

3.2.5 Example (The *I*-adic filtration for modules) Let R be any ring and $I \subset R$ any ideal. Then if we make R into a filtered ring with the I-adic filtration, we can make any R-module M into a filtered R-module by giving M the filtration

$$M \supset IM \supset I^2 M \supset \dots,$$

which is also called the **I-adic filtration.**

The associated graded

We shall now describe a construction that produces graded things from filtered ones.

3.2.6 Definition Given a filtered ring R (with filtration $\{I_n\}$), the **associated graded ring** $\mathrm{gr}(R)$ is the graded ring

$$\mathrm{gr}(R) = \bigoplus_{n=0}^{\infty} I_n/I_{n+1}.$$

This is made into a ring by the following procedure. Given $a \in I_n$ representing a class $\overline{a} \in I_n/I_{n+1}$ and $b \in I_m$ representing a class $\overline{b} \in I_m/I_{m+1}$, we define $\overline{a}\overline{b}$ to be the class in I_{n+m}/I_{n+m+1} represented by ab.

It is easy to check that if different choices of representing elements a, b were made in the above description, the value of $\overline{a}\overline{b}$ thus defined would still be the same, so that the definition is reasonable.

3.2.7 Example Consider $R = \mathbb{Z}_{(p)}$ (the localization at (p)) with the (p)-adic topology. Then $\mathrm{gr}(R) = \mathbb{Z}/p[t]$, as a graded ring. For the successive quotients of ideals are of the form $\mathbb{Z}/p$, and it is easy to check that multiplication lines up in the appropriate form.

In general, as we will see below, when one takes the gr of a noetherian ring with the I-adic topology for some ideal I, one always gets a noetherian ring.

3.2.8 Definition Let R be a filtered ring, and M a filtered R-module (with filtration $\{M_n\}$). We define the **associated graded module** $\mathrm{gr}(M)$ as the graded $\mathrm{gr}(R)$-module

$$\mathrm{gr}(M) = \bigoplus_{n} M_n/M_{n+1}$$

where multiplication by an element of $\mathrm{gr}(R)$ is defined in a similar manner as above.

In other words, we have defined a *functor* gr from the category of filtered R-modules to the category of *graded* $\mathrm{gr}(R)$ modules.

Let R be a filtered ring, and M a finitely generated filtered R-module. In general, $\mathrm{gr}(M)$ *cannot* be expected to be a finitely generated $\mathrm{gr}(R)$-module.

3.2.9 Example Consider the ring $\mathbb{Z}_{(p)}$ (the localization of $\mathbb{Z}$ at p), which we endow with the p^2-adic (i.e., (p^2)-adic) filtration. The associated graded is $\mathbb{Z}/p^2[t]$.

Consider $M = \mathbb{Z}_{(p)}$ with the filtration $M_m = (p^m)$, i.e. the usual (p)-adic topology. The claim is that $\mathrm{gr}(M)$ is *not* a finitely generated $\mathbb{Z}/p^2[t]$-module. This will follow from **??** below, but we can see it directly: multiplication by t acts by zero on $\mathrm{gr}(M)$ (because this corresponds to multiplying by p^2 and shifting the degree by one). However, $\mathrm{gr}(M)$ is nonzero in every degree. If $\mathrm{gr}(M)$ were finitely generated, it would be a finitely generated $\mathbb{Z}/p^2\mathbb{Z}$-module, which it is not.

Topologies

We shall now see that filtered rings and modules come naturally with *topologies* on them.

3.2.10 Definition A **topological ring** is a ring R together with a topology such that the natural maps

$$R \times R \to R, \quad (x, y) \mapsto x + y$$
$$R \times R \to R, \quad (x, y) \mapsto xy$$
$$R \to R, \quad x \mapsto -x$$

are continuous (where $R \times R$ has the product topology).

add: discussion of algebraic objects in categories

In practice, the topological rings that we will be interested will exclusively be *linearly* topologized rings.

3.2.11 Definition A topological ring is **linearly topologized** if there is a neighborhood basis at 0 consisting of open ideals.

Given a filtered ring R with a filtration of ideals $\{I_n\}$, we can naturally linearly topologize R. Namely, we take as a basis the cosets $x + I_n$ for $x \in R, n \in \mathbb{Z}_{\geqslant 0}$. It is then clear that the $\{I_n\}$ form a neighborhood basis at the origin (because any neighborhood $x + I_n$ containing 0 must just be I_n!).

3.2.12 Example For instance, given any ring R and any ideal $I \subset R$, we can consider the **I-adic topology** on R. Here an element is "small" (i.e., close to zero) if it lies in a high power of I.

3.2.13 Proposition *A topology on R defined by the filtration $\{I_n\}$ is Hausdorff if and only if $\bigcap I_n = 0$.*

Proof. Indeed, to say that R is Hausdorff is to say that any two distinct elements $x, y \in R$ can be separated by disjoint neighborhoods. If $\bigcap I_n = 0$, we can find N large such that $x - y \notin I_N$. Then $x + I_N, y + I_N$ are disjoint neighborhoods of x, y. The converse is similar: if $\bigcap I_n \neq 0$, then no neighborhoods can separate a nonzero element in $\bigcap I_n$ from 0. $\square$

Similarly, if M is a filtered R-module with a filtration $\{M_n\}$, we can topologize M by choosing the $\{M_n\}$ to be a neighborhood basis at the origin. Then M becomes a *topological group,* that is a group with a topology such that the group operations are continuous. In the same way, we find:

3.2.14 Proposition *The topology on M is Hausdorff if and only if $\bigcap M_n = 0$.*

Moreover, because of the requirement that $R_m M_n \subset M_{n+m}$, it is easy to see that the map

$$R \times M \to M$$

is itself continuous. Thus, M is a *topological* module.

Here is another example. Suppose M is a linearly topologized module with a basis of submodules $\{M_\alpha\}$ at the origin. Then any submodule $N \subset M$ becomes a linearly topologized module with a basis of submodules $\{N \cap M_\alpha\}$ at the origin with the relative topology.

3.2.15 Proposition *Suppose M is filtered with the $\{M_n\}$. If $N \subset M$ is any submodule, then the closure $\overline{N}$ is the intersection $\bigcap N + M_n$.*

Proof. Recall that $x \in \overline{N}$ is the same as stipulating that every neighborhood of x intersect N. In other words, any basic neighborhood of x has to intersect N. This means that for each n, $x + M_n \cap N \neq \emptyset$, or in other words $x \in M_n + N$. ☐

3.3. The Artin-Rees Lemma

We shall now show that for *noetherian* rings and modules, the I-adic topology is stable under passing to submodules; this useful result, the Artin-Rees lemma, will become indispensable in our analysis of dimension theory in the future.

More precisely, consider the following problem. Let R be a ring and $I \subset R$ an ideal. Then for any R-module M, we can endow M with the I-adic filtration $\{I^n M\}$, which defines a topology on M. If $N \subset M$ is a submodule, then N inherits the subspace topology from M (i.e. that defined by the filtration $\{I^n M \cap N\}$). But N can also be topologized by simply taking the I-adic topology on it. The Artin-Rees lemma states that these two approaches give the same result.

The Artin-Rees Lemma

3.3.1 Theorem (Artin-Rees lemma) *Let R be noetherian, $I \subset R$ an ideal. Suppose M is a finitely generated R-module and $M' \subset M$ a submodule. Then the I-adic topology on M induces the I-adic topology on M'. More precisely, there is a constant c such that*

$$I^{n+c} M \cap M' \subset I^n M'.$$

So the two filtrations $\{I^n M \cap M'\}, \{I^n M'\}$ on M' are equivalent up to a shift.

Proof. The strategy to prove Artin-Rees will be as follows. Call a filtration $\{M_n\}$ on an R-module M (which is expected to be compatible with the I-adic filtration on R, i.e. $I^n M_m \subset M_{m+n}$ for all n, m) ***I*-good** if $IM_n = M_{n+1}$ for large $n \gg 0$. Right now, we have the very I-good filtration $\{I^n M\}$ on M, and the induced filtration $\{I^n M \cap M'\}$ on M'. The Artin-Rees lemma can be rephrased as saying that this filtration on M' is I-good: in fact, this is what we shall prove. It follows that if one has an I-good filtration on M, then the induced filtration on M' is itself I-good.

To do this, we shall give an interpretation of I-goodness in terms of the *blowup algebra*, and use its noetherianness. Recall that this is defined as $S = R \oplus I \oplus I^2 + \dots$, where multiplication is defined in the obvious manner (see example 3.1.6). It can be regarded as a subring of the polynomial ring $R[t]$ where the coefficient of t^i is required to be in I^i. The blowup algebra is clearly a graded ring.

Given a filtration $\{M_n\}$ on an R-module M (compatible with the I-adic filtration of M), we can make $\bigoplus_{n=0}^{\infty} M_n$ into a *graded* S-module in an obvious manner.

Here is the promised interpretation of I-goodness:

3.3.2 Lemma *Then the filtration $\{M_n\}$ of the finitely generated R-module M is I-good if and only if $\bigoplus M_n$ is a finitely generated S-module.*

Proof. Let $S_1 \subset S$ be the subset of elements of degree one. If $\bigoplus M_n$ is finitely generated as an S-module, then $S_1(\bigoplus M_n)$ and $\bigoplus M_n$ agree in large degrees by lemma 3.1.22; however, this means that $IM_{n-1} = M_n$ for $n \gg 0$, which is I-goodness.

Conversely, if $\{M_n\}$ is an I-good filtration, then once the I-goodness starts (say, for $n > N$, we have $IM_n = M_{n+1}$), there is no need to add generators beyond M_N. In fact, we can use R-generators for $M_0, \dots, M_N$ in the appropriate degrees to generate $\bigoplus M_n$ as an R'-module. ∎

Finally, let $\{M_n\}$ be an I-good filtration on the finitely generated R-module M. Let $M' \subset M$ be a submodule; we will, as promised, show that the induced filtration on M' is I-good. Now the associated module $\bigoplus_{n=0}^{\infty}(I^n M \cap M')$ is an S-submodule of $\bigoplus_{n=0}^{\infty} M_n$, which by lemma 3.3.2 is finitely generated. We will show next that S is noetherian, and consequently submodules of finitely generated modules are finitely generated. Applying lemma 3.3.2 again, we will find that the induced filtration must be I-good.

3.3.3 Lemma *Hypotheses as above, the blowup algebra R' is noetherian.*

Proof. Choose generators $x_1, \dots, x_n \in I$; then there is a map $R[y_1, \dots, y_n] \to S$ sending $y_i \to x_i$ (where x_i is in degree one). This is surjective. Hence by the basis theorem (corollary 2.1.13), R' is noetherian. ∎

The Krull intersection theorem

We now prove a useful consequence of the Artin-Rees lemma and Nakayama's lemma. In fancier language, this states that the map from a noetherian local ring into its completion is an *embedding*. A priori, this might not be obvious. For instance, it might be surprising that the inverse limit of the highly torsion groups $\mathbb{Z}/p^n$ turns out to be the torsion-free ring of p-adic integers.

3.3.4 Theorem (Krull intersection theorem) *Let R be a local noetherian ring with maximal ideal $\mathfrak{m}$. Then,*

$$\bigcap \mathfrak{m}^i = (0).$$

Proof. Indeed, the $\mathfrak{m}$-adic topology on $\bigcap \mathfrak{m}^i$ is the restriction of the $\mathfrak{m}$-adic topology of R on $\bigcap \mathfrak{m}^i$ by the Artin-Rees lemma (3.3.1). However, $\bigcap \mathfrak{m}^i$ is contained in every $\mathfrak{m}$-adic neighborhood of 0 in R; the induced topology on $\bigcap \mathfrak{m}^i$ is thus the indiscrete topology.

But to say that the $\mathfrak{m}$-adic topology on a module N is indiscrete is to say that $\mathfrak{m}N = N$, so $N = 0$ by Nakayama. The result is thus clear.

By similar logic, or by localizing at each maximal ideal, we find:

3.3.5 Corollary *If R is a commutative ring and I is contained in the Jacobson radical of R, then $\bigcap I^n = 0$.*

It turns out that the Krull intersection theorem can be proved in the following elementary manner, due to Perdry in **?**. The argument does not use the Artin-Rees lemma. One can prove:

3.3.6 Theorem (?) *Suppose R is a noetherian ring, $I \subset R$ an ideal. Suppose $b \in \bigcap I^n$. Then as ideals $(b) = (b)I$.*

In particular, it follows easily that $\bigcap I^n = 0$ under either of the following conditions:

1. I is contained in the Jacobson radical of R.
2. R is a domain and I is proper.

Proof. Let $a_1, \dots, a_k \in I$ be generators. For each n, the ideal I^n consists of the values of all homogeneous polynomials in $R[x_1, \dots, x_k]$ of degree n evaluated on the tuple $(a_1, \dots, a_k)$, as one may easily see.

It follows that if $b \in \bigcap I^n$, then for each n there is a polynomial $P_n \in R[x_1, \dots, x_k]$ which is homogeneous of degree n and which satisfies

$$P_n(a_1, \dots, a_k) = b.$$

The ideal generated by all the P_n in $R[x_1, \dots, x_k]$ is finitely generated by the Hilbert basis theorem. Thus there is N such that

$$P_N = Q_1 P_1 + Q_2 P_2 + \cdots + Q_{N-1} P_{N-1}$$

for some polynomials $Q_i \in R[x_1, \dots, x_k]$. By taking homogeneous components, we can assume moreover that Q_i is homogeneous of degree $N - i$ for each i. If we evaluate each at $(a_1, \dots, a_k)$ we find

$$b = b(Q_1(a_1, \dots, a_k) + \cdots + Q_{N-1}(a_1, \dots, a_k)).$$

But the $Q_i(a_1, \dots, a_k)$ lie in I as all the a_i do and Q_i is homogeneous of positive degree. Thus b equals b times something in I. $\square$

III.4. Integrality and valuation rings

The notion of integrality is familiar from number theory: it is similar to "algebraic" but with the polynomials involved are required to be monic. In algebraic geometry, integral extensions of rings correspond to correspondingly nice morphisms on the Spec's—when the extension is finitely generated, it turns out that the fibers are finite. That is, there are only finitely many ways to lift a prime ideal to the extension: if $A \to B$ is integral and finitely generated, then $\operatorname{Spec} B \to \operatorname{Spec} A$ has finite fibers.

Integral domains that are *integrally closed* in their quotient field will play an important role for us. Such "normal domains" are, for example, regular in codimension one, which means that the theory of Weil divisors (see section 5.2) applies to them. It is particularly nice because Weil divisors are sufficient to determine whether a function is regular on a normal variety.

A canonical example of an integrally closed ring is a valuation ring; we shall see in this chapter that any integrally closed ring is an intersection of such.

4.1. Integrality

Fundamentals

As stated in the introduction to the chapter, integrality is a condition on rings parallel to that of algebraicity for field extensions.

4.1.1 Definition Let R be a ring, and R' an R-algebra. An element $x \in R'$ is said to be **integral** over R if x satisfies a monic polynomial equation in $R[X]$, say

$$x^n + r_1 x^{n-1} + \cdots + r_n = 0, \quad r_1, \ldots, r_n \in R.$$

We can say that R' is **integral** over R if every $x \in R'$ is integral over R.

Note that in the definition, we are not requiring R to be a *subring* of R'.

4.1.2 Example $\frac{1+\sqrt{-3}}{2}$ is integral over $\mathbb{Z}$; it is in fact a sixth root of unity, thus satisfying the equation $X^6 - 1 = 0$. However, $\frac{1+\sqrt{5}}{2}$ is not integral over $\mathbb{Z}$. To explain this, however, we will need to work a bit more (see proposition 4.1.5 below).

4.1.3 Example Let L/K be a field extension. Then L/K is integral if and only if it is algebraic, since K is a field and we can divide polynomial equations by the leading coefficient to make them monic.

4.1.4 Example Let R be a graded ring. Then the subring $R^{(d)} \subset R$ was defined in definition 3.1.17; recall that this consists of elements of R all of whose nonzero homogeneous components live in degrees that are multiples of d. Then the dth power of any homogeneous element in R is in $R^{(d)}$. As a result, every homogeneous element of R is integral over $R^{(d)}$.

We shall now interpret the condition of integrality in terms of finite generation of certain modules. Suppose R is a ring, and R' an R-algebra. Let $x \in R'$.

4.1.5 Proposition *$x \in R'$ is integral over R if and only if the subalgebra $R[x] \subset R'$ (generated by R, x) is a finitely generated R-module.*

This notation is an abuse of notation (usually $R[x]$ refers to a polynomial ring), but it should not cause confusion.

This result for instance lets us show that $\frac{1+\sqrt{-5}}{2}$ is not integral over $\mathbb{Z}$, because when you keep taking powers, you get arbitrarily large denominators: the arbitrarily large denominators imply that it cannot be integral.

Proof. If $x \in R'$ is integral, then x satisfies

$$x^n + r_1 x^{n-1} + \cdots + r_n = 0, \quad r_i \in R.$$

Then $R[x]$ is generated as an R-module by $1, x, \ldots, x^{n-1}$. This is because the submodule of R' generated by $1, x, \ldots, x^{n-1}$ is closed under multiplication by R and by multiplication by x (by the above equation).

Now suppose x generates a subalgebra $R[x] \subset R'$ which is a finitely generated R-module. Then the increasing sequence of R-modules generated by $\{1\}, \{1, x\}, \{1, x, x^2\}, \ldots$ must stabilize, since the union is $R[x]$.[1] It follows that some x^n can be expressed as a linear combination of smaller powers of x. Thus x is integral over R. $\square$

So, if R' is an R-module, we can say that an element $x \in R'$ is **integral** over R if either of the following equivalent conditions are satisfied:

1. There is a monic polynomial in $R[X]$ which vanishes on x.
2. $R[x] \subset R'$ is a finitely generated R-module.

4.1.6 Example Let F be a field, V a finite-dimensional F-vector space, $T : V \to V$ a linear transformation. Then the ring generated by T and F inside $\mathrm{End}_F(V)$ (which is a noncommutative ring) is finite-dimensional over F. Thus, by similar reasoning, T must satisfy a polynomial equation with coefficients in F (e.g. the characteristic polynomial).

[1] As an easy exercise, one may see that if a finitely generated module M is the union of an increasing sequence of submodules $M_1 \subset M_2 \subset M_3 \subset \ldots$, then $M = M_n$ for some n; we just need to take n large enough such that M_n contains each of the finitely many generators of M.

Of course, if R' is integral over R, R' may not be a finitely generated R-module. For instance, $\overline{\mathbb{Q}}$ is not a finitely generated $\mathbb{Q}$-module, although the extension is integral. As we shall see in the next section, this is always the case if R' is a finitely generated R-*algebra*.

We now will add a third equivalent condition to this idea of "integrality," at least in the case where the structure map is an injection.

4.1.7 Proposition *Let R be a ring, and suppose R is a subring of R'. $x \in R'$ is integral if and only if there exists a finitely generated faithful R-module $M \subset R'$ such that $R \subset M$ and $xM \subset M$.*

A module M is *faithful* if $xM = 0$ implies $x = 0$. That is, the map from R into the $\mathbb{Z}$-endomorphisms of M is injective. If R is a *subring* of R' (i.e. the structure map $R \to R'$ is injective), then R' for instance is a faithful R-module.

Proof. It's obvious that the second condition above (equivalent to integrality) implies the condition of this proposition. Indeed, one could just take $M = R[x]$.

Now let us prove that if there exists such an M which is finitely generated, then x is integral. Just because M is finitely generated, the submodule $R[x]$ is not obviously finitely generated. In particular, this implication requires a bit of proof.

We shall prove that the condition of this proposition implies integrality. Suppose $y_1, \dots, y_k \in M$ generate M as R-module. Then multiplication by x gives an R-module map $M \to M$. In particular, we can write

$$xy_i = \sum a_{ij} y_j$$

where each $a_{ij} \in R$. These $\{a_{ij}\}$ may not be unique, but let us make some choices; we get a k-by-k matrix $A \in M_k(R)$. The claim is that x satisfies the characteristic polynomial of A.

Consider the matrix

$$(x1 - A) \in M_n(R').$$

Note that $(x1 - A)$ annihilates each y_i, by the choice of A. We can consider the adjoint $B = (x1 - A)^{adj}$. Then

$$B(x1 - A) = \det(x1 - A)1.$$

This product of matrices obviously annihilates each vector y_i. It follows that

$$(\det(x1 - A)y_i = 0, \quad \forall i,$$

which implies that $\det(x1 - A)$ kills M. This implies that $\det(x1 - A) = 0$ since M is faithful.

As a result, x satisfies the characteristic polynomial. ∎

4.1.8 Remark (exercise) Let R be a noetherian local domain with maximal ideal $\mathfrak{m}$. As we will define shortly, R is *integrally closed* if every element of the quotient field $K = K(R)$ integral over R belongs to R itself. Then if $x \in K$ and $x\mathfrak{m} \subset \mathfrak{m}$, we have $x \in R$.

4.1.9 Remark (exercise) Let us say that an A-module is *n-generated* if it is generated by at most n elements.

Let A and B be two rings such that $A \subset B$, so that B is an A-module.

Let $n \in \mathbb{N}$. Let $u \in B$. Then, the following four assertions are equivalent:

1. There exists a monic polynomial $P \in A[X]$ with $\deg P = n$ and $P(u) = 0$.
2. There exist a B-module C and an n-generated A-submodule U of C such that $uU \subset U$ and such that every $v \in B$ satisfying $vU = 0$ satisfies $v = 0$. (Here, C is an A-module, since C is a B-module and $A \subset B$.)
3. There exists an n-generated A-submodule U of B such that $1 \in U$ and $uU \subset U$.
4. As an A-module, $A[u]$ is spanned by $1, u, \ldots, u^{n-1}$.

We proved this to show that the set of integral elements is well behaved.

4.1.10 Proposition *Let $R \subset R'$. Let $S = \{x \in R' : x \text{ is integral over } R\}$. Then S is a subring of R'. In particular, it is closed under addition and multiplication.*

Proof. Suppose $x, y \in S$. We can consider the finitely generated modules $R[x], R[y] \subset R'$ generated (as algebras) by x over R. By assumption, these are finitely generated R-modules. In particular, the tensor product

$$R[x] \otimes_R R[y]$$

is a finitely generated R-module (by proposition 4.3.12).

We have a ring-homomorphism $R[x] \otimes_R R[y] \to R'$ which comes from the inclusions $R[x], R[y] \rightarrowtail R'$. Let M be the image of $R[x] \otimes_R R[y]$ in R'. Then M is an R-submodule of R', indeed an R-subalgebra containing x, y. Also, M is finitely generated. Since $x + y, xy \in M$ and M is a subalgebra, it follows that

$$(x + y)M \subset M, \quad xyM \subset M.$$

Thus $x + y, xy$ are integral over R. □

Let us consider the ring $\mathbb{Z}[\sqrt{-5}]$; this is the canonical example of a ring where unique factorization fails. This is because $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. One might ask: what about $\mathbb{Z}[\sqrt{-3}]$? It turns out that $\mathbb{Z}[\sqrt{-3}]$ lacks unique factorization as well. Indeed, here we have

$$(1 - \sqrt{-3})(1 + \sqrt{-3}) = 4 = 2 \times 2.$$

These elements can be factored no more, and $1 - \sqrt{-3}$ and 2 do not differ by units. So in this ring, we have a failure of unique factorization. Nonetheless, the failure of unique factorization in $\mathbb{Z}[\sqrt{-3}]$ is less noteworthy, because $\mathbb{Z}[\sqrt{-3}]$ is not *integrally closed*. Indeed, it turns out that $\mathbb{Z}[\sqrt{-3}]$ is contained in the larger ring $\mathbb{Z}\left[\frac{1+\sqrt{-3}}{2}\right]$, which does have unique factorization, and this larger ring is finite over $\mathbb{Z}[\sqrt{-3}]$.[2] Since being integrally closed is a prerequisite for having unique factorization (see **??** below), the failure in $\mathbb{Z}[\sqrt{-3}]$ is not particularly surprising.

Note that, by contrast, $\mathbb{Z}[\frac{1+\sqrt{-5}}{2}]$ does not contain $\mathbb{Z}[\sqrt{-5}]$ as a finite index subgroup—it cannot be slightly enlarged in the same sense. When one enlarges $\mathbb{Z}[\sqrt{-5}]$, one has to add a lot of stuff. We will see more formally that $\mathbb{Z}[\sqrt{-5}]$ is *integrally closed* in its quotient field, while $\mathbb{Z}[\sqrt{-3}]$ is not. Since unique factorization domains are automatically integrally closed, the failure of $\mathbb{Z}[\sqrt{-5}]$ to be a UFD is much more significant than that of $\mathbb{Z}[\sqrt{-3}]$.

Le sorite for integral extensions

In commutative algebra and algebraic geometry, there are a lot of standard properties that a *morphism* of rings $\phi : R \to S$ can have: it could be of *finite type* (that is, S is finitely generated over $\phi(R)$), it could be *finite* (that is, S is a finite R-module), or it could be *integral* (which we have defined in definition 4.1.1). There are many more examples that we will encounter as we dive deeper into commutative algebra. In algebraic geometry, there are corresponding properties of morphisms of *schemes,* and there are many more interesting ones here.

In these cases, there is usually—for any reasonable property—a standard and familiar list of properties that one proves about them. We will refer to such lists as "sorites," and prove our first one now.

4.1.11 Proposition (Le sorite for integral morphisms) *1. For any ring R and any ideal $I \subset R$, the map $R \to R/I$ is integral.*

2. If $\phi : R \to S$ and $\psi : S \to T$ are integral morphisms, then so is $\psi \circ \phi : R \to T$.

3. If $\phi : R \to S$ is an integral morphism and R' is an R-algebra, then the base-change $R' \to R' \otimes_R S$ is integral.

Proof. The first property is obvious. For the second, the condition of integrality in a morphism of rings depends on the inclusion of the image in the codomain. So we can suppose that $R \subset S \subset T$. Suppose $t \in T$. By assumption, there is a monic polynomial equation

$$t^n + s_1 t^{n-1} + \cdots + s_n = 0$$

that t satisfies, where each $s_i \in S$.

In particular, we find that t is integral over $R[s_1, \dots, s_n]$. As a result, the module $R[s_1, \dots, s_n, t]$ is finitely generated over the ring $R' = R[s_1, \dots, s_n]$. By the following proposition 4.1.12, R' is a finitely generated R-module. In particular, $R[s_1, \dots, s_n, t]$ is a finitely generated R-module (not just a finitely generated R'-module).

Thus the R-module $R[s_1, \dots, s_n, t]$ is a faithful R' module, finitely generated over R, which is preserved under multiplication by t. ∎

[2] In fact, $\mathbb{Z}[\sqrt{-3}]$ is an index two subgroup of $\mathbb{Z}\left[\frac{1+\sqrt{-3}}{2}\right]$, as the ring $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$ can be described as the set of elements $a + b\sqrt{-3}$ where a, b are either both integers or both integers plus $\frac{1}{2}$, as is easily seen: this set is closed under addition and multiplication.

We now prove a result that can equivalently be phrased as "finite type plus integral implies finite" for a map of rings.

4.1.12 Proposition *Let R' be a finitely generated, integral R-algebra. Then R' is a finitely generated R-module: that is, the map $R \to R'$ is finite.*

Proof. Induction on the number of generators of R' as R-algebra. For one generator, this follows from Proposition 4.1.5. In general, we will have $R' = R[\alpha_1, \dots, \alpha_n]$ for some $\alpha_i \in R'$. By the inductive hypothesis, $R[\alpha_1, \dots, \alpha_{n-1}]$ is a finite R-module; by the case of one generator, R' is a finite $R[\alpha_1, \dots, \alpha_{n-1}]$-module. This establishes the result by the next exercise. ∎

4.1.13 Remark (exercise) Let $R \to S, S \to T$ be morphisms of rings. Suppose S is a finite R-module and T a finite T-module. Then T is a finite R-module.

Integral closure

Let R, R' be rings.

4.1.14 Definition If $R \subset R'$, then the set $S = \{x \in R' : x \text{ is integral}\}$ is called the **integral closure** of R in R'. We say that R is **integrally closed in** R' if $S = R'$.

When R is a domain, and K is the quotient field, we shall simply say that R is **integrally closed** if it is integrally closed in K. Alternatively, some people say that R is **normal** in this case.

Integral closure (in, say, the latter sense) is thus an operation that maps integral domains to integral domains. It is easy to see that the operation is *idempotent:* the integral closure of the integral closure is the integral closure.

4.1.15 Example The integers $\mathbb{Z} \subset \mathbb{C}$ have as integral closure (in $\mathbb{C}$) the set of complex numbers satisfying a monic polynomial with integral coefficients. This set is called the set of **algebraic integers**.

For instance, i is an algebraic integer because it satisfies the equation $X^2 + 1 = 0$. $\frac{1-\sqrt{-3}}{2}$ is an algebraic integer, as we talked about last time; it is a sixth root of unity. On the other hand, $\frac{1+\sqrt{-5}}{2}$ is not an algebraic integer.

4.1.16 Example Take $\mathbb{Z} \subset \mathbb{Q}$. The claim is that $\mathbb{Z}$ is integrally closed in its quotient field $\mathbb{Q}$, or simply—integrally closed.

Proof. We will build on this proof later. Here is the point. Suppose $\frac{a}{b} \in \mathbb{Q}$ satisfying an equation

$$P(a/b) = 0, \quad P(t) = t^n + c_1 t^{n-1} + \cdots + c_0, \ \forall c_i \in \mathbb{Z}.$$

Assume that a, b have no common factors; we must prove that b has no prime factors, so is ± 1. If b had a prime factor, say q, then we must obtain a contradiction.

We interrupt with a definition.

4.1.17 Definition The **valuation at** q (or q**-adic valuation**) is the map $v_q : \mathbb{Q}^* \to \mathbb{Z}$ is the function sending $q^k(a/b)$ to k if $q \nmid a, b$. We extend this to all rational numbers via $v(0) = \infty$.

In general, this just counts the number of factors of q in the expression.

Note the general property that

$$v_q(x+y) \geqslant \min(v_q(x), v_q(y)). \tag{4.1.1}$$

If x, y are both divisible by some power of q, so is $x+y$; this is the statement above. We also have the useful property

$$v_q(xy) = v_q(x) + v_q(y). \tag{4.1.2}$$

Now return to the proof that $\mathbb{Z}$ is normal. We would like to show that $v_q(a/b) \geqslant 0$. This will prove that b is not divisible by q. When we show this for all q, it will follow that $a/b \in \mathbb{Z}$.

We are assuming that $P(a/b) = 0$. In particular,

$$\left(\frac{a}{b}\right)^n = -c_1\left(\frac{a}{b}\right)^{n-1} - \cdots - c_0.$$

Apply v_q to both sides:

$$nv_q(a/b) \geqslant \min_{i>0} v_q(c_i(a/b)^{n-i}).$$

Since the $c_i \in \mathbb{Z}$, their valuations are nonnegative. In particular, the right hand side is at least

$$\min_{i>0}(n-i)v_q(a/b).$$

This cannot happen if $v_q(a/b) < 0$, because $n-i < n$ for each $i > 0$. ∎

This argument applies more generally. If K is a field, and $R \subset K$ is a subring "defined by valuations," such as the v_q, then R is integrally closed in its quotient field. More precisely, note the reasoning of the previous example: the key idea was that $\mathbb{Z} \subset \mathbb{Q}$ was characterized by the rational numbers x such that $v_q(x) \geqslant 0$ for all primes q. We can abstract this idea as follows. If there exists a family of functions $\mathcal{V}$ from $K^* \to \mathbb{Z}$ (such as $\{v_q : \mathbb{Q}^* \to \mathbb{Z}\}$) satisfying (4.1.1) and (4.1.2) above such that R is the set of elements such that $v(x) \geqslant 0, v \in \mathcal{V}$ (along with 0), then R is integrally closed in K. We will talk more about this, and about valuation rings, below.

4.1.18 Example We saw earlier (example 4.1.2) that $\mathbb{Z}[\sqrt{-3}]$ is not integrally closed, as $\frac{1+\sqrt{-3}}{2}$ is integral over this ring and in the quotient field, but not in the ring.

We shall give more examples in the next subsec.

Geometric examples

Let us now describe the geometry of a non-integrally closed ring. Recall that finitely generated (reduced) $\mathbb{C}$-algebras are supposed to correspond to affine algebraic varieties. A *smooth* variety (i.e., one that is a complex manifold) will always correspond to an integrally closed ring (though this relies on a deep result that a regular local ring is a factorization domain, and consequently integrally closed): non-normality is a sign of singularities.

4.1.19 Example Here is a ring which is not integrally closed. Take $\mathbb{C}[x,y]/(x^2-y^3)$. Algebraically, this is the subring of the polynomial ring $\mathbb{C}[t]$ generated by t^2 and t^3.

In the complex plane, $\mathbb{C}^2$, this corresponds to the subvariety $C \subset \mathbb{C}^2$ defined by $x^2 = y^3$. In $\mathbb{R}^2$, this can be drawn: it has a singularity at $(x,y) = 0$.

Note that $x^2 = y^3$ if and only if there is a complex number z such that $x = z^3, y = z^2$. This complex number z can be recovered via x/y when $x, y \neq 0$. In particular, there is a map $\mathbb{C} \to C$ which sends $z \to (z^3, z^2)$. At every point other than the origin, the inverse can be recovered using rational functions. But this does not work at the origin.

We can think of $\mathbb{C}[x,y]/(x^2-y^3)$ as the subring R' of $\mathbb{C}[z]$ generated by $\{z^n, n \neq 1\}$. There is a map from $\mathbb{C}[x,y]/(x^2-y^3)$ sending $x \to z^3, y \to z^2$. Since these two domains are isomorphic, and R' is not integrally closed, it follows that $\mathbb{C}[x,y]/(x^2-y^3)$ is not integrally closed. The element z can be thought of as an element of the fraction field of R' or of $\mathbb{C}[x,y]/(x^2-y^3)$. It is integral, though.

The failure of the ring to be integrally closed has to do with the singularity at the origin.

We now give a generalization of the above example.

4.1.20 Example This example is outside the scope of the present course. Say that $X \subset \mathbb{C}^n$ is given as the zero locus of some holomorphic functions $\{f_i : \mathbb{C}^n \to \mathbb{C}\}$. We just gave an example when $n = 2$. Assume that $0 \in X$, i.e. each f_i vanishes at the origin.

Let R be the ring of germs of holomorphic functions 0, in other words holomorphic functions from small open neighborhoods of zero. Each of these f_i becomes an element of R. The ring $R/(\{f_i\})$ is called the ring of germs of holomorphic functions on X at zero.

Assume that R is a domain. This assumption, geometrically, means that near the point zero in X, X can't be broken into two smaller closed analytic pieces. The fraction field of R is to be thought of as the ring of germs of meromorphic functions on X at zero.

We state the following without proof:

4.1.21 Theorem *Let g/g' be an element of the fraction field, i.e. $g, g' \in R$. Then g/g' is integral over R if and only if g/g' is bounded near zero.*

In the previous example of X defined by $x^2 = y^3$, the function x/y (defined near the origin on the curve) is bounded near the origin, so it is integral over the ring of germs of regular functions. The reason it is not defined near the origin is *not* that it blows up. In fact, it extends continuously, but not holomorphically, to the rest of the variety X.

4.2. Lying over and going up

We now interpret integrality in terms of the geometry of Spec. In general, for $R \to S$ a ring-homomorphism, the induced map $\operatorname{Spec} S \to \operatorname{Spec} R$ need not be topologically nice; for instance, even if S is a finitely generated R-algebra, the image of $\operatorname{Spec} S$ in $\operatorname{Spec} R$ need not be either open or closed.[3]

We shall see that under conditions of integrality, more can be said.

Lying over

In general, given a morphism of algebraic varieties $f : X \to Y$, the image of a closed subset $Z \subset X$ is far from closed. For instance, a regular function $f : X \to \mathbb{C}$ that is a closed map would have to be either surjective or constant (if X is connected, say). Nonetheless, under integrality hypotheses, we can say more.

4.2.1 Proposition (Lying over) *If $\phi : R \to R'$ is an integral morphism, then the induced map*

$$\operatorname{Spec} R' \to \operatorname{Spec} R$$

is a closed map; it is surjective if ϕ is injective.

Another way to state the last claim, without mentioning $\operatorname{Spec} R'$, is the following. Assume ϕ is injective and integral. Then if $\mathfrak{p} \subset R$ is prime, then there exists $\mathfrak{q} \subset R'$ such that $\mathfrak{p}$ is the inverse image $\phi^{-1}(\mathfrak{q})$.

Proof. First suppose ϕ injective, in which case we must prove the map $\operatorname{Spec} R' \to \operatorname{Spec} R$ surjective. Let us reduce to the case of a local ring. For a prime $\mathfrak{p} \in \operatorname{Spec} R$, we must show that $\mathfrak{p}$ arises as the inverse image of an element of $\operatorname{Spec} R'$. So we replace R with $R_{\mathfrak{p}}$. We get a map

$$\phi_{\mathfrak{p}} : R_{\mathfrak{p}} \to (R - \mathfrak{p})^{-1} R'$$

which is injective if ϕ is, since localization is an exact functor. Here we have localized both R, R' at the multiplicative subset $R - \mathfrak{p}$.

Note that $\phi_{\mathfrak{p}}$ is an integral extension too. This follows because integrality is preserved by base-change. We will now prove the result for $\phi_{\mathfrak{p}}$; in particular, we will show that there is a prime ideal of $(R-\mathfrak{p})^{-1}R'$ that pulls back to $\mathfrak{p}R_{\mathfrak{p}}$. These will imply that if we pull this prime ideal back to R', it will pull back to $\mathfrak{p}$ in R. In detail, we can consider the diagram

$$\begin{tikzcd} \operatorname{Spec}(R-\mathfrak{p})^{-1}R' \arrow[r] \arrow[d] & \operatorname{Spec} R_{\mathfrak{p}} \arrow[d] \\ \operatorname{Spec} R' \arrow[r] & \operatorname{Spec} R \end{tikzcd}$$

[3] It is, however, true that if R is *noetherian* (see Chapter III.2) and S finitely generated over R, then the image of $\operatorname{Spec} S$ is *constructible,* that is, a finite union of locally closed subsets. **To be added: this result should be added sometime**.

which shows that if $\mathfrak{p}R_{\mathfrak{p}}$ appears in the image of the top map, then $\mathfrak{p}$ arises as the image of something in $\operatorname{Spec} R'$. So it is sufficient for the proposition (that is, the case of ϕ injective) to handle the case of R local, and $\mathfrak{p}$ the maximal ideal.

In other words, we need to show that:

> If R is a *local* ring, $\phi : R \hookrightarrow R'$ an injective integral morphism, then the maximal ideal of R is the inverse image of something in $\operatorname{Spec} R'$.

Assume R is local with maximal ideal $\mathfrak{p}$. We want to find a prime ideal $\mathfrak{q} \subset R'$ such that $\mathfrak{p} = \phi^{-1}(\mathfrak{q})$. Since $\mathfrak{p}$ is already maximal, it will suffice to show that $\mathfrak{p} \subset \phi^{-1}(\mathfrak{q})$. In particular, we need to show that there is a prime ideal $\mathfrak{q}$ such that $\mathfrak{p}R' \subset \mathfrak{q}$. The pull-back of this will be $\mathfrak{p}$.

If $\mathfrak{p}R' \neq R'$, then $\mathfrak{q}$ exists, since every proper ideal of a ring is contained in a maximal ideal. We will in fact show

$$\mathfrak{p}R' \neq R', \tag{4.2.1}$$

or that $\mathfrak{p}$ does not generate the unit ideal in R'. If we prove (4.2.1), we will thus be able to find our $\mathfrak{q}$, and we will be done.

Suppose the contrary, i.e. $\mathfrak{p}R' = R'$. We will derive a contradiction using Nakayama's lemma (lemma 4.1.22). Right now, we cannot apply Nakayama's lemma directly because R' is not a finite R-module. The idea is that we will "descend" the "evidence" that (4.2.1) fails to a small subalgebra of R', and then obtain a contradiction. To do this, note that $1 \in \mathfrak{p}R'$, and we can write

$$1 = \sum x_i \phi(y_i)$$

where $x_i \in R', y_i \in \mathfrak{p}$. This is the "evidence" that (4.2.1) fails, and it involves only a finite amount of data.

Let R'' be the subalgebra of R' generated by $\phi(R)$ and the x_i. Then $R'' \subset R'$ and is finitely generated as an R-algebra, because it is generated by the x_i. However, R'' is integral over R and thus finitely generated as an R-module, by proposition 4.1.12. This is where integrality comes in.

So R'' is a finitely generated R-module. Also, the expression $1 = \sum x_i \phi(y_i)$ shows that $\mathfrak{p}R'' = R''$. However, this contradicts Nakayama's lemma. That brings the contradiction, showing that $\mathfrak{p}$ cannot generate (1) in R', and proving the surjectivity part of lying over theorem.

Finally, we need to show that if $\phi : R \to R'$ is *any* integral morphism, then $\operatorname{Spec} R' \to \operatorname{Spec} R$ is a closed map. Let $X = V(I)$ be a closed subset of $\operatorname{Spec} R'$. Then the image of X in $\operatorname{Spec} R$ is the image of the map

$$\operatorname{Spec} R'/I \to \operatorname{Spec} R$$

obtained from the morphism $R \to R' \to R'/I$, which is integral; thus we are reduced to showing that any integral morphism ϕ has closed image on the Spec. Thus we are reduced to $X = \operatorname{Spec} R'$, if we throw out R' and replace it by R'/I.

In other words, we must prove the following statement. Let $\phi : R \to R'$ be an integral morphism; then the image of $\operatorname{Spec} R'$ in $\operatorname{Spec} R$ is closed. But, quotienting by $\ker \phi$ and taking the map $R/\ker \phi \to R'$, we may reduce to the case of ϕ injective; however, then this follows from the surjectivity result already proved. $\square$

In general, there will be *many* lifts of a given prime ideal. Consider for instance the inclusion $\mathbb{Z} \subset \mathbb{Z}[i]$. Then the prime ideal $(5) \in \operatorname{Spec} \mathbb{Z}$ can be lifted either to $(2+i) \in \operatorname{Spec} \mathbb{Z}[i]$ or $(2-i) \in \operatorname{Spec} \mathbb{Z}[i]$. These are distinct prime ideals: $\frac{2+i}{2-i} \notin \mathbb{Z}[i]$. But note that any element of $\mathbb{Z}$ divisible by $2+i$ is automatically divisible by its conjugate $2-i$, and consequently by their product 5 (because $\mathbb{Z}[i]$ is a UFD, being a euclidean domain).

Nonetheless, the different lifts are incomparable.

4.2.2 Proposition *Let $\phi : R \to R'$ be an integral morphism. Then given $\mathfrak{p} \in \operatorname{Spec} R$, there are no inclusions among the elements $\mathfrak{q} \in \operatorname{Spec} R'$ lifting $\mathfrak{p}$.*

In other words, if $\mathfrak{q}, \mathfrak{q}' \in \operatorname{Spec} R'$ lift $\mathfrak{p}$, then $\mathfrak{q} \not\subset \mathfrak{q}'$.

Proof. We will give a "slick" proof by various reductions. Note that the operations of localization and quotienting only shrink the Spec's: they do not "merge" heretofore distinct prime ideals into one. Thus, by quotienting R by $\mathfrak{p}$, we may assume R is a *domain* and that $\mathfrak{p} = 0$. Suppose we had two primes $\mathfrak{q} \subsetneq \mathfrak{q}'$ of R' lifting $(0) \in \operatorname{Spec} R$. Quotienting R' by $\mathfrak{q}$, we may assume that $\mathfrak{q} = 0$. We could even assume $R \subset R'$, by quotienting by the kernel of ϕ. The next lemma thus completes the proof, because it shows that $\mathfrak{q}' = 0$, contradiction. □

4.2.3 Lemma *Let $R \subset R'$ be an inclusion of integral domains, which is an integral morphism. If $\mathfrak{q} \in \operatorname{Spec} R'$ is a nonzero prime ideal, then $\mathfrak{q} \cap R$ is nonzero.*

Proof. Let $x \in \mathfrak{q}'$ be nonzero. There is an equation

$$x^n + r_1 x^{n-1} + \cdots + r_n = 0, \quad r_i \in R,$$
□

that x satisfies, by assumption. Here we can assume $r_n \neq 0$; then $r_n \in \mathfrak{q}' \cap R$ by inspection, though. So this intersection is nonzero.

4.2.4 Corollary *Let $R \subset R'$ be an inclusion of integral domains, such that R' is integral over R. Then if one of R, R' is a field, so is the other.*

Proof. Indeed, $\operatorname{Spec} R' \to \operatorname{Spec} R$ is surjective by proposition 4.2.1: so if $\operatorname{Spec} R'$ has one element (i.e., R' is a field), the same holds for $\operatorname{Spec} R$ (i.e., R is a field). Conversely, suppose R a field. Then any two prime ideals in $\operatorname{Spec} R'$ pull back to the same element of $\operatorname{Spec} R$. So, by proposition 4.2.2, there can be no inclusions among the prime ideals of $\operatorname{Spec} R'$. But R' is a domain, so it must then be a field. □

4.2.5 Remark (exercise) Let k be a field. Show that $k[\mathbb{Q}_{\geqslant 0}]$ is integral over the polynomial ring $k[T]$. Although this is a *huge* extension, the prime ideal (T) lifts in only one way to $\operatorname{Spec} k[\mathbb{Q}_{\geqslant 0}]$.

4.2.6 Remark (exercise) Suppose $A \subset B$ is an inclusion of rings over a field of characteristic p. Suppose $B^p \subset A$, so that B/A is integral in a very strong sense. Show that the map $\operatorname{Spec} B \to \operatorname{Spec} A$ is a *homeomorphism.*

Going up

Let $R \subset R'$ be an inclusion of rings with R' integral over R. We saw in the lying over theorem (proposition 4.2.1) that any prime $\mathfrak{p} \in \operatorname{Spec} R$ has a prime $\mathfrak{q} \in \operatorname{Spec} R'$ "lying over" $\mathfrak{p}$, i.e. such that $R \cap \mathfrak{q} = \mathfrak{p}$. We now want to show that we can lift finite *inclusions* of primes to R'.

4.2.7 Proposition (Going up) *Let $R \subset R'$ be an integral inclusion of rings. Suppose $\mathfrak{p}_1 \subset \mathfrak{p}_2 \subset \cdots \subset \mathfrak{p}_n \subset R$ is a finite ascending chain of prime ideals in R. Then there is an ascending chain $\mathfrak{q}_1 \subset \mathfrak{q}_2 \subset \cdots \subset \mathfrak{q}_n$ in $\operatorname{Spec} R'$ lifting this chain.*

Moreover, $\mathfrak{q}_1$ can be chosen arbitrarily so as to lift $\mathfrak{p}_1$.

Proof. By induction and lying over (proposition 4.2.1), it suffices to show:

> Let $\mathfrak{p}_1 \subset \mathfrak{p}_2$ be an inclusion of primes in $\operatorname{Spec} R$. Let $\mathfrak{q}_1 \in \operatorname{Spec} R'$ lift $\mathfrak{p}_1$. Then there is $\mathfrak{q}_2 \in \operatorname{Spec} R'$, which satisfies the dual conditions of lifting $\mathfrak{p}_2$ and containing $\mathfrak{q}_1$.

To show that this is true, we apply proposition 4.2.1 to the inclusion $R/\mathfrak{p}_1 \hookrightarrow R'/\mathfrak{q}_1$. There is an element of $\operatorname{Spec} R'/\mathfrak{q}_1$ lifting $\mathfrak{p}_2/\mathfrak{p}_1$; the corresponding element of $\operatorname{Spec} R'$ will do for $\mathfrak{q}_2$. ∎

4.3. Valuation rings

A valuation ring is a special type of local ring. Its distinguishing characteristic is that divisibility is a "total preorder." That is, two elements of the quotient field are never incompatible under divisibility. We shall see in this section that integrality can be detected using valuation rings only.

Geometrically, the valuation ring is something like a local piece of a smooth curve. In fact, in algebraic geometry, a more compelling reason to study valuation rings is provided by the valuative criteria for separatedness and properness (cf. **?** or **?**). One key observation about valuation rings that leads the last results is that any local domain can be "dominated" by a valuation ring with the same quotient field (i.e. mapped into a valuation ring via local homomorphism), but valuation rings are the maximal elements in this relation of domination.

Definition

4.3.1 Definition A **valuation ring** is a domain R such that for every pair of elements $a, b \in R$, either $a \mid b$ or $b \mid a$.

4.3.2 Example $\mathbb{Z}$ is not a valuation ring. It is neither true that 2 divides 3 nor that 3 divides 2.

4.3.3 Example $\mathbb{Z}_{(p)}$, which is the set of all fractions of the form $a/b \in \mathbb{Q}$ where $p \nmid b$, is a valuation ring. To check whether a/b divides a'/b' or vice versa, one just has to check which is divisible by the larger power of p.

4.3.4 Proposition *Let R be a domain with quotient field K. Then R is a valuation ring if and only if for every $x \in K$, either x or x^{-1} lies in R.*

Proof. Indeed, if $x = a/b$, $a, b \in R$, then either $a \mid b$ or $b \mid a$, so either x or $x^{-1} \in R$. This condition is equivalent to R's being a valuation ring. $\square$

Valuations

The reason for the name "valuation ring" is provided by the next definition. As we shall see, any valuation ring comes from a "valuation."

By definition, an *ordered abelian group* is an abelian group A together with a set of *positive elements* $A_+ \subset A$. This set is required to be closed under addition and satisfy the property that if $x \in A$, then precisely one of the following is true: $x \in A_+$, $-x \in A_+$, and $x = 0$. This allows one to define an ordering $<$ on A by writing $x < y$ if $y - x \in A_+$. Given A, we often formally adjoin an element ∞ which is bigger than every element in A.

4.3.5 Definition Let K be a field. A **valuation** on K is a map $v : K \to A \cup \{\infty\}$ for some ordered abelian group A satisfying:

1. $v(0) = \infty$ and $v(K^*) \subset A$.
2. For $x, y \in K^*$, $v(xy) = v(x) + v(y)$. That is, $v|_{K^*}$ is a homomorphism.
3. For $x, y \in K$, $v(x + y) \geqslant \min(v(x), v(y))$.

Suppose that K is a field and $v : K \to A \cup \{\infty\}$ is a valuation (i.e. $v(0) = \infty$). Define $R = \{x \in K : v(x) \geqslant 0\}$.

4.3.6 Proposition *R as just defined is a valuation ring.*

Proof. First, we prove that R is a ring. R is closed under addition and multiplication by the two conditions

$$v(xy) = v(x) + v(y)$$

and

$$v(x + y) \geqslant \min v(x), v(y),$$

so if $x, y \in R$, then $x + y, xy$ have nonnegative valuations.

Note that $0 \in R$ because $v(0) = \infty$. Also $v(1) = 0$ since $v : K^* \to A$ is a homomorphism. So $1 \in R$ too. Finally, $-1 \in R$ because $v(-1) = 0$ since A is totally ordered. It follows that R is also a group.

Let us now show that R is a valuation ring. If $x \in K^*$, either $v(x) \geqslant 0$ or $v(x^{-1}) \geqslant 0$ since A is totally ordered.[4] So either $x, x^{-1} \in R$. $\square$

In particular, the set of elements with nonnegative valuation is a valuation ring. The converse also holds. Whenever you have a valuation ring, it comes about in this manner.

[4]Otherwise $0 = v(x) + v(x^{-1}) < 0$, contradiction.

4.3.7 Proposition *Let R be a valuation ring with quotient field K. There is an ordered abelian group A and a valuation $v : K^* \to A$ such that R is the set of elements with nonnegative valuation.*

Proof. First, we construct A. In fact, it is the quotient of K^* by the subgroup of units R^* of R. We define an ordering by saying that $x \leqslant y$ if $y/x \in R$—this doesn't depend on the representatives in K^* chosen. Note that either $x \leqslant y$ or $y \leqslant x$ must hold, since R is a valuation ring. The combination of $x \leqslant y$ and $y \leqslant x$ implies that x, y are equivalent classes. The nonnegative elements in this group are those whose representatives in K^* belong to R.

It is easy to see that K^*/R^* in this way is a totally ordered abelian group with the image of 1 as the unit. The reduction map $K^* \to K^*/R^*$ defines a valuation whose corresponding ring is just R. We have omitted some details; for instance, it should be checked that the valuation of $x + y$ is at least the minimum of $v(x), v(y)$. $\square$

To summarize:

> Every valuation ring R determines a valuation v from the fraction field of R into $A \cup \{\infty\}$ for A a totally ordered abelian group such that R is just the set of elements of K with nonnegative valuation. As long as we require that $v : K^* \to A$ is surjective, then A is uniquely determined as well.

4.3.8 Definition A valuation ring R is **discrete** if we can choose A to be $\mathbb{Z}$.

4.3.9 Example $\mathbb{Z}_{(p)}$ is a discrete valuation ring.

The notion of a valuation ring is a useful one.

General remarks

Let R be a commutative ring. Then $\operatorname{Spec} R$ is the set of primes of R, equipped with a certain topology. The space $\operatorname{Spec} R$ is almost never Hausdorff. It is almost always a bad idea to apply the familiar ideas from elementary topology (e.g. the fundamental group) to $\operatorname{Spec} R$. Nonetheless, it has some other nice features that substitute for its non-Hausdorffness.

For instance, if $R = \mathbb{C}[x, y]$, then $\operatorname{Spec} R$ corresponds to $\mathbb{C}^2$ with some additional nonclosed points. The injection of $\mathbb{C}^2$ with its usual topology into $\operatorname{Spec} R$ is continuous. While in $\operatorname{Spec} R$ you don't want to think of continuous paths, you can in $\mathbb{C}^2$.

Suppose you had two points $x, y \in \mathbb{C}^2$ and their images in $\operatorname{Spec} R$. Algebraically, you can still think about algebraic curves passing through x, y. This is a subset of x, y defined by a single polynomial equation. This curve will have what's called a "generic point," since the ideal generated by this curve will be a prime ideal. The closure of this generic point will be precisely this algebraic curve—including x, y.

4.3.10 Remark If $\mathfrak{p}, \mathfrak{p}' \in \operatorname{Spec} R$, then

$$\mathfrak{p}' \in \overline{\{\mathfrak{p}\}}$$

iff

$$\mathfrak{p}' \supset \mathfrak{p}.$$

Why is this? Well, the closure of $\{\mathfrak{p}\}$ is just $V(\mathfrak{p})$, since this is the smallest closed subset of $\operatorname{Spec} R$ containing $\mathfrak{p}$.

The point of this discussion is that instead of paths, one can transmit information from point to point in $\operatorname{Spec} R$ by having one point be in a closure of another. However, we will show that this relation is contained by the theory of valuation rings.

4.3.11 Theorem *Let R be a domain containing a prime ideal $\mathfrak{p}$. Let K be the fraction field of R.*

Then there is a valuation v on K defining a valuation ring $R' \subset K$ such that

1. $R \subset R'$.
2. $\mathfrak{p} = \{x \in R : v(x) > 0\}$.

Let us motivate this by the remark:

4.3.12 Remark A valuation ring is automatically a local ring. A local ring is a ring where either $x, 1-x$ is invertible for all x in the ring. Let us show that this is true for a valuation ring.

If x belongs to a valuation ring R with valuation v, it is invertible if $v(x) = 0$. So if $x, 1-x$ were both noninvertible, then both would have positive valuation. However, that would imply that $v(1) \geqslant \min v(x), v(1-x)$ is positive, contradiction.

> If R' is any valuation ring (say defined by a valuation v), then R' is local with maximal ideal consisting of elements with positive valuation.

The theorem above says that there's a good supply of valuation rings. In particular, if R is any domain, $\mathfrak{p} \subset R$ a prime ideal, then we can choose a valuation ring $R' \supset R$ such that $\mathfrak{p}$ is the intersection of the maximal ideal of R' intersected with R. So the map $\operatorname{Spec} R' \to \operatorname{Spec} R$ contains $\mathfrak{p}$.

Proof. Without loss of generality, replace R by $R_{\mathfrak{p}}$, which is a local ring with maximal ideal $\mathfrak{p}R_{\mathfrak{p}}$. The maximal ideal intersects R only in $\mathfrak{p}$.

So, we can assume without loss of generality that

1. R is local.
2. $\mathfrak{p}$ is maximal.

Let P be the collection of all subrings $R' \subset K$ such that $R' \supset R$ but $\mathfrak{p}R' \neq R'$. Then P is a poset under inclusion. The poset is nonempty, since $R \in P$. Every totally ordered chain in P has an upper bound. If you have a totally ordered subring of elements in P, then you can take the union. We invoke:

4.3.13 Lemma *Let R_α be a chain in P and $R' = \bigcup R_\alpha$. Then $R' \in P$.*

Proof. Indeed, it is easy to see that this is a subalgebra of K containing R. The thing to observe is that

$$\mathfrak{p}R' = \bigcup_\alpha \mathfrak{p}R_\alpha;$$

since by assumption, $1 \notin \mathfrak{p}R_\alpha$ (because each $R_\alpha \in P$), $1 \notin \mathfrak{p}R'$. In particular, $R' \notin P$. □

By the lemma, Zorn's lemma to the poset P. In particular, P has a maximal element R'. By construction, R' is some subalgebra of K and $\mathfrak{p}R' \neq R'$. Also, R' is maximal with respect to these properties.

We show first that R' is local, with maximal ideal $\mathfrak{m}$ satisfying

$$\mathfrak{m} \cap R = \mathfrak{p}.$$ □

The second part is evident from locality of R', since $\mathfrak{m}$ must contain the proper ideal $\mathfrak{p}R'$, and $\mathfrak{p} \subset R$ is a maximal ideal.

Suppose that $x \in R'$; we show that either $x, 1-x$ belongs to R'^* (i.e. is invertible). Take the ring $R'[x^{-1}]$. If x is noninvertible, this properly contains R'. By maximality, it follows that $\mathfrak{p}R'[x^{-1}] = R'[x^{-1}]$.

And we're out of time. We'll pick this up on Monday.

Let us set a goal.

First, recall the notion introduced last time. A **valuation ring** is a domain R where for all x in the fraction field of R, either x or x^{-1} lies in R. We saw that if R is a valuation ring, then R is local. That is, there is a unique maximal ideal $\mathfrak{m} \subset R$, automatically prime. Moreover, the zero ideal (0) is prime, as R is a domain. So if you look at the spectrum $\operatorname{Spec} R$ of a valuation ring R, there is a unique closed point $\mathfrak{m}$, and a unique generic point (0). There might be some other prime ideals in $\operatorname{Spec} R$; this depends on where the additional valuation lives.

4.3.14 Example Suppose the valuation defining the valuation ring R takes values in $\mathbb{R}$. Then the only primes are $\mathfrak{m}$ and zero.

Let R now be any ring, with $\operatorname{Spec} R$ containing prime ideals $\mathfrak{p} \subset \mathfrak{q}$. In particular, $\mathfrak{q}$ lies in the closure of $\mathfrak{p}$. As we will see, this implies that there is a map

$$\phi : R \to R'$$

such that $\mathfrak{p} = \phi^{-1}(0)$ and $\mathfrak{q} = \phi^{-1}(\mathfrak{m})$, where $\mathfrak{m}$ is the maximal ideal of R'. This statement says that the relation of closure in $\operatorname{Spec} R$ is always controlled by valuation rings. In yet another phrasing, in the map

$$\operatorname{Spec} R' \to \operatorname{Spec} R$$

the closed point goes to $\mathfrak{q}$ and the generic point to $\mathfrak{p}$. This is our eventual goal.

To carry out this goal, we need some more elementary facts. Let us discuss things that don't have any obvious relation to it.

Back to the goal

Now we return to the goal of the lecture. Again, R was any ring, and we had primes $\mathfrak{p} \subset \mathfrak{q} \subset R$. We wanted a valuation ring R' and a map $\phi : R \to R'$ such that zero pulled back to $\mathfrak{p}$ and the maximal ideal pulled back to $\mathfrak{q}$.

What does it mean for $\mathfrak{p}$ to be the inverse image of $(0) \subset R'$? This means that $\mathfrak{p} = \ker \phi$. So we get an injection

$$R/\mathfrak{p} \rightarrowtail R'.$$

We will let R' be a subring of the quotient field K of the domain $R/\mathfrak{p}$. Of course, this subring will contain $R/\mathfrak{p}$.

In this case, we will get a map $R \to R'$ such that the pull-back of zero is $\mathfrak{p}$. What we want, further, to be true is that R' is a valuation ring and the pull-back of the maximal ideal is $\mathfrak{q}$.

This is starting to look at the problem we discussed last time. Namely, let's throw out R, and replace it with $R/\mathfrak{p}$. Moreover, we can replace R with $R_{\mathfrak{q}}$ and assume that R is local with maximal ideal $\mathfrak{q}$. What we need to show is that a valuation ring R' contained in the fraction field of R, containing R, such that the intersection of the maximal ideal of R' with R is equal to $\mathfrak{q} \subset R$. If we do this, then we will have accomplished our goal.

4.3.15 Lemma *Let R be a local domain. Then there is a valuation subring R' of the quotient field of R that* dominates *R, i.e .the map $R \to R'$ is a* local *homomorphism.*

Let's find R' now.

Choose R' maximal such that $\mathfrak{q}R' \neq R'$. Such a ring exists, by Zorn's lemma. We gave this argument at the end last time.

4.3.16 Lemma R' *as described is local.*

Proof. Look at $\mathfrak{q}R' \subset R'$; it is a proper subset, too, by assumption. In particular, $\mathfrak{q}R'$ is contained in some maximal ideal $\mathfrak{m} \subset R'$. Replace R' by $R'' = R'_{\mathfrak{m}}$. Note that

$$R' \subset R''$$

and

$$\mathfrak{q}R'' \neq R''$$

because $\mathfrak{m}R'' \neq R''$. But R' is maximal, so $R' = R''$, and R'' is a local ring. So R' is a local ring. ∎

Let $\mathfrak{m}$ be the maximal ideal of R'. Then $\mathfrak{m} \supset \mathfrak{q}R$, so $\mathfrak{m} \cap R = \mathfrak{q}$. All that is left to prove now is that R' is a valuation ring.

4.3.17 Lemma R' *is integrally closed.*

Proof. Let R'' be its integral closure. Then $\mathfrak{m}R'' \neq R''$ by lying over, since $\mathfrak{m}$ (the maximal ideal of R') lifts up to R''. So R'' satisfies

$$\mathfrak{q}R'' \neq R''$$

and by maximality, we have $R'' = R'$. ∎

To summarize, we know that R' is a local, integrally closed subring of the quotient field of R, such that the maximal ideal of R' pulls back to $\mathfrak{q}$ in R. All we now need is:

4.3.18 Lemma *R' is a valuation ring.*

Proof. Let x lie in the fraction field. We must show that either x or $x^{-1} \in R'$. Say $x \notin R'$. This means by maximality of R' that $R'' = R'[x]$ satisfies

$$\mathfrak{q}R'' = R''.$$

In particular, we can write

$$1 = \sum q_i x^i, \quad q_i \in \mathfrak{q}R' \subset R'.$$

This implies that

$$(1 - q_0) + \sum_{i>0} -q_i x^i = 0.$$

But $1 - q_0$ is invertible in R', since R' is local. We can divide by the highest power of x:

$$x^{-N} + \sum_{i>0} \frac{-q_i}{1 - q_0} x^{-N+i} = 0.$$

In particular, $1/x$ is integral over R'; this implies that $1/x \in R'$ since R' is integrally closed and q_0 is a nonunit. So R' is a valuation ring. ∎

We can state the result formally.

4.3.19 Theorem *Let R be a ring, $\mathfrak{p} \subset \mathfrak{q}$ prime ideals. Then there is a homomorphism $\phi : R \to R'$ into a valuation ring R' with maximal ideal $\mathfrak{m}$ such that*

$$\phi^{-1}(0) = \mathfrak{p}$$

and

$$\phi^{-1}(\mathfrak{m}) = \mathfrak{q}.$$

There is a related fact which we now state.

4.3.20 Theorem *Let R be any domain. Then the integral closure of R in the quotient field K is the intersection*

$$\bigcap R_\alpha$$

of all valuation rings $R_\alpha \subset K$ containing R.

So an element of the quotient field is integral over R if and only if its valuation is nonnegative at every valuation which is nonnegative on R.

Proof. The $\subset$ argument is easy, because one can check that a valuation ring is integrally closed. (Exercise.) The interesting direction is to assume that $v(x) \geqslant 0$ for all v nonnegative on R.

Let us suppose x is nonintegral. Suppose $R' = R[1/x]$ and I be the ideal $(x^{-1}) \subset R'$. There are two cases:

1. $I = R'$. Then in the ring R', x^{-1} is invertible. In particular, $x^{-1}P(x^{-1}) = 1$. Multiplying by a high power of x shows that x is integral over R. Contradiction.
2. Suppose $I \subsetneq R'$. Then I is contained in a maximal ideal $\mathfrak{q} \subset R'$. There is a valuation subring $R'' \subset K$, containing R', such that the corresponding valuation is positive on $\mathfrak{q}$. In particular, this valuation is positive on x^{-1}, so it is negative on x, contradiction. $\square$

So the integral closure has this nice characterization via valuation rings. In some sense, the proof that $\mathbb{Z}$ is integrally closed has the property that every integrally closed ring is integrally closed for that reason: it's the common nonnegative locus for some valuations.

4.4. The Hilbert Nullstellensatz

The Nullstellensatz is the basic algebraic fact, which we have invoked in the past to justify various examples, that connects the idea of the Spec of a ring to classical algebraic geometry.

Statement and initial proof of the Nullstellensatz

There are several ways in which the Nullstellensatz can be stated. Let us start with the following very concrete version.

4.4.1 Theorem *All maximal ideals in the polynomial ring $R = \mathbb{C}[x_1, \dots, x_n]$ come from points in $\mathbb{C}^n$. In other words, if $\mathfrak{m} \subset R$ is maximal, then there exist $a_1, \dots, a_n \in \mathbb{C}$ such that $\mathfrak{m} = (x_1 - a_1, \dots, x_n - a_n)$.*

The maximal spectrum of $R = \mathbb{C}[x_1, \dots, x_n]$ is thus identified with $\mathbb{C}^n$.

We shall now reduce Theorem 4.4.1 to an easier claim. Let $\mathfrak{m} \subset R$ be a maximal ideal. Then there is a map

$$\mathbb{C} \to R \to R/\mathfrak{m}$$

where $R/\mathfrak{m}$ is thus a finitely generated $\mathbb{C}$-algebra, as R is. The ring $R/\mathfrak{m}$ is also a field by maximality.

We would like to show that $R/\mathfrak{m}$ is a finitely generated $\mathbb{C}$-vector space. This would imply that $R/\mathfrak{m}$ is integral over $\mathbb{C}$, and there are no proper algebraic extensions of $\mathbb{C}$. Thus, if we prove this, it will follow that the map $\mathbb{C} \to R/\mathfrak{m}$ is an isomorphism. If $a_i \in \mathbb{C}$ $(1 \leqslant i \leqslant n)$ is the image of x_i in $R/\mathfrak{m} = \mathbb{C}$, it will follow that $(x_1 - a_1, \dots, x_n - a_n) \subset \mathfrak{m}$, so $(x_1 - a_1, \dots, x_n - a_n) = \mathfrak{m}$.

Consequently, the Nullstellensatz in this form would follow from the next claim:

4.4.2 Proposition *Let k be a field, L/k an extension of fields. Suppose L is a finitely generated k-algebra. Then L is a finite k-vector space.*

This is what we will prove.

We start with an easy proof in the special case:

4.4.3 Lemma *Assume k is uncountable (e.g. $\mathbb{C}$, the original case of interest). Then the above proposition is true.*

Proof. Since L is a finitely generated k-algebra, it suffices to show that L/k is algebraic. If not, there exists $x \in L$ which isn't algebraic over k. So x satisfies no nontrivial polynomials. I claim now that the uncountably many elements $\frac{1}{x-\lambda}, \lambda \in K$ are linearly independent over K. This will be a contradiction as L is a finitely generated k-algebra, hence at most countably dimensional over k. (Note that the polynomial ring is countably dimensional over k, and L is a quotient.)

So let's prove this. Suppose not. Then there is a nontrivial linear dependence

$$\sum \frac{c_i}{x-\lambda_i} = 0, \quad c_i, \lambda_i \in K.$$

Here the λ_j are all distinct to make this nontrivial. Clearing denominators, we find

$$\sum_i c_i \prod_{j \neq i} (x - \lambda_j) = 0.$$

Without loss of generality, $c_1 \neq 0$. This equality was in the field L. But x is transcendental over k. So we can think of this as a polynomial ring relation. Since we can think of this as a relation in the polynomial ring, we see that doing so, all but the $i = 1$ term in the sum is divisible by $x - \lambda_1$ as a polynomial. It follows that, as polynomials in the indeterminate x,

$$x - \lambda_1 \mid c_1 \prod_{j \neq 1} (x - \lambda_j).$$

This is a contradiction since all the λ_i are distinct. ∎

This is kind of a strange proof, as it exploits the fact that $\mathbb{C}$ is uncountable. This shouldn't be relevant.

The normalization lemma

Let's now give a more algebraic proof. We shall exploit the following highly useful fact in commutative algebra:

4.4.4 Theorem (Noether normalization lemma) *Let k be a field, and $R = k[x_1, \dots, x_n]/\mathfrak{p}$ be a finitely generated domain over k (where $\mathfrak{p}$ is a prime ideal in the polynomial ring).*

Then there exists a polynomial subalgebra $k[y_1, \dots, y_m] \subset R$ such that R is integral over $k[y_1, \dots, y_m]$.

Later we will see that m is the *dimension* of R.

There is a geometric picture here. Then $\operatorname{Spec} R$ is some irreducible algebraic variety in k^n (plus some additional points), with a smaller dimension than n if $\mathfrak{p} \neq 0$. Then there exists a *finite map* to k^m. In particular, we can map surjectively $\operatorname{Spec} R \to k^m$ which is integral. The fibers are in fact finite, because integrality implies finite fibers. (We have not actually proved this yet.)

How do we actually find such a finite projection? In fact, in characteristic zero, we just take a vector space projection $\mathbb{C}^n \to \mathbb{C}^m$. For a "generic" projection onto a subspace of the appropriate dimension, the projection will will do as our finite map. In characteristic p, this may not work.

Proof. First, note that m is uniquely determined as the transcendence degree of the quotient field of R over k.

Among the variables $x_1, \dots, x_n \in R$ (which we think of as in R by an abuse of notation), choose a maximal subset which is algebraically independent. This subset has no nontrivial polynomial relations. In particular, the ring generated by that subset is just the polynomial ring on that subset. We can permute these variables and assume that

$$\{x_1, \dots, x_m\}$$

is the maximal subset. In particular, R contains the *polynomial ring* $k[x_1, \dots, x_m]$ and is generated by the rest of the variables. The rest of the variables are not adjoined freely though.

The strategy is as follows. We will implement finitely many changes of variable so that R becomes integral over $k[x_1, \dots, x_m]$.

The essential case is where $m = n - 1$. Let us handle this. So we have

$$R_0 = k[x_1, \dots, x_m] \subset R = R_0[x_n]/\mathfrak{p}.$$

Since x_n is not algebraically independent, there is a nonzero polynomial $f(x_1, \dots, x_m, x_n) \in \mathfrak{p}$.

We want f to be monic in x_n. This will buy us integrality. A priori, this might not be true. We will modify the coordinate system to arrange that, though. Choose $N \gg 0$. Define for $1 \leqslant i \leqslant m$,

$$x_i' = x_i + x_n^{N^i}.$$

Then the equation becomes:

$$0 = f(x_1, \dots, x_m, x_n) = f(\left\{x_i' - x_n^{N^i}\right\}, x_n).$$

Now $f(x_1, \dots, x_n, x_{n+1})$ looks like some sum

$$\sum \lambda_{a_1 \dots b} x_1^{a_1} \dots x_m^{a_m} x_n^b, \quad \lambda_{a_1 \dots b} \in k.$$

But N is really really big. Let us expand this expression in the x_i' and pay attention to the largest power of x_n we see. We find that

$$f(\{x_i' - x_n^{N_i}\}, x_n)$$

has the largest power of x_n precisely where, in the expression for f, a_m is maximized first, then a_{m-1}, and so on. The largest exponent would have the form

$$x_n^{a_m N^m + a_{m-1}N^{m-1}+\cdots+b}.$$

We can't, however, get any exponents of x_n in the expression $f(\{x_i' - x_n^{N_i}\}, x_n)$ other than these. If N is super large, then all these exponents will be different from each other. In particular, each power of x_n appears precisely once in the expansion of f. We see in particular that x_n is integral over $x_1', \dots, x_n'$. Thus each x_i is as well.

So we find

R is integral over $k[x_1', \dots, x_m']$.

We have thus proved the normalization lemma in the codimension one case. What about the general case? We repeat this. Say we have

$$k[x_1, \dots, x_m] \subset R.$$

Let R' be the subring of R generated by $x_1, \dots, x_m, x_{m+1}$. The argument we just gave implies that we can choose $x_1', \dots, x_m'$ such that R' is integral over $k[x_1', \dots, x_m']$, and the x_i' are algebraically independent. We know in fact that $R' = k[x_1', \dots, x_m', x_{m+1}]$.

Let us try repeating the argument while thinking about x_{m+2}. Let $R'' = k[x_1', \dots, x_m', x_{m+2}]$ modulo whatever relations that x_{m+2} has to satisfy. So this is a subring of R. The same argument shows that we can change variables such that $x_1'', \dots, x_m''$ are algebraically independent and R'' is integral over $k[x_1'', \dots, x_m'']$. We have furthermore that $k[x_1'', \dots, x_m'', x_{m+2}] = R''$.

Having done this, let us give the argument where $m = n - 2$. You will then see how to do the general case. Then I claim that:

R is integral over $k[x_1'', \dots, x_m'']$.

For this, we need to check that x_{m+1}, x_{m+2} are integral (because these together with the x_i'' generate $R''[x_{m+2}][x_{m+2}] = R$. But x_{m+2} is integral over this by construction. The integral closure of $k[x_1'', \dots, x_m'']$ in R thus contains

$$k[x_1'', \dots, x_m'', x_{m+2}] = R''. \qquad \square$$

However, R'' contains the elements $x_1', \dots, x_m'$. But by construction, x_{m+1} is integral over the $x_1', \dots, x_m'$. The integral closure of $k[x_1'', \dots, x_m'']$ must contain x_{m+2}. This completes the proof in the case $m = n - 2$. The general case is similar; we just make several changes of variables, successively.

Back to the Nullstellensatz

Consider a finitely generated k-algebra R which is a field. We need to show that R is a finite k-module. This will prove the proposition. Well, note that R is integral over a polynomial ring $k[x_1, \dots, x_m]$ for some m. If $m > 0$, then this polynomial ring has more than one prime. For instance, (0) and $(x_1, \dots, x_m)$. But these must lift to primes in R. Indeed, we have seen that whenever you have an integral extension, the induced map on spectra is surjective. So

$$\operatorname{Spec} R \to \operatorname{Spec} k[x_1, \dots, x_m]$$

is surjective. If R is a field, this means $\operatorname{Spec} k[x_1, \dots, x_m]$ has one point and $m = 0$. So R is integral over k, thus algebraic. This implies that R is finite as it is finitely generated. This proves one version of the Nullstellensatz.

Another version of the Nullstellensatz, which is more precise, says:

4.4.5 Theorem *Let $I \subset \mathbb{C}[x_1, \dots, x_n]$. Let $V \subset \mathbb{C}^n$ be the subset of $\mathbb{C}^n$ defined by the ideal I (i.e. the zero locus of I).*

Then $\operatorname{Rad}(I)$ *is precisely the collection of f such that $f|_V = 0$. In particular,*

$$\operatorname{Rad}(I) = \bigcap_{\mathfrak{m} \supset I, \mathfrak{m} \text{ maximal}} \mathfrak{m}.$$

In particular, there is a bijection between radical ideals and algebraic subsets of $\mathbb{C}^n$.

The last form of the theorem, which follows from the expression of maximal ideals in the polynomial ring, is very similar to the result

$$\operatorname{Rad}(I) = \bigcap_{\mathfrak{p} \supset I, \mathfrak{p} \text{ prime}} \mathfrak{p},$$

true in any commutative ring. However, this general result is not necessarily true.

4.4.6 Example The intersection of all primes in a DVR is zero, but the intersection of all maximals is nonzero.

Proof of theorem 4.4.5. It now suffices to show that for every $\mathfrak{p} \subset \mathbb{C}[x_1, \dots, x_n]$ prime, we have

$$\mathfrak{p} = \bigcap_{\mathfrak{m} \supset I \text{ maximal}} \mathfrak{m}$$

since every radical ideal is an intersection of primes.

Let $R = \mathbb{C}[x_1, \dots, x_n]/\mathfrak{p}$. This is a domain finitely generated over $\mathbb{C}$. We want to show that the intersection of maximal ideals in R is zero. This is equivalent to the above displayed equality.

So fix $f \in R - \{0\}$. Let R' be the localization $R' = R_f$. Then R' is also an integral domain, finitely generated over $\mathbb{C}$. R' has a maximal ideal $\mathfrak{m}$ (which a priori could be zero). If we look at the map $R' \to R'/\mathfrak{m}$, we get a map into a field finitely generated over $\mathbb{C}$, which is thus $\mathbb{C}$. The composite map

$$R \to R' \to R'/\mathfrak{m}$$

is just given by an n-tuple of complex numbers, i.e. to a point in $\mathbb{C}^n$ which is even in V as it is a map out of R. This corresponds to a maximal ideal in R. This maximal ideal does not contain f by construction. □

4.4.7 Remark (exercise) Prove the following result, known as "Zariski's lemma" (which easily implies the Nullstellensatz): if k is a field, k' a field extension of k which is a finitely generated k-*algebra*, then k' is finite algebraic over k. Use the following argument of McCabe (in **?**):

1. k' contains a subring S of the form $S = k[x_1, \ldots, x_t]$ where the $x_1, \ldots, x_t$ are algebraically independent over k, and k' is algebraic over the quotient field of S (which is a polynomial ring).
2. If k' is not algebraic over k, then $S \neq k$ is not a field.
3. Show that there is $y \in S$ such that k' is integral over S_y. Deduce that S_y is a field.
4. Since $\operatorname{Spec}(S_y) = \{0\}$, argue that y lies in every non-zero prime ideal of $\operatorname{Spec} S$. Conclude that $1 + y \in k$, and S is a field—contradiction.

A little affine algebraic geometry

In what follows, let k be algebraically closed, and let A be a finitely generated k-algebra. Recall that $\operatorname{Spec}_{\max} A$ denotes the set of maximal ideals in A. Consider the natural k-algebra structure on $\operatorname{Funct}(\operatorname{Spec}_{\max} A, k)$. We have a map

$$A \to \operatorname{Funct}(\operatorname{Spec}_{\max} A, k)$$

which comes from the Weak Nullstellensatz as follows. Maximal ideals $\mathfrak{m} \subset A$ are in bijection with maps $\varphi_{\mathfrak{m}} : A \to k$ where $\ker(\varphi_{\mathfrak{m}}) = \mathfrak{m}$, so we define $a \longmapsto [\mathfrak{m} \longmapsto \varphi_{\mathfrak{m}}(a)]$. If A is reduced, then this map is injective because if $a \in A$ maps to the zero function, then $a \in \cap\, \mathfrak{m} \to a$ is nilpotent $\to a = 0$.

4.4.8 Definition A function $f \in \operatorname{Funct}(\operatorname{Spec}_{\max} A, k)$ is called **algebraic** if it is in the image of A under the above map. (Alternate words for this are **polynomial** and **regular**.)

Let A and B be finitely generated k-algebras and $\phi : A \to B$ a homomorphism. This yields a map $\Phi : \operatorname{Spec}_{\max} B \to \operatorname{Spec}_{\max} A$ given by taking pre-images.

4.4.9 Definition A map $\Phi : \operatorname{Spec}_{\max} B \to \operatorname{Spec}_{\max} A$ is called **algebraic** if it comes from a homomorphism ϕ as above.

To demonstrate how these definitions relate to one another we have the following proposition.

4.4.10 Proposition *A map $\Phi : \operatorname{Spec}_{\max} B \to \operatorname{Spec}_{\max} A$ is algebraic if and only if for any algebraic function $f \in \operatorname{Funct}(\operatorname{Spec}_{\max} A, k)$, the pullback $f \circ \Phi \in \operatorname{Funct}(\operatorname{Spec}_{\max} B, k)$ is algebraic.*

Proof. Suppose that Φ is algebraic. It suffices to check that the following diagram is commutative:

$$\begin{tikzcd} \mathrm{Funct}(\mathrm{Spec}_{\max} A, k) \arrow[r, "- \circ \Phi"] & \mathrm{Funct}(\mathrm{Spec}_{\max} B, k) \\ A \arrow[u] \arrow[r, "\phi"'] & B \arrow[u] \end{tikzcd}$$

where $\phi : A \to B$ is the map that gives rise to Φ.

[$\Leftarrow$] Suppose that for all algebraic functions $f \in \mathrm{Funct}(\mathrm{Spec}_{\max} A, k)$, the pull-back $f \circ \Phi$ is algebraic. Then we have an induced map, obtained by chasing the diagram counter-clockwise:

$$\begin{tikzcd} \mathrm{Funct}(\mathrm{Spec}_{\max} A, k) \arrow[r, "- \circ \Phi"] & \mathrm{Funct}(\mathrm{Spec}_{\max} B, k) \\ A \arrow[u] \arrow[r, dashed, "\phi"'] & B \arrow[u] \end{tikzcd}$$

From ϕ, we can construct the map $\Phi' : \mathrm{Spec}_{\max} B \to \mathrm{Spec}_{\max} A$ given by $\Phi'(\mathfrak{m}) = \phi^{-1}(\mathfrak{m})$. I claim that $\Phi = \Phi'$. If not, then for some $\mathfrak{m} \in \mathrm{Spec}_{\max} B$ we have $\Phi(\mathfrak{m}) \neq \Phi'(\mathfrak{m})$. By definition, for all algebraic functions $f \in \mathrm{Funct}(\mathrm{Spec}_{\max} A, k)$, $f \circ \Phi = f \circ \Phi'$ so to arrive at a contradiction we show the following lemma:
Given any two distinct points in $\mathrm{Spec}_{\max} A = V(I) \subset k^n$, there exists some algebraic f that separates them. This is trivial when we realize that any polynomial function is algebraic, and such polynomials separate points. $\square$

4.5. Serre's criterion and its variants

We are going to now prove a useful criterion for a noetherian ring to be a product of normal domains, due to Serre: it states that a (noetherian) ring is normal if and only if most of the localizations at prime ideals are discrete valuation rings (this corresponds to the ring being *regular* in codimension one, though we have not defined regularity yet) and a more technical condition that we will later interpret in terms of *depth.* One advantage of this criterion is that it does *not* require the ring to be a product of domains a priori.

Reducedness

There is a "baby" version of Serre's criterion for testing whether a ring is reduced, which we star with.

Recall:

4.5.1 Definition A ring R is **reduced** if it has no nonzero nilpotents.

4.5.2 Proposition *If R is noetherian, then R is reduced if and only if it satisfies the following conditions:*

1. *Every associated prime of R is minimal (no embedded primes).*
2. *If $\mathfrak{p}$ is minimal, then $R_{\mathfrak{p}}$ is a field.*

Proof. First, assume R reduced. What can we say? Say $\mathfrak{p}$ is a minimal prime; then $R_{\mathfrak{p}}$ has precisely one prime ideal (namely, $\mathfrak{m} = \mathfrak{p}R_{\mathfrak{p}}$). It is in fact a local artinian ring, though we don't need that fact. The radical of $R_{\mathfrak{p}}$ is just $\mathfrak{m}$. But R was reduced, so $R_{\mathfrak{p}}$ was reduced; it's an easy argument that localization preserves reducedness. So $\mathfrak{m} = 0$. The fact that 0 is a maximal ideal in $R_{\mathfrak{p}}$ says that it is a field.

On the other hand, we still have to do part 1. R is reduced, so $\mathrm{Rad}(R) = \bigcap_{\mathfrak{p} \in \mathrm{Spec}\, R} \mathfrak{p} = 0$. In particular,

$$\bigcap_{\mathfrak{p} \text{ minimal}} \mathfrak{p} = 0.$$

The map

$$R \to \prod_{\mathfrak{p} \text{ minimal}} R/\mathfrak{p}$$

is injective. The associated primes of the product, however, are just the minimal primes. So $\mathrm{Ass}(R)$ can contain only minimal primes.

That's one direction of the proposition. Let us prove the converse now. Assume R satisfies the two conditions listed. In other words, $\mathrm{Ass}(R)$ consists of minimal primes, and each $R_{\mathfrak{p}}$ for $\mathfrak{p} \in \mathrm{Ass}(R)$ is a field. We would like to show that R is reduced. Primary decomposition tells us that there is an injection

$$R \hookrightarrow \prod_{\mathfrak{p}_i \text{ minimal}} M_i, \quad M_i \;\; \mathfrak{p}_i - \text{primary}.$$

In this case, each M_i is primary with respect to a minimal prime. We have a map

$$R \hookrightarrow \prod M_i \to \prod (M_i)_{\mathfrak{p}_i},$$

which is injective, because when you localize a primary module at its associated prime, you don't kill anything by definition of primariness. Since we can draw a diagram

$$\begin{tikzcd} R \arrow[r] \arrow[d] & \prod M_i \arrow[d] \\ \prod R_{\mathfrak{p}_i} \arrow[r] & \prod (M_i)_{\mathfrak{p}_i} \end{tikzcd}$$

and the map $R \to \prod (M_i)_{\mathfrak{p}_i}$ is injective, the downward arrow on the right injective. Thus R can be embedded in a product of the fields $\prod R_{\mathfrak{p}_i}$, so is reduced. ∎

This proof actually shows:

4.5.3 Proposition (Scholism) *A noetherian ring R is reduced iff it injects into a product of fields. We can take the fields to be the localizations at the minimal primes.*

4.5.4 Example Let $R = k[X]$ be the coordinate ring of a variety X in $\mathbb{C}^n$. Assume X is reduced. Then MaxSpec$R$ is a union of irreducible components X_i, which are the closures of the minimal primes of R. The fields you get by localizing at minimal primes depend only on the irreducible components, and in fact are the rings of meromorphic functions on X_i. Indeed, we have a map

$$k[X] \to \prod k[X_i] \to \prod k(X_i).$$

If we don't assume that R is radical, this is **not** true.

There is a stronger condition than being reduced we could impose. We could say:

4.5.5 Proposition *If R is a noetherian ring, then R is a domain iff*

1. *R is reduced.*
2. *R has a unique minimal prime.*

Proof. One direction is obvious. A domain is reduced and (0) is the minimal prime.

The other direction is proved as follows. Assume 1 and 2. Let $\mathfrak{p}$ be the unique minimal prime of R. Then $\mathrm{Rad}(R) = 0 = \mathfrak{p}$ as every prime ideal contains $\mathfrak{p}$. As (0) is a prime ideal, R is a domain. ∎

We close by making some remarks about this embedding of R into a product of fields.

4.5.6 Definition Let R be any ring, not necessarily a domain. Let $K(R)$ be the localized ring $S^{-1}R$ where S is the multiplicatively closed set of non-zero-divisors in R. $K(R)$ is called the **total ring of fractions** of R.

When R is a field, this is the quotient field.

First, to get a feeling for this, we show:

4.5.7 Proposition *Let R be noetherian. The set of non-zero-divisors S can be described by $S = R - \bigcup_{\mathfrak{p} \in \mathrm{Ass}(R)} \mathfrak{p}$.*

Proof. If $x \in \mathfrak{p} \in \mathrm{Ass}(R)$, then x must kill something in R as it is in an associated prime. So x is a zero divisor.

Conversely, suppose x is a zero divisor, say $xy = 0$ for some $y \in R - \{0\}$. In particular, $x \in \mathrm{Ann}(y)$. We have an injection $R/\mathrm{Ann}(y) \hookrightarrow R$ sending 1 to y. But $R/\mathrm{Ann}(y)$ is nonzero, so it has an associated prime $\mathfrak{p}$ of $R/\mathrm{Ann}(y)$, which contains $\mathrm{Ann}(y)$ and thus x. But $\mathrm{Ass}(R/\mathrm{Ann}(y)) \subset \mathrm{Ass}(R)$. So x is contained in a prime in $\mathrm{Ass}(R)$. ∎

Assume now that R is reduced. Then $K(R) = S^{-1}R$ where S is the complement of the union of the minimal primes. At least, we can claim:

4.5.8 Proposition *Let R be reduced and noetherian. Then $K(R) = \prod_{\mathfrak{p}_i \text{ minimal}} R_{\mathfrak{p}_i}$.*

So $K(R)$ is the product of fields into which R embeds.

We now continue the discussion begun last time. Let R be noetherian and M a finitely generated R-module. We would like to understand very rough features of M. We can embed M into a larger R-module. Here are two possible approaches.

1. $S^{-1}M$, where S is a large multiplicatively closed subset of M. Let us take S to be the set of all $a \in R$ such that $M \xrightarrow{a} M$ is injective, i.e. a is not a zero divisor on M. Then the map
$$M \to S^{-1}M$$
is an injection. Note that S is the complement of the union of $\mathrm{Ass}(R)$.

2. Another approach would be to use a *primary decomposition*
$$M \hookrightarrow \prod M_i,$$
where each M_i is $\mathfrak{p}_i$-primary for some prime $\mathfrak{p}_i$ (and these primes range over $\mathrm{Ass}(M)$). In this case, it is clear that anything not in each $\mathfrak{p}_i$ acts injectively. So we can draw a commutative diagram
$$\begin{tikzcd} M \arrow[r] \arrow[d] & \prod M_i \arrow[d] \\ \prod M_{\mathfrak{p}_i} \arrow[r] & \prod (M_i)_{\mathfrak{p}_i} \end{tikzcd} .$$
The map going right and down is injective. It follows that M injects into the product of its localizations at associated primes.

The claim is that these constructions agree if M has no embedded primes. I.e., if there are no nontrivial containments among the associated primes of M, then $S^{-1}M$ (for $S = R - \bigcup_{\mathfrak{p} \in \mathrm{Ass}(M)} \mathfrak{p}$) is just $\prod M_{\mathfrak{p}}$. To see this, note that any element of S must act invertibly on $\prod M_{\mathfrak{p}}$. We thus see that there is always a map
$$S^{-1}M \to \prod_{\mathfrak{p} \in \mathrm{Ass}(M)} M_{\mathfrak{p}}.$$

4.5.9 Proposition *This is an isomorphism if M has no embedded primes.*

Proof. Let us go through a series of reductions. Let $I = \mathrm{Ann}(M) = \{a : aM = 0\}$. Without loss of generality, we can replace R by R/I. This plays nice with the associated primes.

The assumption is now that $\mathrm{Ass}(M)$ consists of the minimal primes of R.

Without loss of generality, we can next replace R by $S^{-1}R$ and M by $S^{-1}M$, because that doesn't affect the conclusion; localization plays nice with associated primes.

Now, however, R is artinian: i.e., all primes of R are minimal (or maximal). Why is this? Let R be *any* noetherian ring and $S = R - \bigcup_{\mathfrak{p} \text{ minimal}} \mathfrak{p}$. Then I claim that $S^{-1}R$ is artinian. We'll prove this in a moment.

So R is artinian, hence a product $\prod R_i$ where each R_i is local artinian. Without loss of generality, we can replace R by R_i by taking products. The condition we are trying to prove is now that

$$S^{-1}M \to M_{\mathfrak{m}}$$

for $\mathfrak{m} \subset R$ the maximal ideal. But S is the complement of the union of the minimal primes, so it is $R - \mathfrak{m}$ as R has one minimal (and maximal) ideal. This is obviously an isomorphism: indeed, both are M. ☐

To be added: proof of artianness

4.5.10 Corollary *Let R be a noetherian ring with no embedded primes (i.e.* $\mathrm{Ass}(R)$ *consists of minimal primes). Then $K(R) = \prod_{\mathfrak{p}_i \text{ minimal}} R_{\mathfrak{p}_i}$.*

If R is reduced, we get the statement made last time: there are no embedded primes, and $K(R)$ is a product of fields.

The image of $M \to S^{-1}M$

Let's ask now the following question. Let R be a noetherian ring, M a finitely generated R-module, and S the set of non-zero-divisors on M, i.e. $R - \bigcup_{\mathfrak{p} \in \mathrm{Ass}(M)} \mathfrak{p}$. We have seen that there is an imbedding

$$\phi : M \hookrightarrow S^{-1}M.$$

What is the image? Given $x \in S^{-1}M$, when does it belong to the imbedding above.

To answer such a question, it suffices to check locally. In particular:

4.5.11 Proposition *x belongs to the image of M in $S^{-1}M$ iff for every $\mathfrak{p} \in \operatorname{Spec} R$, the image of x in $(S^{-1}M)_{\mathfrak{p}}$ lies inside $M_{\mathfrak{p}}$.*

This isn't all that interesting. However, it turns out that you can check this at a smaller set of primes.

4.5.12 Proposition *In fact, it suffices to show that x is in the image of $\phi_{\mathfrak{p}}$ for every $\mathfrak{p} \in$* $\mathrm{Ass}(M/sM)$ *where $s \in S$.*

This is a little opaque; soon we'll see what it actually means. The proof is very simple.

Proof. Remember that $x \in S^{-1}M$. In particular, we can write $x = y/s$ where $y \in M, s \in S$. What we'd like to prove that $x \in M$, or equivalently that $y \in sM$.[5] In particular, we want to know that y maps to zero in M/sM. If not, there exists an associated prime $\mathfrak{p} \in \mathrm{Ass}(M/sM)$ such that y does not get killed in $(M/sM)_{\mathfrak{p}}$. We have assumed, however, for every associated prime $\mathfrak{p} \in \mathrm{Ass}(M)$, $x \in (S^{-1}M)_{\mathfrak{p}}$ lies in the image of $M_{\mathfrak{p}}$. This states that the image of y in this quotient $(M/sM)_{\mathfrak{p}}$ is zero, or that y is divisible by s in this localization. ☐

[5] In general, this would be equivalent to $ty \in tsM$ for some $t \in S$; but S consists of non-zero-divisors on M.

The case we actually care about is the following:

Take R as a noetherian domain and $M = R$. Then $S = R - \{0\}$ and $S^{-1}M$ is just the fraction field $K(R)$. The goal is to describe R as a subset of $K(R)$. What we have proven is that R is the intersection in the fraction field

$$\boxed{R = \bigcap_{\mathfrak{p} \in \mathrm{Ass}(R/s), s \in R - 0} R_{\mathfrak{p}}.}$$

So to check that something belongs to R, we just have to check that in a *certain set of localizations.*

Let us state this as a result:

4.5.13 Theorem *If R is a noetherian domain*

$$R = \bigcap_{\mathfrak{p} \in \mathrm{Ass}(R/s), s \in R - 0} R_{\mathfrak{p}}$$

Serre's criterion

We can now state a result.

4.5.14 Theorem (Serre) *Let R be a noetherian domain. Then R is integrally closed iff it satisfies*

1. *For any $\mathfrak{p} \subset R$ of height one, $R_{\mathfrak{p}}$ is a DVR.*
2. *For any $s \neq 0$, R/s has no embedded primes (i.e. all the associated primes of R/s are height one).*

Here is the non-preliminary version of the Krull theorem.

4.5.15 Theorem (Algebraic Hartogs) *Let R be a noetherian integrally closed ring. Then*

$$R = \bigcap_{\mathfrak{p} \text{ height one}} R_{\mathfrak{p}},$$

where each $R_{\mathfrak{p}}$ is a DVR.

Proof. Now evident from the earlier result theorem 4.5.13 and Serre's criterion. ☐

Earlier in the class, we proved that a domain was integrally closed if and only if it could be described as an intersection of valuation rings. We have now shown that when R is noetherian, we can take *discrete* valuation rings.

4.5.16 Remark In algebraic geometry, say $R = \mathbb{C}[x_1, \ldots, x_n]/I$. Its maximal spectrum is a subset of $\mathbb{C}^n$. If I is prime, and R a domain, this variety is irreducible. We are trying to describe R inside its field of fractions.

The field of fractions are like the "meromorphic functions"; R is like the holomorphic functions. Geometrically, this states to check that a meromorphic function is holomorphic, you can just check this by computing the "poleness" along each codimension one subvariety. If the function doesn't blow up on each of the codimension one subvarieties, and R is normal, then you can extend it globally.

This is an algebraic version of Hartog's theorem: this states that a holomorphic function on $\mathbb{C}^2 - (0,0)$ extends over the origin, because this has codimension > 1.

All the obstructions of extending a function to all of $\operatorname{Spec} R$ are in codimension one.

Now, we prove Serre's criterion.

Proof. Let us first prove that R is integrally closed if 1 and 2 occur. We know that

$$R = \bigcap_{\mathfrak{p} \in \operatorname{Ass}(R/x), x \neq 0} R_{\mathfrak{p}};$$

by condition 1, each such $\mathfrak{p}$ is of height one, and $R_{\mathfrak{p}}$ is a DVR. So R is the intersection of DVRs and thus integrally closed.

The hard part is going in the other direction. Assume R is integrally closed. We want to prove the two conditions. In R, consider the following conditions on a prime ideal $\mathfrak{p}$:

1. $\mathfrak{p}$ is an associated prime of R/x for some $x \neq 0$.
2. $\mathfrak{p}$ is height one.
3. $\mathfrak{p}_{\mathfrak{p}}$ is principal in $R_{\mathfrak{p}}$.

First, 3 implies 2 implies 1. 3 implies that $\mathfrak{p}$ contains an element x which generates $\mathfrak{p}$ after localizing. It follows that there can be no prime between (x) and $\mathfrak{p}$ because that would be preserved under localization. Similarly, 2 implies 1 is easy. If $\mathfrak{p}$ is minimal over (x), then $\mathfrak{p} \in \operatorname{Ass} R/(x)$ since the minimal primes in the support are always associated.

We are trying to prove the inverse implications. In that case, the claims of the theorem will be proved. We have to show that 1 implies 3. This is an argument we really saw last time, but let's see it again. Say $\mathfrak{p} \in \operatorname{Ass}(R/x)$. We can replace R by $R_{\mathfrak{p}}$ so that we can assume that $\mathfrak{p}$ is maximal. We want to show that $\mathfrak{p}$ is generated by one element.

What does the condition $\mathfrak{p} \in \operatorname{Ass}(R/x)$ buy us? It tells us that there is $\overline{y} \in R/x$ such that $\operatorname{Ann}(\overline{y}) = \mathfrak{p}$. In particular, there is $y \in R$ such that $\mathfrak{p}y \subset (x)$ and $y \notin (x)$. We have the element $y/x \in K(R)$ which sends $\mathfrak{p}$ into R. That is,

$$(y/x)\mathfrak{p} \subset R.$$

There are two cases to consider, as in last time:

1. $(y/x)\mathfrak{p} = R$. Then $\mathfrak{p} = R(x/y)$ so $\mathfrak{p}$ is principal.

2. $(y/x)\mathfrak{p} \neq R$. In particular, $(y/x)\mathfrak{p} \subset \mathfrak{p}$. Then since $\mathfrak{p}$ is finitely generated, we find that y/x is integral over R, hence in R. This is a contradiction as $y \notin (x)$.

Only the first case is now possible. So $\mathfrak{p}$ is in fact principal. $\square$

III.5. Unique factorization and the class group

Commutative rings in general do not admit unique factorization. Nonetheless, for many rings ("integrally closed" rings), which includes the affine coordinate rings one obtains in algebraic geometry when one studies smooth varieties, there is an invariant called the "class group" that measures the failure of unique factorization. This "class group" is a certain quotient of codimension one primes (geometrically, codimension one subvarieties) modulo rational equivalence.

Many even nicer rings have the convenient property that their localizations at prime ideals *factorial*, a key example being the coordinate ring of an affine nonsingular variety. For these even nicer rings, an alternative method of defining the class group can be given: the class group corresponds to the group of isomorphism classes of *invertible modules*. Geometrically, such invertible modules are line bundles on the associated variety (or scheme).

5.1. Unique factorization

Definition

We begin with the nicest of all possible cases, when the ring itself admits unique factorization.

Let R be a domain.

5.1.1 Definition A nonzero element $x \in R$ is **prime** if (x) is a prime ideal.

In other words, x is not a unit, and if $x \mid ab$, then either $x \mid a$ or $x \mid b$.

We restate the earlier definition 2.7.7 slightly.

5.1.2 Definition A domain R is **factorial** (or a **unique factorization domain**, or a **UFD**) if every nonzero noninvertible element $x \in R$ factors as a product $x_1 \dots x_n$ where each x_i is prime.

Recall that a *principal ideal domain* is a UFD (theorem 2.7.9), as is a *euclidean* domain (theorem 2.7.11); actually, a euclidean domain is a PID. Previously, we imposed something seemingly slightly stronger: that the factorization be unique. We next show that we get that for free.

5.1.3 Proposition (The fundamental theorem of arithmetic) *This factorization is essentially unique, that is, up to multiplication by units.*

Proof. Let $x \in R$ be a nonunit. Say $x = x_1 \dots x_n = y_1 \dots y_m$ were two different prime factorizations. Then $m, n > 0$.

We have that $x_1 \mid y_1 \dots y_m$, so $x_1 \mid y_i$ for some i. But y_i is prime. So x_1 and y_i differ by a unit. By removing each of these, we can get a smaller set of nonunique factorizations. Namely, we find that

$$x_2 \dots x_n = y_1 \dots \hat{y}_i \dots y_m$$

and then we can induct on the number of factors. □

The motivating example is of course:

5.1.4 Example $\mathbb{Z}$ is factorial. This is the fundamental theorem of arithmetic, and follows because $\mathbb{Z}$ is a euclidean domain. The same observation applies to a polynomial ring over a field by proposition 2.7.12.

Gauß's lemma

We now show that factorial rings are closed under the operation of forming polynomial rings.

5.1.5 Theorem (Gauß's lemma) *If R is factorial, so is the polynomial ring $R[X]$.*

In general, if R is a PID, $R[X]$ will *not* be a PID. For instance, $\mathbb{Z}[X]$ is not a PID: the prime ideal $(2, X)$ is not principal.

Proof. In the course of this proof, we shall identify the prime elements in $R[X]$. We start with a lemma that allows us to compare factorizations in $K[X]$ (for K the quotient field) and $R[X]$; the advantage is that we already know the polynomial ring over a *field* to be a UFD.

5.1.6 Lemma *Suppose R is a unique factorization domain with quotient field K. Suppose $f \in R[X]$ is irreducible in $R[X]$ and there is no nontrivial common divisor of the coefficients of f. Then f is irreducible in $K[X]$.*

With this in mind, we say that a polynomial in $R[X]$ is **primitive** if the coefficients have no common divisor in R.

Proof. Indeed, suppose we had a factorization

$$f = gh, \quad g, h \in K[X],$$

where g, h have degree $\geqslant 1$. Then we can clear denominators to find a factorization

$$rf = g'h'$$

where $r \in R - \{0\}$ and $g', h' \in R[X]$. By clearing denominators as little as possible, we may assume that g', h' are primitive. To be precise, we divide g', h' by their *contents*. Let us define:

5.1.7 Definition The **content** $\mathrm{Cont}(f)$ of a polynomial $f \in R[X]$ is the greatest common divisor of its coefficients. The content of an element f in $K[X]$ is defined by considering $r \in R$ such that $rf \in R[X]$, and taking $\mathrm{Cont}(rf)/r$. This is well-defined, modulo elements of R^*, and we have $\mathrm{Cont}(sf) = s \,\mathrm{Cont}\, f$ if $s \in K$.

To say that the content lies in R is to say that the polynomial is in $R[X]$; to say that the content is a unit is to say that the polynomial is primitive. Note that a monic polynomial in $R[X]$ is primitive.

So we have:

5.1.8 Lemma *Any element of* $K[X]$ *is a product of* $\mathrm{Cont}(f)$ *and something primitive in* $R[X]$.

Proof. Indeed, $f/\mathrm{Cont}(f)$ has content a unit. It therefore cannot have anything in the denominator. Indeed, if it had a term $r/p^i X^n$ where $r, p \in R$ and $p \nmid r$ is prime, then the content would divide r/p^i. It thus could not be in R. $\square$

5.1.9 Lemma $\mathrm{Cont}(fg) = \mathrm{Cont}(f)\,\mathrm{Cont}(g)$ *if* $f, g \in K[X]$.

Proof. By dividing f, g by their contents, it suffices to show that the product of two primitive polynomials in $R[X]$ (i.e. those with no common divisor of all their coefficients) is itself primitive. Indeed, suppose f, g are primitive and $p \in R$ is a prime. Then $\overline{f}, \overline{g} \in R/(p)[X]$ are nonzero. Their product $\overline{f}\overline{g}$ is also not zero because $R/(p)[X]$ is a domain, p being prime. In particular, p is not a common factor of the coefficients of fg. Since p was arbitrary, this completes the proof. $\square$

So return to the main proof. We know that $f = gh$. We divided g, h by their contents to get $g', h' \in R[X]$. We had then

$$rf = g'h', \quad r \in K^*.$$

Taking the contents, and using the fact that f, g', h' are primitive, we have then:

$$r = \mathrm{Cont}(g')\,\mathrm{Cont}(h') = 1 \quad (\text{modulo } R^*).$$

But then $f = r^{-1}g'h'$ shows that f is not irreducible in $R[X]$, contradiction. $\square$

Let R be a ring. Recall that an element is **irreducible** if it admits no nontrivial factorization. The product of an irreducible element and a unit is irreducible. Call a ring **finitely irreducible** if every element in the ring admits a factorization into finitely many irreducible elements.

5.1.10 Lemma *A ring R is finitely irreducible if every ascending sequence of* principal *ideals in R stabilizes.*

A ring such that every ascending sequence of ideals (not necessarily principal) stabilizes is said to be *noetherian;* this is a highly useful finiteness condition on a ring.

Proof. Suppose R satisfies the ascending chain condition on principal ideals. Then let $x \in R$. We would like to show it can be factored as a product of irreducibles. So suppose x is not the product of finitely many irreducibles. In particular, it is reducible: $x = x_1 x_1'$, where neither factor is a unit. One of this cannot be written as a finite product of irreducibles. Say it is x_1. Similarly, we can write $x_1 = x_2 x_2''$ where one of the factors, wlog x_2, is not the product of finitely many irreducibles. Repeating inductively gives the ascending sequence

$$(x) \subset (x_1) \subset (x_2) \subset \dots,$$

and since each factorization is nontrivial, the inclusions are each nontrivial. This is a contradiction. ∎

5.1.11 Lemma *Suppose R is a UFD. Then every ascending sequence of principal ideals in $R[X]$ stabilizes. In particular, $R[X]$ is finitely irreducible.*

Proof. Suppose $(f_1) \subset (f_2) \subset \cdots \in R[X]$. Then each $f_{i+1} \mid f_i$. In particular, the degrees of f_i are nonincreasing, and consequently stabilize. Thus for $i \gg 0$, we have $\deg f_{i+1} = \deg f_i$. We can thus assume that all the degrees are the same. In this case, if $i \gg 0$ and $k > 0$, $f_i/f_{i+k} \in R[X]$ must actually lie in R as R is a domain. In particular, throwing out the first few elements in the sequence if necessary, it follows that our sequence looks like

$$f, f/r_1, f/(r_1 r_2), \dots$$

where the $r_i \in R$. However, we can only continue this a finite amount of time before the r_i's will have to become units since R is a UFD. (Or $f = 0$.) So the sequence of ideals stabilizes. ∎

5.1.12 Lemma *Every element in $R[X]$ can be factored into a product of irreducibles.*

Proof. Now evident from the preceding lemmata. ∎

Suppose P is an irreducible element in $R[X]$. I claim that P is prime. There are two cases:

1. $P \in R$ is a prime in R. Then we know that $P \mid f$ if and only if the coefficients of f are divisible by P. In particular, $P \mid f$ iff $P \mid \mathrm{Cont}(f)$. It is now clear that $P \mid fg$ if and only if P divides one of $\mathrm{Cont}(f), \mathrm{Cont}(g)$ (since $\mathrm{Cont}(fg) = \mathrm{Cont}(f)\,\mathrm{Cont}(g)$).
2. P does not belong to R. Then P must have content a unit or it would be divisible by its content. So P is irreducible in $K[X]$ by the above reasoning.

   Say we have an expression
   $$P \mid fg, \quad f, g \in R[X].$$
   Since P is irreducible, hence prime, in the UFD (even PID) $K[X]$, we have that P divides one of f, g in $K[X]$. Say we can write
   $$f = qP, q \in K[X].$$
   Then taking the content shows that $\mathrm{Cont}(q) = \mathrm{Cont}(f) \in R$, so $q \in R[X]$. It follows that $P \mid f$ in $R[X]$.

We have shown that every element in $R[X]$ factors into a product of prime elements. From this, it is clear that $R[X]$ is a UFD. ∎

5.1.13 Corollary *The polynomial ring $k[X_1, \dots, X_n]$ for k a field is factorial.*

Proof. Induction on n. ∎

Factoriality and height one primes

We now want to give a fancier criterion for a ring to be a UFD, in terms of the lattice structure on $\operatorname{Spec} R$. This will require a notion from dimension theory (to be developed more fully later).

5.1.14 Definition Let R be a domain. A prime ideal $\mathfrak{p} \subset R$ is said to be of **height one** if $\mathfrak{p}$ is minimal among ideals containing x for some nonzero $x \in R$.

So a prime of height one is not the zero prime, but it is as close to zero as possible, in some sense. When we later talk about dimension theory, we will talk about primes of any height. In a sense, $\mathfrak{p}$ is "almost" generated by one element.

5.1.15 Theorem *Let R be a noetherian domain. The following are equivalent:*

1. *R is factorial.*
2. *Every height one prime is principal.*

Proof. Let's first show 1) implies 2). Assume R is factorial and $\mathfrak{p}$ is height one, minimal containing (x) for some $x \neq 0 \in R$. Then x is a nonunit, and it is nonzero, so it has a prime factorization

$$x = x_1 \dots x_n, \quad \text{each } x_i \text{ prime.}$$

Some $x_i \in \mathfrak{p}$ because $\mathfrak{p}$ is prime. In particular,

$$\mathfrak{p} \supset (x_i) \supset (x).$$

But (x_i) is prime itself, and it contains (x). The minimality of $\mathfrak{p}$ says that $\mathfrak{p} = (x_i)$.

Conversely, suppose every height one prime is principal. Let $x \in R$ be nonzero and a nonunit. We want to factor x as a product of primes. Consider the ideal $(x) \subsetneq R$. As a result, (x) is contained in a prime ideal. Since R is noetherian, there is a minimal prime ideal $\mathfrak{p}$ containing (x). Then $\mathfrak{p}$, being a height one prime, is principal—say $\mathfrak{p} = (x_1)$. It follows that $x_1 \mid x$ and x_1 is prime. Say

$$x = x_1 x_1'.$$

If x_1' is a nonunit, repeat this process to get $x_1' = x_2 x_2'$ with x_2 a prime element. Keep going; inductively we have

$$x_k = x_{k+1} x_{k+1}'.$$

If this process stops, with one of the x_k' a unit, we get a prime factorization of x. Suppose the process continues forever. Then we would have

$$(x) \subsetneq (x_1') \subsetneq (x_2') \subsetneq (x_3') \subsetneq \dots,$$

which is impossible by noetherianness. ∎

We have seen that unique factorization can be formulated in terms of prime ideals.

Factoriality and normality

We next state a generalization of the "rational root theorem" as in high school algebra.

5.1.16 Proposition *A factorial domain is integrally closed.*

Proof. **To be added: proof – may be in the queue already** ☐

5.2. Weil divisors

Definition

We start by discussing Weil divisors.

5.2.1 Definition A **Weil divisor** for R is a formal linear combination $\sum n_i[\mathfrak{p}_i]$ where the $\mathfrak{p}_i$ range over height one primes of R. So the group of Weil divisors is the free abelian group on the height one primes of R. We denote this group by $\mathrm{Weil}(R)$.

The geometric picture behind Weil divisors is that a Weil divisor is like a hypersurface: a subvariety of codimension one.

Valuations

Nagata's lemma

We finish with a fun application of the exact sequence of Weil divisors to a purely algebraic statement about factoriality.

5.2.2 Lemma *Let A be a normal noetherian domain.*

5.2.3 Theorem *Let A be a noetherian domain, $x \in A - \{0\}$. Suppose (x) is prime and A_x is factorial. Then A is factorial.*

Proof. We first show that A is normal (hence regular in codimension one). Indeed, A_x is normal. So if $t \in K(A)$ is integral over A, it lies in A_x. So we need to check that if $a/x^n \in A_x$ is integral over A and $x \nmid x$, then $n = 0$. Suppose we had an equation

$$(a/x^n)^N + b_1(a/x^n)^{N-1} + \cdots + b_N = 0.$$

Multiplying both sides by x^{nN} gives that

$$a^N \in xR,$$

so $x \mid a$ by primality.

Now we use the exact sequence

$$(x) \to \mathrm{Cl}(A) \to \mathrm{Cl}(A_x) \to 0.$$

The end is zero, and the image of the first map is zero. So $\mathrm{Cl}(A) = 0$. Thus A is a UFD. ☐

5.3. Locally factorial domains

Definition

5.3.1 Definition A noetherian domain R is said to be **locally factorial** if $R_{\mathfrak{p}}$ is factorial for each $\mathfrak{p}$ prime.

5.3.2 Example The coordinate ring $\mathbb{C}[x_1, \dots, x_n/I$ of an algebraic variety is locally factorial if the variety is smooth. We may talk about this later.

5.3.3 Example (Nonexample) Let R be $\mathbb{C}[A, B, C, D]/(AD - BC)$. The spectrum of R has maximal ideals consisting of 2-by-2 matrices of determinant zero. This variety is very singular at the origin. It is not even locally factorial at the origin.

The failure of unique factorization comes from the fact that

$$AD = BC$$

in this ring R. This is a prototypical example of a ring without unique factorization. The reason has to do with the fact that the variety has a singularity at the origin.

The Picard group

5.3.4 Definition Let R be a commutative ring. An R-module I is **invertible** if there exists J such that

$$I \otimes_R J \simeq R.$$

Invertibility is with respect to the tensor product.

5.3.5 Remark In topology, one is often interested in classifying *vector bundles* on spaces. In algebraic geometry, a module M over a ring R gives (as in **??**) a sheaf of abelian groups over the topological space $\operatorname{Spec} R$; this is supposed to be an analogy with the theory of vector bundles. (It is not so implausible since the Serre-Swan theorem (**??**) gives an equivalence of categories between the vector bundles over a compact space X and the projective modules over the ring $C(X)$ of continuous functions.) In this analogy, the invertible modules are the *line bundles.* The definition has a counterpart in the topological setting: for instance, a vector bundle $\mathcal{E} \to X$ over a space X is a line bundle (that is, of rank one) if and only if there is a vector bundle $\mathcal{E}' \to X$ such that $\mathcal{E} \otimes \mathcal{E}'$ is the trivial bundle $X \times \mathbb{R}$.

There are many equivalent characterizations.

5.3.6 Proposition *Let R be a ring, I an R-module. TFAE:*

1. *I is invertible.*
2. *I is finitely generated and $I_{\mathfrak{p}} \simeq R_{\mathfrak{p}}$ for all primes $\mathfrak{p} \subset R$.*

3. *I is finitely generated and there exist $a_1, \dots, a_n \in R$ which generate (1) in R such that*

$$I[a_i^{-1}] \simeq R[a_i^{-1}].$$

Proof. First, we show that if I is invertible, then I is finitely generated. Suppose $I \otimes_R J \simeq R$. This means that $1 \in R$ corresponds to an element

$$\sum i_k \otimes j_k \in I \otimes_R J.$$

Thus, there exists a finitely generated submodule $I_0 \subset I$ such that the map $I_0 \otimes J \to I \otimes J$ is surjective. Tensor this with I, so we get a surjection

$$I_0 \simeq I_0 \otimes J \otimes I \to I \otimes J \otimes I \simeq I$$

which leads to a surjection $I_0 \twoheadrightarrow I$. This implies that I is finitely generated

Step 1: 1 implies 2. We now show 1 implies 2. Note that if I is invertible, then $I \otimes_R R'$ is an invertible R' module for any R-algebra R'; to get an inverse of $I \otimes_R R'$, tensor the inverse of I with R'. In particular, $I_{\mathfrak{p}}$ is an invertible $R_{\mathfrak{p}}$-module for each $\mathfrak{p}$. As a result,

$$I_{\mathfrak{p}}/\mathfrak{p}I_{\mathfrak{p}}$$

is invertible over the *field* $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$. This means that $I_{\mathfrak{p}}/\mathfrak{p}I_{\mathfrak{p}}$ is a one-dimensional vector space over the residue field. (The invertible modules over a vector space are the one-dimensional spaces.) Choose an element $x \in I_{\mathfrak{p}}$ which generates $I_{\mathfrak{p}}/\mathfrak{p}I_{\mathfrak{p}}$. Since $I_{\mathfrak{p}}$ is finitely generated, Nakayama's lemma shows that x generates $I_{\mathfrak{p}}$.

We get a surjection $\alpha : R_{\mathfrak{p}} \twoheadrightarrow I_{\mathfrak{p}}$ carrying $1 \to x$. We claim that this map is injective. This will imply that $I_{\mathfrak{p}}$ is free of rank 1. So, let J be an inverse of I among R-modules, so that $I \otimes_R J = R$; the same argument as above provides a surjection $\beta : R_{\mathfrak{p}} \to J_{\mathfrak{p}}$. Then $\beta' = \beta \otimes 1_{I_{\mathfrak{p}}} : I_{\mathfrak{p}} \to R_{\mathfrak{p}}$ is also a surjection. Composing, we get a surjective map

$$R_{\mathfrak{p}} \overset{\alpha}{\twoheadrightarrow} I_{\mathfrak{p}} \overset{\beta'}{\twoheadrightarrow} R_{\mathfrak{p}}$$

whose composite must be multiplication by a unit, since the ring is local. Thus the composite is injective and α is injective. It follows that α is an isomorphism, so that $I_{\mathfrak{p}}$ is free of rank one.

Step 2: 2 implies 3. Now we show 2 implies 3. Suppose I is finitely generated with generators $\{x_1, \dots, x_n\} \subset I$ and $I_{\mathfrak{p}} \simeq R_{\mathfrak{p}}$ for all $\mathfrak{p}$. Then for each $\mathfrak{p}$, we can choose an element x of $I_{\mathfrak{p}}$ generating $I_{\mathfrak{p}}$ as $R_{\mathfrak{p}}$-module. By multiplying by the denominator, we can assume that $x \in I$. By assumption, we can then find $a_i, s_i \in R$ with

$$s_i x_i = a_i x \in R$$

for some $s_i \notin \mathfrak{p}$ as x generates $I_{\mathfrak{p}}$. This means that x generates I after inverting the s_i. It follows that $I[1/a] = R[1/a]$ where $a = \prod s_i \notin \mathfrak{p}$. In particular, we find that there is an open covering $\{\operatorname{Spec} R[1/a_{\mathfrak{p}}]\}$ of $\operatorname{Spec} R$ (where $a_{\mathfrak{p}} \notin \mathfrak{p}$) on which I is isomorphic to R. To say that these cover $\operatorname{Spec} R$ is to say that the $a_{\mathfrak{p}}$ generate 1.

Finally, let's do the implication 3 implies 1. Assume that we have the situation of $I[1/a_i] \simeq R[1/a_i]$. We want to show that I is invertible. We start by showing that I is **finitely presented**. This means that there is an exact sequence

$$R^m \to R^n \to I \to 0,$$

i.e. I is the cokernel of a map between free modules of finite rank. To see this, first, we've assumed that I is finitely generated. So there is a surjection

$$R^n \twoheadrightarrow I$$

with a kernel $K \rightarrowtail R^n$. We must show that K is finitely generated. Localization is an exact functor, so $K[1/a_i]$ is the kernel of $R[1/a_i]^n \to I[1/a_i]$. However, we have an exact sequence

$$K[1/a_i] \rightarrowtail R[1/a_i]^n \twoheadrightarrow R[1/a_i]$$

by the assumed isomorphism $I[1/a_i] \simeq R[1/a_i]$. But since a free module is projective, this sequence splits and we find that $K[1/a_i]$ is finitely generated. If it's finitely generated, it's generated by finitely many elements in K. As a result, we find that there is a map

$$R^N \to K$$

such that the localization to $\operatorname{Spec} R[1/a_i]$ is surjective. This implies by the homework that $R^N \to K$ is surjective.[1] Thus K is finitely generated.

In any case, we have shown that the module I is finitely presented. **Define** $J = \hom_R(I, R)$ as the candidate for its dual. This construction is compatible with localization. We can choose a finite presentation $R^m \to R^n \to I \to 0$, which leads to a sequence

$$0 \to J \to \hom(R^n, R) \to \hom(R^m, R).$$

It follows that the formation of J commutes with localization. In particular, this argument shows that

$$J[1/a] = \hom_{R[1/a]}(I[1/a], R[1/a]).$$

One can check this by using the description of J. By construction, there is a canonical map $I \otimes J \to R$. I claim that this map is invertible.

For the proof, we use the fact that one can check for an isomorphism locally. It suffices to show that

$$I[1/a] \otimes J[1/a] \to R[1/a]$$

is an isomorphism for some collection of a's that generate the unit ideal. However, we have $a_1, \dots, a_n$ that generate the unit ideal such that $I[1/a_i]$ is free of rank 1, hence so is $J[1/a_i]$. It thus follows that $I[1/a_i] \otimes J[1/a_i]$ is an isomorphism. $\square$

5.3.7 Definition Let R be a commutative ring. We define the **Picard group** $\mathrm{Pic}(R)$ to be the set of isomorphism classes of invertible R-modules. This is an abelian group; the addition law is defined so that the sum of the classes represented by M, N is $M \otimes_R N$. The identity element is given by R.

[1] To check that a map is surjective, just check at the localizations at any maximal ideal.

The Picard group is thus analogous (cf. **??**) to the set of isomorphism classes of line bundles on a topological space (which is also an abelian group). While the latter can often be easily computed (for a nice space X, the line bundles are classified by elements of $H^2(X, \mathbb{Z})$), the interpretation in the algebraic setting is more difficult.

Cartier divisors

Assume furthermore that R is a domain. We now introduce:

5.3.8 Definition A **Cartier divisor** for R is a submodule $M \subset K(R)$ such that M is invertible.

In other words, a Cartier divisor is an invertible fractional ideal. Alternatively, it is an invertible R-module M with a nonzero map $M \to K(R)$. **Once this map is nonzero, it is automatically injective,** since injectivity can be checked at the localizations, and any module-homomorphism from a domain into its quotient field is either zero or injective (because it is multiplication by some element).

We now make this into a group.

5.3.9 Definition Given $(M, a : M \hookrightarrow K(R))$ and $(N, b : N \hookrightarrow K(R))$, we define the sum to be

$$(M \otimes N, a \otimes b : M \otimes N \hookrightarrow K(R)).$$

The map $a \otimes b$ is nonzero, so by what was said above, it is an injection. Thus the Cartier divisors from an abelian group $\mathrm{Cart}(R)$.

By assumption, there is a homomorphism

$$\mathrm{Cart}(R) \to \mathrm{Pic}(R)$$

mapping $(M, M \hookrightarrow K(R)) \to M$.

5.3.10 Proposition *The map* $\mathrm{Cart}(R) \to \mathrm{Pic}(R)$ *is surjective. In other words, any invertible R-module can be embedded in $K(R)$.*

Proof. Let M be an invertible R-module. Indeed, we know that $M_{(0)} = M \otimes_R K(R)$ is an invertible $K(R)$-module, so a one-dimensional vector space over $K(R)$. In particular, $M_{(0)} \simeq K(R)$. There is a nonzero homomorphic map

$$M \to M_{(0)} \simeq K(R),$$

which is automatically injective by the discussion above. □

What is the kernel of $\mathrm{Cart}(R) \to \mathrm{Pic}(R)$? This is the set of Cartier divisors which are isomorphic to R itself. In other words, it is the set of $(R, R \hookrightarrow K(R))$. This data is the same thing as the data of a nonzero element of $K(R)$. So the kernel of

$$\mathrm{Cart}(R) \to \mathrm{Pic}(R)$$

has kernel isomorphic to $K(R)^*$. We have a short exact sequence

$$K(R)^* \to \mathrm{Cart}(R) \to \mathrm{Pic}(R) \to 0.$$

Weil divisors and Cartier divisors

Now, we want to assume $\mathrm{Cart}(R)$ if R is "good." The "goodness" in question is to assume that R is locally factorial, i.e. that $R_{\mathfrak{p}}$ is factorial for each $\mathfrak{p}$. This is true, for instance, if R is the coordinate ring of a smooth algebraic variety.

5.3.11 Proposition *If R is locally factorial and noetherian, then the group* $\mathrm{Cart}(R)$ *is a free abelian group. The generators are in bijection with the height one primes of R.*

Now assume that R is a locally factorial, noetherian domain. We shall produce an isomorphism

$$\mathrm{Weil}(R) \simeq \mathrm{Cart}(R)$$

that sends $[\mathfrak{p}_i]$ to that height one prime $\mathfrak{p}_i$ together with the imbedding $\mathfrak{p}_i \hookrightarrow R \to K(R)$.

We first check that this is well-defined. Since $\mathrm{Weil}(R)$ is free, all we have to do is check that each $\mathfrak{p}_i$ is a legitimate Cartier divisor. In other words, we need to show that:

5.3.12 Proposition *If $\mathfrak{p} \subset R$ is a height one prime and R locally factorial, then $\mathfrak{p}$ is invertible.*

Proof. In the last lecture, we gave a criterion for invertibility: namely, being locally trivial. We have to show that for any prime $\mathfrak{q}$, we have that $\mathfrak{p}_{\mathfrak{q}}$ is isomorphic to $R_{\mathfrak{q}}$. If $\mathfrak{p} \not\subset \mathfrak{q}$, then $\mathfrak{p}_{\mathfrak{q}}$ is the entire ring $R_{\mathfrak{q}}$, so this is obvious. Conversely, suppose $\mathfrak{p} \subset \mathfrak{q}$. Then $\mathfrak{p}_{\mathfrak{q}}$ is a height one prime of $R_{\mathfrak{q}}$: it is minimal over some element in $R_{\mathfrak{q}}$.

Thus $\mathfrak{p}_{\mathfrak{q}}$ is principal, in particular free of rank one, since $R_{\mathfrak{q}}$ is factorial. We saw last time that being factorial is equivalent to the principalness of height one primes. □

We need to define the inverse map

$$\mathrm{Cart}(R) \to \mathrm{Weil}(R).$$

In order to do this, start with a Cartier divisor $(M, M \hookrightarrow K(R))$. We then have to describe which coefficient to assign a height one prime. To do this, we use a local criterion.

Let's first digress a bit. Consider a locally factorial domain R and a prime $\mathfrak{p}$ of height one. Then $R_{\mathfrak{p}}$ is factorial. In particular, its maximal ideal $\mathfrak{p}R_{\mathfrak{p}}$ is height one, so principal. It is the principal ideal generated by some $t \in R_{\mathfrak{p}}$. Now we show:

5.3.13 Proposition *Every nonzero ideal in $R_{\mathfrak{p}}$ is of the form (t^n) for some unique $n \geqslant 0$.*

Proof. Let $I_0 \subset R_{\mathfrak{p}}$ be nonzero. If $I_0 = R_{\mathfrak{p}}$, then we're done—it's generated by t^0. Otherwise, $I_0 \subsetneq R_{\mathfrak{p}}$, so contained in $\mathfrak{p}R_{\mathfrak{p}} = (t)$. So let $I_1 = \{x \in R_{\mathfrak{p}} : tx \in I_0\}$. Thus

$$I_1 = t^{-1} I_0.$$

I claim now that $I_1 \neq I_0$, i.e. that there exists $x \in R_{\mathfrak{p}}$ such that $x \notin I_0$ but $tx \in I_0$. The proof comes from the theory of associated primes. Look at $R_{\mathfrak{p}}/I_0$; it has at least one associated prime as it is nonzero.

Since it is a torsion module, this associated prime must be $\mathfrak{p}R_{\mathfrak{p}}$ since the only primes in $R_{\mathfrak{p}}$ are (0) and (t), **which we have not yet shown**. So there exists an element in the quotient R/I_0 whose annihilator is precisely (t). Lifting this gives an element in R which when multiplied by (t) is in I_0 but which is not in I_0. So $I_0 \subsetneq I_1$.

Proceed as before now. Define $I_2 = \{x \in R_{\mathfrak{p}} : tx \in I_1\}$. This process must halt since we have assumed noetherianness. We must have $I_m = I_{m+1}$ for some m, which would imply that some $I_m = R_{\mathfrak{p}}$ by the above argument. It then follows that $I_0 = (t^m)$ since each I_i is just tI_{i+1}. ∎

We thus have a good structure theory for ideals in R localized at a height one prime. Let us make a more general claim.

5.3.14 Proposition *Every nonzero finitely generated $R_{\mathfrak{p}}$-submodule of the fraction field $K(R)$ is of the form (t^n) for some $n \in \mathbb{Z}$.*

Proof. Say that $M \subset K(R)$ is such a submodule. Let $I = \{x \in R_{\mathfrak{p}}, xM \subset R_{\mathfrak{p}}\}$. Then $I \neq 0$ as M is finitely generated M is generated over $R_{\mathfrak{p}}$ by a finite number of fractions $a_i/b_i, b_i \in R$. Then the product $b = \prod b_i$ brings M into $R_{\mathfrak{p}}$.

We know that $I = (t^m)$ for some m. In particular, $t^m M$ is an ideal in R. In particular,

$$t^m M = t^p R$$ ∎

for some p, in particular $M = t^{p-m} R$.

Now let's go back to the main discussion. R is a noetherian locally factorial domain; we want to construct a map

$$\mathrm{Cart}(R) \to \mathrm{Weil}(R).$$

Given $(M, M \hookrightarrow K(R))$ with M invertible, we want to define a formal sum $\sum n_i[\mathfrak{p}_i]$. For every height one prime $\mathfrak{p}$, let us look at the local ring $R_{\mathfrak{p}}$ with maximal ideal generated by some $t_{\mathfrak{p}} \in R_{\mathfrak{p}}$. Now $M_{\mathfrak{p}} \subset K(R)$ is a finitely generated $R_{\mathfrak{p}}$-submodule, so generated by some $t_{\mathfrak{p}}^{n_{\mathfrak{p}}}$. So we map $(M, M \hookrightarrow K(R))$ to

$$\sum_{\mathfrak{p}} n_{\mathfrak{p}}[\mathfrak{p}].$$

First, we have to check that this is well-defined. In particular, we have to show:

5.3.15 Proposition *For almost all height one $\mathfrak{p}$, we have $M_{\mathfrak{p}} = R_{\mathfrak{p}}$. In other words, the integers $n_{\mathfrak{p}}$ are almost all zero.*

Proof. We can always assume that M is actually an ideal. Indeed, choose $a \in R$ with $aM = I \subset R$. As Cartier divisors, we have $M = I - (a)$. If we prove the result for I and (a), then we will have proved it for M (note that the $n_{\mathfrak{p}}$'s are additive invariants[2]). So because of this additivity, it is sufficient to prove the proposition for actual (i.e. nonfractional) ideals.

[2] To see this, localize at $\mathfrak{p}$—then if M is generated by t^a, N generated by t^b, then $M \otimes N$ is generated by t^{a+b}.

Assume thus that $M \subset R$. All of these $n_{\mathfrak{p}}$ associated to M are at least zero because M is actually an ideal. What we want is that $n_{\mathfrak{p}} \leqslant 0$ for almost all $\mathfrak{p}$. In other words, we must show that

$$M_{\mathfrak{p}} \supset R_{\mathfrak{p}} \quad \text{almost all } \mathfrak{p}.$$

To do this, just choose any $x \in M - 0$. There are finitely many minimal primes containing (x) (by primary decomposition applied to $R/(x)$). Every other height one prime $\mathfrak{q}$ does not contain (x).[3] This states that $M_{\mathfrak{q}} \supset x/x = 1$, so $M_{\mathfrak{q}} \supset R_{\mathfrak{q}}$.

The key claim we've used in this proof is the following. If $\mathfrak{q}$ is a height one prime in a domain R containing some nonzero element (x), then $\mathfrak{q}$ is minimal among primes containing (x). In other words, we can test the height one condition at any nonzero element in that prime. Alternatively:

5.3.16 Lemma *There are no nontrivial containments among height one primes.*

Anyway, we have constructed maps between $\mathrm{Cart}(R)$ and $\mathrm{Weil}(R)$. The map $\mathrm{Cart}(R) \to \mathrm{Weil}(R)$ takes $M \to \sum n_{\mathfrak{p}}[\mathfrak{p}]$. The other map $\mathrm{Weil}(R) \to \mathrm{Cart}(R)$ takes $[\mathfrak{p}] \to \mathfrak{p} \subset K(R)$. The composition $\mathrm{Weil}(R) \to \mathrm{Weil}(R)$ is the identity. Why is that? Start with a prime $\mathfrak{p}$; that goes to the Cartier divisor $\mathfrak{p}$. Then we need to finitely generatedre the multiplicities at other height one primes. But if $\mathfrak{p}$ is height one and $\mathfrak{q}$ is a height one prime, then if $\mathfrak{p} \neq \mathfrak{q}$ the lack of nontrivial containment relations implies that the multiplicity of $\mathfrak{p}$ at $\mathfrak{q}$ is zero. We have shown that

$$\mathrm{Weil}(R) \to \mathrm{Cart}(R) \to \mathrm{Weil}(R)$$

is the identity.

Now we have to show that $\mathrm{Cart}(R) \to \mathrm{Weil}(R)$ is injective. Say we have a Cartier divisor $(M, M \hookrightarrow K(R))$ that maps to zero in $\mathrm{Weil}(R)$, i.e. all its multiplicities $n_{\mathfrak{p}}$ are zero at height one primes. We show that $M = R$.

First, assume $M \subset R$. It is sufficient to show that at any maximal ideal $\mathfrak{m} \subset R$, we have

$$M_{\mathfrak{m}} = R_{\mathfrak{m}}.$$

What can we say? Well, $M_{\mathfrak{m}}$ is principal as M is invertible, being a Cartier divisor. Let it be generated by $x \in R_{\mathfrak{m}}$; suppose x is a nonunit (or we're already done). But $R_{\mathfrak{m}}$ is factorial, so $x = x_1 \dots x_n$ for each x_i prime. If $n > 0$, then however M has nonzero multiplicity at the prime ideal $(x_i) \subset R_{\mathfrak{m}}$. This is a contradiction.

The general case of M not really a subset of R can be handled similarly: then the generating element x might lie in the fraction field. So x, if it is not a unit in R, is a product of some primes in the numerator and some primes in the denominator. The nonzero primes that occur lead to nonzero multiplicities.

[3] Again, we're using something about height one primes not proved yet.

Recap and a loose end

Last time, it was claimed that if R is a locally factorial domain, and $\mathfrak{p} \subset R$ is of height one, then every prime ideal of $R_{\mathfrak{p}}$ is either maximal or zero. This follows from general dimension theory. This is equivalent to the following general claim about height one primes:

> There are no nontrivial inclusions among height one primes for R a locally factorial domain.

Proof. Suppose $\mathfrak{q} \subsetneq \mathfrak{p}$ is an inclusion of height one primes.

Replace R by $R_{\mathfrak{p}}$. Then R is local with some maximal ideal $\mathfrak{m}$, which is principal with some generator x. Then we have an inclusion

$$0 \subset \mathfrak{q} \subset \mathfrak{m}.$$

This inclusion is proper. However, $\mathfrak{q}$ is principal since it is height one in the factorial ring $R_{\mathfrak{p}}$. This cannot be since every element is a power of x times a unit. (Alright, this wasn't live TEXed well.) ☐

Last time, we were talking about $\mathrm{Weil}(R)$ and $\mathrm{Cart}(R)$ for R a locally factorial noetherian domain.

1. $\mathrm{Weil}(R)$ is free on the height one primes.
2. $\mathrm{Cart}(R)$ is the group of invertible submodules of $K(R)$.

We produced an isomorphism
$$\mathrm{Weil}(R) \simeq \mathrm{Cart}(R).$$

5.3.17 Remark Geometrically, what is this? Suppose $R = \mathbb{C}[X_1, \dots, X_n]/I$ for some ideal I. Then the maximal ideals, or closed points in $\operatorname{Spec} R$, are certain points in $\mathbb{C}^n$; they form an irreducible variety if R is a domain. The locally factorial condition is satisfied, for instance, if the variety is *smooth*. In this case, the Weil divisors correspond to sums of irreducible varieties of codimension one—which correspond to the primes of height one. The Weil divisors are free on the set of irreducible varieties of codimension one.

The Cartier divisors can be thought of as "linear combinations" of subvarieties which are locally defined by one equation. It is natural to assume that the condition of being defined by one equation corresponds to being codimension one. This is true by the condition of R locally factorial.

In general, we can always construct a map

$$\mathrm{Cart}(R) \to \mathrm{Weil}(R),$$

but it is not necessarily an isomorphism.

Further remarks on Weil(R) **and** Cart(R)

Recall that the Cartier group fits in an exact sequence:

$$K(R)^* \to \mathrm{Cart}(R) \to \mathrm{Pic}(R) \to 0,$$

because every element of $\mathrm{Cart}(R)$ determines its isomorphism class, and every element of $K(R)^*$ determines a free module of rank one. Contrary to what was stated last time, it is **not true** that exactness holds on the right. In fact, the kernel is the group R^* of units of R. So the exact sequence runs

$$0 \to R^* \to K(R)^* \to \mathrm{Cart}(R) \to \mathrm{Pic}(R) \to 0.$$

This is true for *any* domain R. For R locally factorial and noetherian, we know that $\mathrm{Cart}(R) \simeq \mathrm{Weil}(R)$, though.

We can think of this as a generalization of unique factorization.

5.3.18 Proposition *R is factorial if and only if R is locally factorial and* $\mathrm{Pic}(R) = 0$.

Proof. Assume R is locally factorial and $\mathrm{Pic}(R) = 0$. Then every prime ideal of height one (an element of $\mathrm{Weil}(R)$, hence of $\mathrm{Cart}(R)$) is principal, which implies that R is factorial. And conversely. □

In general, we can think of the exact sequence above as a form of unique factorization for a locally factorial domain: any invertible fractional ideal is a product of height one prime ideals.

Let us now give an example. **To be added: ?**

III.6. Dedekind domains

The notion of a Dedekind domain allows one to generalize the usual unique factorization in principal ideal domains as in $\mathbb{Z}$ to settings such as the ring of integers in an algebraic number field. In general, a Dedekind domain does not have unique factorization, but the *ideals* in a Dedekind domain do factor uniquely into a product of prime ideals. We shall see that Dedekind domains have a short characterization in terms of the characteristics we have developed.

After this, we shall study the case of an *extension* of Dedekind domains $A \subset B$. It will be of interest to determine how a prime ideal of A factors in B. This should provide background for the study of basic algebraic number theory, e.g. a rough equivalent of the first chapter of **?** or **?**.

6.1. Discrete valuation rings

Definition

We start with the simplest case of a *discrete valuation ring,* which is the local version of a Dedekind domain. Among the one-dimensional local noetherian rings, these will be the nicest.

6.1.1 Theorem *Let R be a noetherian local domain whose only prime ideals are (0) and the maximal ideal $\mathfrak{m} \neq 0$. Then, the following are equivalent:*

1. *R is factorial.*
2. *$\mathfrak{m}$ is principal.*
3. *R is integrally closed.*
4. *R is a valuation ring with value group $\mathbb{Z}$.*

6.1.2 Definition A ring satisfying these conditions is called a **discrete valuation ring** (**DVR**). A discrete valuation ring necessarily has only two prime ideals, namely $\mathfrak{m}$ and (0).

Alternatively, we can say that a noetherian local domain is a DVR if and only if it is of dimension one and integrally closed.

Proof. Assume 1: that is, suppose R is factorial. Then every prime ideal of height one is principal by theorem 5.1.15. But $\mathfrak{m}$ is the only prime that can be height one: it is minimal over any nonzero nonunit of R, so $\mathfrak{m}$ is principal. Thus 1 implies 2, and similarly 2 implies 1 by theorem 5.1.15.

1 implies 3 is true for any R: a factorial ring is always integrally closed, by proposition 5.1.16.

4 implies 2 is easy as well. Indeed, suppose R is a valuation ring with value group $\mathbb{Z}$. Then, one chooses an element $x \in R$ such that the valuation of x is one. It is easy to see that x generates $\mathfrak{m}$: if $y \in \mathfrak{m}$ is a non-unit, then the valuation of y is at least one, so $y/x \in R$ and $y \in (x)$.

The proof that 2 implies 4 is also straightforward. Suppose $\mathfrak{m}$ is principal, generated by t. In this case, we claim that any $x \in R$ is associate (i.e. differs by a unit from) a power of t. Indeed, since $\bigcap \mathfrak{m}^n = 0$ by the Krull intersection theorem (**??**), it follows that there exists n such that x is divisible by t^n but not by t^{n+1}. In particular, if we write $x = ut^n$, then $u \notin (t)$ is a unit. This proves the claim.

With this in mind, we need to show that R is a valuation ring with value group $\mathbb{Z}$. If $x \in R$, we define the valuation of x to be the nonnegative integer n such that $(x) = (t^n)$. One can easily check that this is a valuation on R, which extends to the quotient field by additivity.

The interesting part of the argument is the claim that 3 implies 2. Suppose R is integrally closed, noetherian, and of dimension one; we claim that $\mathfrak{m}$ is principal. Choose $x \in \mathfrak{m} - \{0\}$. If $(x) = \mathfrak{m}$, we are done.

Otherwise, we can look at $\mathfrak{m}/(x) \neq 0$. The module $\mathfrak{m}/(x)$ is finitely generated module a noetherian ring which is nonzero, so it has an associated prime. That associated prime is either zero or $\mathfrak{m}$ because R has dimension one. But 0 is not an associated prime because every element in the module is killed by x. So $\mathfrak{m}$ is an associated prime of $\mathfrak{m}/(x)$.

There is $\overline{y} \in \mathfrak{m}/(x)$ whose annihilator is $\mathfrak{m}$. Thus, there is $y \in \mathfrak{m}$ such that $y \notin (x)$ and $\mathfrak{m}y \subset (x)$. In particular, $y/x \in K(R) - R$, but

$$(y/x)\mathfrak{m} \subset R.$$

There are two cases:

1. Suppose $(y/x)\mathfrak{m} = R$. Then we can write $\mathfrak{m} = R(x/y)$. So $\mathfrak{m}$ is principal. (This argument shows that $x/y \in R$.)
2. The other possibility is that $y/x\mathfrak{m} \subsetneq R$. In this case, $(y/x)\mathfrak{m}$ is an ideal, so

   $$(y/x)\mathfrak{m} \subset \mathfrak{m}.$$

   In particular, multiplication by y/x carries $\mathfrak{m}$ to itself, and stabilizes the finitely generated *faithful* module $\mathfrak{m}$. By proposition 4.1.7, we see that y/x is integral over R. In particular, we find that $y/x \in R$, as R was integrally closed, a contradiction as $y \notin (x)$. □

Let us give several examples of DVRs.

6.1.3 Example The localization $\mathbb{Z}_{(p)}$ at any prime ideal $(p) \neq 0$ is a DVR. The associated valuation is the p-adic valuation.

6.1.4 Example Although we shall not prove (or define) this, the local ring of an algebraic curve at a smooth point is a DVR. The associated valuation measures the extent to which a function (or germ thereof) has a zero (or pole) at that point.

6.1.5 Example The formal power series ring $\mathbb{C}[[T]]$ is a discrete valuation ring, with maximal ideal (T).

Another approach

In the proof of theorem 6.1.1, we freely used the notion of associated primes, and thus some of the results of chapter III.2. However, we can avoid all that and give a more "elementary approach," as in **?**.

Let us suppose that R is an integrally closed, local noetherian domain of dimension one. We shall prove that the maximal ideal $\mathfrak{m} \subset R$ is principal. This was the hard part of theorem 6.1.1, and the only part where we used associated primes earlier.

Proof. We will show that $\mathfrak{m}$ is principal, by showing it is *invertible* (as will be seen below). We divide the proof into steps:

Step one For a nonzero ideal $I \subset R$, let $I^{-1} := \{x \in K(R) : xI \subset R\}$, where $K(R)$ is the quotient field of R. Then clearly $I^{-1} \supset R$ and I^{-1} is an R-module, but in general we cannot say that $I^{-1} \neq R$ even if I is proper. Nevertheless, we claim that in the present situation, we have

$$\mathfrak{m}^{-1} \neq R.$$

This is the conclusion of Step one.

The proof runs across a familiar line: we show that any maximal element in the set of ideals $I \subset R$ with $I^{-1} \neq R$ is prime. The set of such ideals is nonempty: it contains any (a) for $a \in \mathfrak{m}$ (in which case $(a)^{-1} = Ra^{-1} \neq R$). There must be a maximal element in this set of ideals by noetherianness, which as we will see is prime; thus, that maximal element must be $\mathfrak{m}$, which proves our claim.

So to fill in the missing link, we must prove:

6.1.6 Lemma *If S is a noetherian domain, any maximal element in the set of ideals $I \subset S$ with $I^{-1} \neq S$ is prime.*

Proof. Let J be a maximal element, and suppose we have $ab \in J$, with $a, b \notin J$. I claim that if $z \in J^{-1} - S$, then $za, zb \in J^{-1} - S$. The J^{-1} part follows since J^{-1} is a S-module.

By symmetry it is enough to prove the other half for a, namely that $za \notin S$; but then if $za \in S$, we would have $z((a) + J) \subset S$, which implies $((a) + J)^{-1} \neq S$, contradiction, for J was maximal.

Then it follows that $z(ab) = (za)b \in J^{-1} - S$, by applying the claim just made twice. But $ab \in J$, so $z(ab) \in S$, contradiction. ☐

Step two In the previous step, we have established that $\mathfrak{m}^{-1} \neq R$.

We now claim that $\mathfrak{m}\mathfrak{m}^{-1} = R$. First, we know of course that $\mathfrak{m}\mathfrak{m}^{-1} \subset R$ by definition of inverses, and equally $\mathfrak{m} \subset \mathfrak{m}\mathfrak{m}^{-1}$ too. So $\mathfrak{m}\mathfrak{m}^{-1}$ is an ideal sandwiched between $\mathfrak{m}$ and R. Thus we only need to prove that $\mathfrak{m}\mathfrak{m}^{-1} = \mathfrak{m}$ is impossible. If this were the case, we could choose some $a \in \mathfrak{m}^{-1} - R$ which must satisfy $a\mathfrak{m} \subset \mathfrak{m}$. Then a would integral over R. As R is integrally closed, this is impossible.

Step three Finally, we claim that $\mathfrak{m}$ is principal, which is the final step of the proof. In fact, let us prove a more general claim.

6.1.7 Proposition *Let $(R, \mathfrak{m})$ be a local noetherian domain such that $\mathfrak{m}\mathfrak{m}^{-1} = R$. Then $\mathfrak{m}$ is principal.*

Proof. Indeed, since $\mathfrak{m}\mathfrak{m}^{-1} = R$, write

$$1 = \sum m_i n_i, \quad m_i \in \mathfrak{m},\ n_i \in \mathfrak{m}^{-1}.$$

At least one $m_j n_j$ is invertible, since R is local. It follows that there are $x \in \mathfrak{m}$ and $y \in \mathfrak{m}^{-1}$ whose product xy is a unit in R. We may even assume $xy = 1$.

Then we claim $\mathfrak{m} = (x)$. Indeed, we need only prove $\mathfrak{m} \subset (x)$. For this, if $q \in \mathfrak{m}$, then $qy \in R$ by definition of $\mathfrak{m}^{-1}$, so

$$q = x(qy) \in (x).$$ □

So we are done in this case too. Taking stock, we have an effective way to say whether a ring is a DVR. These three conditions are much easier to check in practice (noetherianness is usually easy, integral closure is usually automatic, and the last one is not too hard either for reasons that will follow) than the existence of an absolute value.

6.2. Dedekind rings

Definition

We now introduce a closely related notion.

6.2.1 Definition A **Dedekind ring** is a noetherian domain R such that

1. R is integrally closed.
2. Every nonzero prime ideal of R is maximal.

6.2.2 Remark If R is Dedekind, then any nonzero element is height one. This is evident since every nonzero prime is maximal.

If R is Dedekind, then R is locally factorial. In fact, the localization of R at a nonzero prime $\mathfrak{p}$ is a DVR.

Proof. $R_{\mathfrak{p}}$ has precisely two prime ideals: (0) and $\mathfrak{p}R_{\mathfrak{p}}$. As a localization of an integrally closed domain, it is integrally closed. So $R_{\mathfrak{p}}$ is a DVR by the above result (hence factorial). □

Assume R is Dedekind now. We have an exact sequence

$$0 \to R^* \to K(R)^* \to \mathrm{Cart}(R) \to \mathrm{Pic}(R) \to 0.$$

Here $\mathrm{Cart}(R) \simeq \mathrm{Weil}(R)$. But $\mathrm{Weil}(R)$ is free on the nonzero primes, or equivalently maximal ideals, R being Dedekind. In fact, however, $\mathrm{Cart}(R)$ has a simpler description.

6.2.3 Proposition *Suppose R is Dedekind. Then* $\mathrm{Cart}(R)$ *consists of all nonzero finitely generated submodules of $K(R)$ (i.e.* ***fractional ideals****).*

This is the same thing as saying as every nonzero finitely generated submodule of $K(R)$ is invertible.

Proof. Suppose $M \subset K(R)$ is nonzero and finitely generated It suffices to check that M is invertible after localizing at every prime, i.e. that $M_{\mathfrak{p}}$ is an invertible—or equivalently, trivial, $R_{\mathfrak{p}}$-module. At the zero prime, there is nothing to check. We might as well assume that $\mathfrak{p}$ is maximal. Then $R_{\mathfrak{p}}$ is a DVR and $M_{\mathfrak{p}}$ is a finitely generated submodule of $K(R_{\mathfrak{p}}) = K(R)$.

Let S be the set of integers n such that there exists $x \in M_{\mathfrak{p}}$ with $v(x) = n$, for v the valuation of $R_{\mathfrak{p}}$. By finite generation of M, S is bounded below. Thus S has a least element k. There is an element of $M_{\mathfrak{p}}$, call it x, with valuation k.

It is easy to check that $M_{\mathfrak{p}}$ is generated by x, and is in fact free with generator x. The reason is simply that x has the smallest valuation of anything in $M_{\mathfrak{p}}$. ☐

What's the upshot of this?

6.2.4 Theorem *If R is a Dedekind ring, then any nonzero ideal $I \subset R$ is invertible, and therefore uniquely described as a product of powers of (nonzero) prime ideals, $I = \prod \mathfrak{p}_i^{n_i}$.*

Proof. This is simply because I is in $\mathrm{Cart}(R) = \mathrm{Weil}(R)$ by the above result. ☐

This is Dedekind's generalization of unique factorization.

We now give the standard examples:

6.2.5 Example

1. Any PID (in particular, any DVR) is Dedekind.
2. If K is a finite extension of $\mathbb{Q}$, and set R to be the integral closure of $\mathbb{Z}$ in K, then R is a Dedekind ring. The ring of integers in any number field is a Dedekind ring.
3. If R is the coordinate ring of an algebraic variety which is smooth and irreducible of dimension one, then R is Dedekind.
4. Let X be a compact Riemann surface, and let $S \subset X$ be a nonempty finite subset. Then the ring of meromorphic functions on X with poles only in S is Dedekind. The maximal ideals in this ring are precisely those corresponding to points of $X - S$.

A more elementary approach

We would now like to give a more elementary approach to the unique factorization of ideals in Dedekind domains, one which does not use the heavy machinery of Weil and Cartier divisors.

In particular, we can encapsulate what has already been proved as:

6.2.6 Theorem *Let A be a Dedekind domain with quotient field K. Then there is a bijection between the discrete valuations of K that assign nonnegative orders to elements of A and the nonzero prime ideals of A.*

Proof. Indeed, every valuation gives a prime ideal of elements of positive order; every prime ideal $\mathfrak{p}$ gives a discrete valuation on $A_{\mathfrak{p}}$, hence on K. ☐

This result, however trivial to prove, is the main reason we can work essentially interchangeably with prime ideals in Dedekind domains and discrete valuations.

Now assume A is Dedekind. A finitely generated A-submodule of the quotient field F is called a **fractional ideal**; by multiplying by some element of A, we can always pull a fractional ideal into A, when it becomes an ordinary ideal. The sum and product of two fractional ideals are fractional ideals.

6.2.7 Theorem (Invertibility) *If I is a nonzero fractional ideal and $I^{-1} := \{x \in F : xI \subset A\}$, then I^{-1} is a fractional ideal and $II^{-1} = A$.*

Thus, the nonzero fractional ideals are an *abelian group* under multiplication.

Proof. To see this, note that invertibility is preserved under localization: for a multiplicative set S, we have $S^{-1}(I^{-1}) = (S^{-1}I)^{-1}$, where the second ideal inverse is with respect to $S^{-1}A$; this follows from the fact that I is finitely generated. Note also that invertibility is true for discrete valuation rings: this is because the only ideals are principal, and principal ideals (in any integral domain) are obviously invertible.

So for all primes $\mathfrak{p}$, we have $(II^{-1})_{\mathfrak{p}} = A_{\mathfrak{p}}$, which means the inclusion of A-modules $II^{-1} \to A$ is an isomorphism at each localization. Therefore it is an isomorphism, by general algebra. ☐

The next result says we have unique factorization of **ideals**:

6.2.8 Theorem (Factorization) *Each ideal $I \subset A$ can be written uniquely as a product of powers of prime ideals.*

Proof. Let's use the pseudo-inductive argument to obtain existence of a prime factorization. Let I be the maximal ideal which can't be written in such a manner, which exists since A is Noetherian. Then I isn't prime (evidently), so it's contained in some prime $\mathfrak{p}$. But $I = (I\mathfrak{p}^{-1})\mathfrak{p}$, and $I\mathfrak{p}^{-1} \neq I$ can be written as a product of primes, by the inductive assumption. Whence so can I, contradiction.

Uniqueness of factorization follows by localizing at each prime. ☐

6.2.9 Definition Let P be the subgroup of nonzero principal ideals in the group I of nonzero ideals. The quotient I/P is called the **ideal class group**.

The ideal class group of the integers, for instance (or any principal ideal domain) is clearly trivial. In general, this is not the case, because Dedekind domains do not generally admit unique factorization.

6.2.10 Proposition *Let A be a Dedekind domain. Then A is a UFD if and only if its ideal class group is trivial.*

Proof. If the ideal class group is trivial, then A is a principal ideal domain, hence a UFD by elementary algebra. Conversely, suppose A admits unique factorization. Then, by the following lemma, every prime ideal is principal. Hence every ideal is principal, in view of the unique factorization of ideals. $\square$

6.2.11 Lemma *Let R be a UFD, and let $\mathfrak{p}$ be a prime ideal which contains no proper prime sub-ideal except for 0. Then $\mathfrak{p}$ is principal.*

The converse holds as well; a domain is a UFD if and only if every prime ideal of height one is principal, by Theorem 5.1.15.

Proof. First, $\mathfrak{p}$ contains an element $x \neq 0$, which we factor into irreducibles $\pi_1 \dots \pi_k$. One of these, say π_j, belongs to $\mathfrak{p}$, so $\mathfrak{p} \supset (\pi_j)$. Since $\mathfrak{p}$ is minimal among nonzero prime ideals, we have $\mathfrak{p} = (\pi_j)$. (Note that (π_j) is prime by unique factorization.) $\square$

6.2.12 Remark (exercise) This exercise is from **?**. If A is the integral closure of $\mathbb{Z}$ in a number field (so that A is a Dedekind domain), then it is known (cf. **?** for a proof) that the ideal class group of A is *finite*. From this, show that every open subset of $\operatorname{Spec} A$ is a principal open set $D(f)$. Scheme-theoretically, this means that every open subscheme of $\operatorname{Spec} A$ is affine (which is not true for general rings).

Modules over Dedekind domains

Let us now consider some properties of Dedekind domains.

6.2.13 Proposition *Let A be a Dedekind domain, and let M be a finitely generated A module. Then M is projective (or equivalently flat, or locally free) if and only if it is torsion-free.*

Proof. If M is projective, then it is a direct summand of a free module, so it is torsion-free. So we need to show that if M is torsion-free, then it is projective. Recall that to show M is projective, it suffices to show that $M_{\mathfrak{p}}$ is projective for any prime $\mathfrak{p} \subset M$. But note that $A_{\mathfrak{p}}$ is a PID so a module over it is torsion free if and only if it is flat, by Lemma **??**. However, it is also a local Noetherian ring, so a module is flat if and only if it is projective. So $M_{\mathfrak{p}}$ is projective if and only if it is torsion-free, so it now suffices to show that it is torsion-free.

However for any multiplicative set $S \subset A$, if M is torsion-free then M_S is also torsion-free. This is because if

$$\frac{a}{s'} \cdot \frac{m}{s} = 0$$

then there is t such that $tam = 0$, as desired. $\square$

6.2.14 Proposition *Let A be a Dedekind domain. Then any finitely generated module M over it has (not canonically) a decomposition $M = M^{tors} \oplus M^{tors-free}$.*

Proof. Note that by Lemma **??**, we have a short exact sequence

$$0 \to M^{tors} \to M \to M^{tors-free} \to 0$$

but by proposition 6.2.13 the torsion free part is projective, so M can be split, not necessarily canonically as $M^{tors} \oplus M^{tors-free}$, as desired. ☐

Note that we may give further information about the torsion free part of the module:

$$M^{tors} = \bigoplus_{\mathfrak{p}} M_{\mathfrak{p}}^{tors}$$

First note that there is a map

$$M^{tors} \to \bigoplus_{\mathfrak{p}} M_{\mathfrak{p}}^{tors}$$

because M is torsion, every element is supported at finitely many points, so the image of f in $M_{\mathfrak{p}}^{tors}$ is only nonzero for finitely many $\mathfrak{p}$. It is an isomorphism, because it is an isomorphism after every localization.

So we have pretty much specified what the torsion part is. We can in fact also classify the torsion free part; in particular, we have

$$M^{tors-free} \simeq \oplus \mathcal{L}$$

where $\mathcal{L}$ are locally free modules of rank 1. This is because we know from above that the torsion free module is projective, we may apply Problem Set 10, Problem 12, and then since L is a line bundle, and I_{-D} is also, $L \otimes I_{-D}$ is a line bundle, and then $M/L \otimes I_{-D}$ is flat, so it is projective, so we may split it off.

6.2.15 Lemma *For A a Dedekind Domain, and $I \subset A$ an ideal, then I is a locally free module of rank 1.*

Proof. First note that I is torsion-free and therefore projective by 6.2.13, and it is also finitely generated, because A is Noetherian. But for a finitely generated module over a Noetherian ring, we know that it is projective if and only if it is locally free, so we have shown that it is locally free.

Also recall that for a module which is locally free, the rank is well defined, i.e, any localization which makes it free makes it free of the same rank. So to test the rank, it suffices to show that if we tensor with the field of fractions K, it is free of rank 1. But note that since K, being a localization of A is flat over A so we have short exact sequence

$$0 \to I \otimes_A K \to A \otimes_A K \to (A/I) \otimes_A K \to 0$$

However, note that $\mathrm{supp}(A/I) = V(\mathrm{Ann}(A/I)) = V(I)$, and the prime (0) is not in $V(I)$, so $A/I \otimes_A K$, which is the localization of A/I at (0) vanishes, so we have

$$I \otimes_A K \simeq A \otimes_A K$$

but this is one-dimensional as a free K module, so the rank is 1, as desired. ☐

We close by listing a collection of useful facts about Dedekind domains. A dozen things every Good Algebraist shou
R is a Dedekind domain.

1. R is local $\iff R$ is a field or a DVR.
2. R semi-local $\implies$ it is a PID.
3. R is a PID $\iff$ it is a UFD $\iff C(R) = \{1\}$
4. R is the full ring of integers of a number field $K \implies |C(R)| < \infty$, and this number is the *class number* of K.
5. $C(R)$ can be any abelian group. This is Clayborn's Theorem.
6. For any non-zero prime $\mathfrak{p} \in \operatorname{Spec} R$, $\mathfrak{p}^n/\mathfrak{p}^{n+1} \cong R/\mathfrak{p}$ as an R-module.
7. "To contain is to divide", i.e. if $A, B \subset R$, then $A \subset B \iff A = BC$ for some $C \subset R$.
8. (Generation of ideals) Every non-zero ideal $B \subset R$ is generated by two elements. Moreover, one of the generators can be taken to be any non-zero element of B.
9. (Factor rings) If $A \subset R$ is non-zero, then R/A is a PIR (**principal ideal ring**).
10. (Steinitz Isomorphism Theorem) If $A, B \subset R$ are non-zero ideals, then $A \oplus B \cong {}_R R \oplus AB$ as R-modules.
11. If M is a finitely generated torsion-free R-module of rank n,[1] then it is of the form $M \cong R^{n-1} \oplus A$, where A is a non-zero ideal, determined up to isomorphism.
12. If M is a finitely generated torsion R-module, then M is uniquely of the form $M \cong R/A_1 \oplus \cdots \oplus R/A_n$ with $A_1 \subsetneq A_2 \subsetneq \cdots \subsetneq A_n \subsetneq R$.

To be added: eventually, proofs of these should be added

6.3. Extensions

In this section, we will essentially consider the following question: if A is a Dedekind domain, and L a finite extension of the quotient field of A, is the integral closure of A in L a Dedekind domain? The general answer turns out to be yes, but the result is somewhat simpler for the case of a separable extension, with which we begin.

[1] The rank is defined as $rk(M) = \dim_{K(R)} M \otimes_R K(R)$ where $K(R)$ is the quotient field.

Integral closure in a finite separable extension

One of the reasons Dedekind domains are so important is

6.3.1 Theorem *Let A be a Dedekind domain with quotient field K, L a finite separable extension of K, and B the integral closure of A in L. Then B is Dedekind.*

This can be generalized to the Krull-Akizuki theorem below (**??**).

First let us give an alternate definition of "separable". For a finite field extension k' of k, we may consider the bilinear pairing $k' \otimes_k k' \to k$ given by $x, y \mapsto \mathrm{tr}_{k'/k}(xy)$. Which is to say $xy \in k'$ can be seen as a k-linear map of finite dimensional vector spaces $k' \to k'$, and we are considering the trace of this map. Then we claim that k' is separable if and only if the bilinear pairing $k' \times k' \to k$ is non-degenerate.

To show the above claim, first note that the pairing is non-degenerate if and only if it is non-degenerate after tensoring with the algebraic closure. This is because if $\mathrm{tr}(xy) = 0$ for all $y \in k'$, then $\mathrm{tr}((x \otimes 1_{\overline{k}})y) = 0$ for all $y \in k' \otimes_k \overline{k}$, which we may see to be true by decomposing into pure tensors. The other direction is obtained by selecting a basis of $\overline{k}$ over k, and then noting that for y_i basis elements, if $\mathrm{tr}(\sum xy_i) = 0$ then $\mathrm{tr}(xy_i) = 0$ for each i.

So now we just need to show that $X = k' \otimes_k \overline{k}$ is reduced if and only if the map $X \otimes_{\overline{k}} X \to \overline{k}$ given by $a \otimes b \mapsto \mathrm{tr}(ab)$ is non-degenerate. To do this, we show that elements of the kernel of the bilinear map are exactly the nilpotents. But note that X is a finite dimensional algebra over $\overline{k}$, and we may elements as matrices. Then if $\mathrm{tr}(AB) = 0$ for all B if and only if $\mathrm{tr}(PAP^{-1}PBP^{-1}) = 0$ for all B, so we may assume A is in Upper Triangular Form. From this, the claim becomes clear.

Proof. We need to check that B is Noetherian, integrally, closed, and of dimension 1.

- Noetherian. Indeed, B is a finitely generated A-module, which obviously implies Noetherianness. To see this, note that the K-linear map $(.,.) : L \times L \to K$, $a, b \to \mathrm{Tr}(ab)$ is nondegenerate since L is separable over K (**??**). Let $F \subset B$ be a free module spanned by a K-basis for L. Then since traces preserve integrality and A is integrally closed, we have $B \subset F^*$, where $F^* := \{x \in K : (x, F) \subset A\}$. Now F^* is A-free on the dual basis of F though, so B is a submodule of a finitely generated A module, hence a finitely generated A-module.
- Integrally closed. B is the integral closure of A in L, so it is integrally closed (integrality being transitive).
- Dimension 1. Indeed, if $A \subset B$ is an integral extension of domains, then $\dim A = \dim B$. This follows essentially from the theorems of "lying over" and "going up." Cf. **?**.

So, consequently the ring of algebraic integers (integral over $\mathbb{Z}$) in a number field (finite extension of $\mathbb{Q}$) is Dedekind. ∎

Note that the above proof actually implied (by the argument about traces) the following useful fact:

6.3.2 Proposition *Let A be a noetherian integrally closed domain with quotient field K. Let L be a finite separable extension and B the ring of integers. Then B is a finitely generated A-module.*

We shall give another, more explicit proof of Proposition 6.3.2 whose technique will be useful in the sequel. Let $\alpha \in B$ be a generator of L/K. Let $n = [L : K]$ and $\sigma_1, \dots, \sigma_n$ the distinct embeddings of L into the algebraic closure of K. Define the **discriminant** of α to be

$$D(\alpha) = \left(\det \begin{bmatrix} 1 & \sigma_1\alpha & (\sigma_1\alpha)^2 & \dots \\ 1 & \sigma_2\alpha & (\sigma_2\alpha)^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right)^2 .$$

This maps to the same element under each σ_i, so is in K^* (and even A^* by integrality); it is nonzero by basic facts about vanderMonde determinants since each σ_i maps α to a different element. The next lemma clearly implies that B is contained in a finitely generated A-module, hence is finitely generated (since A is noetherian).

6.3.3 Lemma *We have $B \subset D(\alpha)^{-1}A[\alpha]$.*

Proof. Indeed, suppose $x \in B$. We can write $x = c_0(1) + c_1(\alpha) + \dots c_{n-1}(\alpha^{n-1})$ where each $c_i \in K$. We will show that in fact, each $c_i \in D(\alpha)^{-1}A$, which will prove the lemma. Applying each σ_i, we have for each i, $\sigma_i x = c_0(1) + c_1(\sigma_i\alpha) + \dots + c_{n-1}(\sigma_i\alpha^{n-1})$. Now by Cramer's lemma, each c_i can be written as a quotient of determinants of matrices involving $\sigma_j x$ and the α^j. The denominator determinant is in fact $D(\alpha)$. The numerator is in K and must be integral, hence is in A. This proves the claim and the lemma. ∎

The above technique may be illustrated with an example.

6.3.4 Example Let p^i be a power of a prime p and consider the extension $\mathbb{Q}(\zeta_{p^i})/\mathbb{Q}$ for ζ_{p^i} a primitive p^i-th root of unity. This is a special case of a cyclotomic extension, an important example in the subject. We claim that the ring of integers (that is, the integral closure of $\mathbb{Z}$) in $\mathbb{Q}(\zeta_{p^i})$ is precisely $\mathbb{Z}[\zeta_{p^i}]$. This is true in fact for all cyclotomic extensions, but we will not be able to prove it here.

First of all, ζ_{p^i} satisfies the equation $X^{p^{i-1}(p-1)} + X^{p^{i-1}(p-2)} + \dots + 1 = 0$. This is because if ζ_p is a p-th root of unity, $(\zeta_p - 1)(1 + \zeta_p + \dots + \zeta_p^{p-1}) = \zeta_p^p - 1 = 0$. In particular, $X - \zeta_{p^i} \mid X^{p^{i-1}(p-1)} + X^{p^{i-1}(p-2)} + \dots + 1$, and consequently (taking $X = 1$), we find that $1 - \zeta_{p^i}$ divides p in the ring of integers in $\mathbb{Q}(\zeta_{p^i})/\mathbb{Q}$. This is true for *any* primitive p^i-th root of unity for *any* p^i. Thus the norm to $\mathbb{Q}$ of $1 - \zeta_{p^i}^j$ for any j is a power of p.

I claim that this implies that the discriminant $D(\zeta_{p^i})$ is a power of p, up to sign. But by the vanderMonde formula, this discriminant is a product of terms of the form $\prod(1 - \zeta_{p^i}^j)$ up to roots of unity. The norm to $\mathbb{Q}$ of each factor is thus a power of p, and the discriminant itself plus or minus a power of p.

By the lemma, it follows that the ring of integers is contained in $\mathbb{Z}[p^{-1}, \zeta_{p^i}]$. To get down further to $\mathbb{Z}[\zeta_{p^i}]$ requires a bit more work. **To be added: this proof**

The Krull-Akizuki theorem

We are now going to prove a general theorem that will allow us to remove the separability hypothesis in **??**. Let us say that a noetherian domain has **dimension at most one** if every nonzero prime ideal is maximal; we shall later generalize this notion of "dimension."

6.3.5 Theorem (Krull-Akizuki) *Suppose A is a noetherian domain of dimension at most one. Let L be a finite extension of the quotient field $K(A)$, and suppose $B \subset L$ is a domain containing A. Then B is noetherian of dimension at most one.*

From this, it is clear:

6.3.6 Theorem *The integral closure of a Dedekind domain in any finite extension of the quotient field is a Dedekind domain.*

Proof. Indeed, by Krull-Akizuki, this integral closure is noetherian and of dimension $\leqslant 1$; it is obviously integrally closed as well, hence a Dedekind domain. □

Now let us prove Krull-Akizuki. **To be added: we need to introduce material about length**

Proof. We are going to show that for any $a \in A - \{0\}$, the A-module B/aB has finite length. (This is quite nontrivial, since B need not even be finitely *generated* as an A-module.) From this it will be relatively easy to deduce the result.

Indeed, if $I \subset B$ is any nonzero ideal, then I contains a nonzero element of A; to see this, we need only choose an element $b \in I$ and consider an irreducible polynomial

$$a_0 X^n + \cdots + a_n \in K[X]$$

that it satisfies. We can assume that all the $a_i \in A$ by clearing denominators. It then follows that $a_n \in A \cap I$. So choose some $a \in (A \cap I) - \{0\}$. We then know by the previous paragraph (though we have not proved it yet) that B/aB has finite length as an A-module (and a fortiori as a B-module); in particular, the submodule I/aB is finitely generated as a B-module. The exact sequence

$$0 \to aB \to I \to I/aB \to 0$$

shows that I must be finitely generated as a B-module, since the two outer terms are. Thus any ideal of B is finitely generated, so B is noetherian.

To be added: B has dimension at most one

To prove the Krull-Akizuki theorem, we are going to prove:

6.3.7 Lemma (Finite length lemma) *If A is a noetherian domain of dimension at most one, then for any torsion-free A-module M such that $K(A) \otimes_A M$ is finite-dimensional (alternatively: M has finite rank) and $a \neq 0$, M/aM has finite length.*

Proof. We are going to prove something stronger. If M has rank n and is torsion-free, then will show

$$\ell(M/aM) \leqslant n\ell(A/aA). \tag{6.3.1}$$

Note that A/aA has finite length. This follows because there is a filtration of A/aA whose quotients are of the form $A/\mathfrak{p}$ for $\mathfrak{p}$ prime; but these $\mathfrak{p}$ cannot be zero as A/aA is torsion. So these primes are maximal, and A/aA has a filtration whose quotients are *simple*. Thus $\ell(A/aA) < \infty$. In fact, we see thus that *any torsion, finitely-generated module has finite length*; this will be used in the sequel.

There are two cases:

1. M is finitely generated. We can choose generators $m_1, \dots, m_n$ in M of $K(A) \otimes_A M$; we then from these generators get a map

   $$A^n \to M$$

   which becomes an isomorphism after localizing at $A - \{0\}$. In particular, the kernel and cokernel are torsion modules. The kernel must be trivial (A being a domain), and $A^n \to M$ is thus injective. Thus we have found a finite free submodule $F \subset M$ such that M/F is a torsion module T, which is also finitely generated.

   We have an exact sequence

   $$0 \to F/(aM \cap F) \to M/aM \to T/aT \to 0.$$

   Here the former has length at most $\ell(F/aF) = n\ell(A/aA)$, and we get the bound $\ell(M/aM) \leqslant n\ell(A/aA) + \ell(T/aT)$. However, we have the annoying last term to contend with, which makes things somewhat inconvenient. Thus, we use a trick: for each $t > 0$, we consider the exact sequence

   $$0 \to F/(a^tM \cap F) \to M/a^tM \to T/a^tT \to 0.$$

   This gives

   $$\ell(M/a^tM) \leqslant tn\ell(A/aA) + \ell(T/a^tT) \leqslant tn\ell(A/aA) + \ell(T).$$

   However, $\ell(T) < \infty$ as T is torsion (cf. the first paragraph). If we divide by t, we get the inequality

   $$\frac{1}{t}\ell(M/a^tM) \leqslant n\ell(A/aA) + \frac{\ell(T)}{t}. \tag{6.3.2}$$

   However, the filtration $a^tM \subset a^{t-1}M \subset \cdots \subset aM \subset M$ whose quotients are all isomorphic to M/aM (M being torsion-free) shows that $\ell(M/a^tM) = t\ell(M/aM)$ In particular, letting $t \to \infty$ in (6.3.2) gives (6.3.1) in the case where M is finitely generated.

2. M is not finitely generated. Now we can use a rather cheeky argument. M is the inductive limit of its finitely generated submodules $M_F \subset M$, each of which is itself torsion free and of rank at most n. Thus M/aM is the inductive limit of its submodules $M_F/(aM \cap M_F)$ as M_F ranges over We know that $\ell(M_F/(aM \cap M_F)) \leqslant n\ell(A/aA)$ for each finitely generated $M_F \subset M$ by the first case above (and the fact that $M_F/(aM \cap M_F)$ is a quotient of M_F/aM_F).

But if M/aM is the inductive limit of *submodules* of length at most $n\ell(A/aA)$, then it itself can have length at most $n\ell(A/aA)$. For M/aM must be in fact equal to the submodule $M_F/(aM \cap M_F)$ that has the largest length (no other submodule $M_{F'}/(aM \cap M_{F'})$ can properly contain this). ∎

With this lemma proved, it is now clear that Krull-Akizuki is proved as well.

Extensions of discrete valuations

As a result, we find:

6.3.8 Theorem *Let K be a field, L a finite separable extension. Then a discrete valuation on K can be extended to one on L.*

To be added: This should be clarified — what is a discrete valuation?

Proof. Indeed, let $R \subset K$ be the ring of integers of the valuation, that is the subset of elements of nonnegative valuation. Then R is a DVR, hence Dedekind, so the integral closure $S \subset L$ is Dedekind too (though in general it is not a DVR—it may have several non-zero prime ideals) by Theorem 6.3.1. Now as above, S is a finitely generated R-module, so if $\mathfrak{m} \subset R$ is the maximal ideal, then

$$\mathfrak{m}S \neq S$$

by Nakayama's lemma (cf. for instance **?**). So $\mathfrak{m}S$ is contained in a maximal ideal $\mathfrak{M}$ of S with, therefore, $\mathfrak{M} \cap R = \mathfrak{m}$. (This is indeed the basic argument behind lying over, which I could have just invoked.) Now $S_{\mathfrak{M}} \supset R_{\mathfrak{m}}$ is a DVR as it is the localization of a Dedekind domain at a prime ideal, and one can appeal to **??**. So there is a discrete valuation on $S_{\mathfrak{M}}$. Restricted to R, it will be a power of the given R-valuation, because its value on a uniformizer π is < 1. However, a power of a discrete valuation is a discrete valuation too. So we can adjust the discrete valuation on $S_{\mathfrak{M}}$ if necessary to make it an extension.

This completes the proof. ∎

Note that there is a one-to-one correspondence between extensions of the valuation on K and primes of S lying above $\mathfrak{m}$. Indeed, the above proof indicated a way of getting valuations on L from primes of S. For an extension of the valuation on K to L, let $\mathfrak{M} := \{x \in S : |x| < 1\}$.

6.4. Action of the Galois group

Suppose we have an integral domain (we don't even have to assume it Dedekind) A with quotient field K, a finite Galois extension L/K, with B the integral closure in L. Then the Galois group $G = G(L/K)$ acts on B; it preserves B because it preserves equations in $A[X]$. In particular, if $\mathfrak{P} \subset B$ is a prime ideal, so is $\sigma\mathfrak{P}$, and the set $\operatorname{Spec} B$ of prime ideals in B becomes a G-set.

The orbits of the Galois group

It is of interest to determine the orbits; this question has a very clean answer.

6.4.1 Proposition *The orbits of G on the prime ideals of B are in bijection with the primes of A, where a prime ideal $\mathfrak{p} \subset A$ corresponds to the set of primes of B lying over A.*[2] *Alternatively, any two primes $\mathfrak{P}, \mathfrak{Q} \subset B$ lying over A are conjugate by some element of G.*

In other words, under the natural map $\operatorname{Spec} B \to \operatorname{Spec} A = \operatorname{Spec} B^G$, the latter space is the quotient under the action of G, while $A = B^G$ is the ring of invariants in B.[3]

Proof. We need only prove the second statement. Let S be the multiplicative set $A - \mathfrak{p}$. Then $S^{-1}B$ is the integral closure of $S^{-1}A$, and in $S^{-1}A = A_{\mathfrak{p}}$, the ideal $\mathfrak{p}$ is maximal. Let $\mathfrak{Q}, \mathfrak{P}$ lie over $\mathfrak{p}$; then $S^{-1}\mathfrak{Q}, S^{-1}\mathfrak{P}$ lie over $S^{-1}\mathfrak{p}$ and are maximal (to be added). If we prove that $S^{-1}\mathfrak{Q}, S^{-1}\mathfrak{P}$ are conjugate under the Galois group, then $\mathfrak{Q}, \mathfrak{P}$ must also be conjugate by the properties of localization. *In particular, we can reduce to the case of* $\mathfrak{p}, \mathfrak{Q}, \mathfrak{P}$ *all maximal.*

The rest of the proof is now an application of the Chinese remainder theorem. Suppose that, for all $\sigma \in G$, we have $\sigma\mathfrak{P} \neq \mathfrak{Q}$. Then the ideals $\sigma\mathfrak{P}, \mathfrak{Q}$ are distinct maximal ideals, so by the remainder theorem, we can find $x \equiv 1 \mod \sigma\mathfrak{P}$ for all $\sigma \in G$ and $x \equiv 0 \mod \mathfrak{Q}$. Now, consider the norm $N_K^L(x)$; the first condition implies that it is congruent to 1 modulo $\mathfrak{p}$. But the second implies that the norm is in $\mathfrak{Q} \cap K = \mathfrak{p}$, contradiction. ∎

The decomposition and inertia groups

Now, let's zoom in on a given prime $\mathfrak{p} \subset A$. We know that G acts transitively on the set $\mathfrak{P}_1, \dots, \mathfrak{P}_g$ of primes lying above $\mathfrak{p}$; in particular, there are at most $[L : K]$ of them.

6.4.2 Definition If $\mathfrak{P}$ is any one of the $\mathfrak{P}_i$, then the stabilizer in G of this prime ideal is called the **decomposition group** $G_{\mathfrak{P}}$.

We have, clearly, $(G : G_{\mathfrak{P}}) = g$.

Now if $\sigma \in G_{\mathfrak{P}}$, then σ acts on the residue field $B/\mathfrak{P}$ while fixing the subfield $A/\mathfrak{p}$. In this way, we get a homomorphism $\sigma \to \overline{\sigma}$ from G into the automorphism group of $B/\mathfrak{P}$ over $A/\mathfrak{p}$) (we don't call it a Galois group because we don't yet know whether the extension is Galois).

The following result will be crucial in constructing the so-called "Frobenius elements" of crucial use in class field theory.

6.4.3 Proposition *Suppose $A/\mathfrak{p}$ is perfect. Then $B/\mathfrak{P}$ is Galois over $A/\mathfrak{p}$, and the homomorphism $\sigma \to \overline{\sigma}$ is surjective from $G_{\mathfrak{P}} \to G(B/\mathfrak{P}/A/\mathfrak{p})$.*

[2] It is useful to note here that the lying over theorem works for arbitrary integral extensions.
[3] The reader who does not know about the Spec of a ring can disregard these remarks.

Proof. In this case, the extension $B/\mathfrak{P}/A/\mathfrak{p}$ is separable, and we can choose $\overline{x} \in B/\mathfrak{P}$ generating it by the primitive element theorem. We will show that $\overline{x}$ satisfies a polynomial equation $\overline{P}(X) \in A/\mathfrak{p}[X]$ all of whose roots lie in $B/\mathfrak{P}$, which will prove that the residue field extension is Galois. Moreover, we will show that all the nonzero roots of $\overline{P}$ in $B/\mathfrak{P}$ are conjugates of $\overline{x}$ under elements of $G_{\mathfrak{P}}$. This latter will imply surjectivity of the homomorphism $\sigma \to \overline{\sigma}$, because it shows that any conjugate of $\overline{x}$ under $G(B/\mathfrak{P}/A/\mathfrak{p})$ is a conjugate under $G_{\mathfrak{P}}$.

We now construct the aforementioned polynomial. Let $x \in B$ lift $\overline{x}$. Choose $y \in B$ such that $y \equiv x \mod \mathfrak{P}$ but $y \equiv 0 \mod \mathfrak{Q}$ for the other primes $\mathfrak{Q}$ lying over $\mathfrak{p}$. We take $P(X) = \prod_{\sigma \in G}(X - \sigma(y)) \in A[X]$. Then the reduction $\overline{P}$ satisfies $\overline{P}(\overline{x}) = \overline{P}(\overline{y}) = 0$, and $\overline{P}$ factors completely (via $\prod_{\sigma}(X - \overline{\sigma(t)})$) in $B/\mathfrak{P}[X]$. This implies that the residue field extension is Galois, as already stated. But it is also clear that the polynomial $\overline{P}(X)$ has roots of zero and $\sigma(\overline{y}) = \sigma(\overline{x})$ for $\sigma \in G_{\mathfrak{P}}$. This completes the proof of the other assertion, and hence the proposition. □

6.4.4 Definition The kernel of the map $\sigma \to \overline{\sigma}$ is called the **inertia group** $T_{\mathfrak{P}}$. Its fixed field is called the **inertia field**.

These groups will resurface significantly in the future.

6.4.5 Remark Although we shall never need this in the future, it is of interest to see what happens when the extension L/K is *purely inseparable.*[4] Suppose A is integrally closed in K, and B is the integral closure in L. Let the characteristic be p, and the degree $[L : K] = p^i$. In this case, $x \in B$ if and only if $x^{p^i} \in A$. Indeed, it is clear that the condition mentioned implies integrality. Conversely, if x is integral, then so is x^{p^i}, which belongs to K (by basic facts about purely inseparable extensions). Since A is integrally closed, it follows that $x^{p^i} \in A$.

Let now $\mathfrak{p} \subset A$ be a prime ideal. I claim that there is precisely one prime ideal $\mathfrak{P}$ of B lying above A, and $\mathfrak{P}^{p^i} = \mathfrak{p}$. Namely, this ideal consists of $x \in B$ with $x^{p^i} \in \mathfrak{p}$! The proof is straightforward; if $\mathfrak{P}$ is *any* prime ideal lying over $\mathfrak{p}$, then $x \in \mathfrak{P}$ iff $x^{p^i} \in L \cap \mathfrak{P} = \mathfrak{p}$. In a terminology to be explained later, $\mathfrak{p}$ is *totally ramified.*

[4] Cf. **?**, for instance.

III.7. Dimension theory

Dimension theory assigns to each commutative ring—say, noetherian—an invariant called the dimension. The most standard definition, that of Krull dimension (which we shall not adopt at first), defines the dimension in terms of the maximal lengths of ascending chains of prime ideals. In general, however, the geometric intuition behind dimension is that it should assign to an affine ring—say, one of the form $\mathbb{C}[x_1, \ldots, X_n]/I$—something like the "topological dimension" of the affine variety in $\mathbb{C}^n$ cut out by the ideal I.

In this chapter, we shall obtain three different expressions for the dimension of a noetherian local ring $(R, \mathfrak{m})$, each of which will be useful at different times in proving results.

7.1. The Hilbert function and the dimension of a local ring

Integer-valued polynomials

It is now necessary to do a small amount of general algebra.

Let $P \in \mathbb{Q}[t]$. We consider the question of when P maps the integers $\mathbb{Z}$, or more generally the sufficiently large integers, into $\mathbb{Z}$. Of course, any polynomial in $\mathbb{Z}[t]$ will do this, but there are others: consider $\frac{1}{2}(t^2 - t)$, for instance.

7.1.1 Proposition *Let $P \in \mathbb{Q}[t]$. Then $P(m)$ is an integer for $m \gg 0$ integral if and only if P can be written in the form*

$$P(t) = \sum_n c_n \binom{t}{n}, \quad c_n \in \mathbb{Z}.$$

In particular, $P(\mathbb{Z}) \subset \mathbb{Z}$.

So P is a $\mathbb{Z}$-linear function of binomial coefficients.

Proof. Note that the set $\left\{\binom{t}{n}\right\}_{n \in \mathbb{Z}_{\geqslant 0}}$ forms a basis for the set of polynomials $\mathbb{Q}[t]$. It is thus clear that $P(t)$ can be written as a rational combination $\sum c_n \binom{t}{n}$ for the $c_n \in \mathbb{Q}$. We need to argue that the $c_n \in \mathbb{Z}$ in fact.

Consider the operator Δ defined on functions $\mathbb{Z} \to \mathbb{C}$ as follows:

$$(\Delta f)(m) = f(m) - f(m-1).$$

It is obvious that if f takes integer values for $m \gg 0$, then so does Δf. It is also easy to check that $\Delta \binom{t}{n} = \binom{t}{n-1}$.

By looking at the function $\Delta P = \sum c_n \binom{t}{n-1}$ (which takes values in $\mathbb{Z}$), it is easy to see that the $c_n \in \mathbb{Z}$ by induction on the degree. It is also easy to see directly that the binomial coefficients take values in $\mathbb{Z}$ at *all* arguments. ☐

Definition and examples

Let R be a ring.

7.1.2 Remark (Question) What is a good definition for $\dim(R)$? Actually, more generally, what is the dimension of R at a given "point" (i.e. prime ideal)?

Geometrically, think of $\operatorname{Spec} R$, for any ring; pick some point corresponding to a maximal ideal $\mathfrak{m} \subset R$. We want to define the **dimension of** R at $\mathfrak{m}$. This is to be thought of kind of like "dimension over the complex numbers," for algebraic varieties defined over $\mathbb{C}$. But it should be purely algebraic. What might you do?

Here is an idea. For a topological space X to be n-dimensional at $x \in X$, there should be n coordinates at the point x. In other words, the point x should be uniquely defined by the zero locus of n points on the space. Motivated by this, we could try defining $\dim_{\mathfrak{m}} R$ to be the number of generators of $\mathfrak{m}$. However, this is a bad definition, as $\mathfrak{m}$ may not have the same number of generators as $\mathfrak{m}R_{\mathfrak{m}}$. In other words, it is not a truly *local* definition.

7.1.3 Example Let R be a noetherian integrally closed domain which is not a UFD. Let $\mathfrak{p} \subset R$ be a prime ideal which is minimal over a principal ideal but which is not itself principal. Then $\mathfrak{p}R_{\mathfrak{p}}$ is generated by one element, as we will eventually see, but $\mathfrak{p}$ is not.

We want our definition of dimension to be local. So this leads us to:

7.1.4 Definition If R is a (noetherian) *local* ring with maximal ideal $\mathfrak{m}$, then the **embedding dimension** of R, denoted $\operatorname{Emdim} R$ is the minimal number of generators for $\mathfrak{m}$. If R is a noetherian ring and $\mathfrak{p} \subset R$ a prime ideal, then the **embedding dimension at** $\mathfrak{p}$ is that of the local ring $R_{\mathfrak{p}}$.

In the above definition, it is clearly sufficient to study what happens for local rings, and we impose that restriction for now. By Nakayama's lemma, the embedding dimension is the minimal number of generators of $\mathfrak{m}/\mathfrak{m}^2$, or the $R/\mathfrak{m}$-dimension of that vector space:

$$\operatorname{Emdim} R = \dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2.$$

In general, however, the embedding dimension is not going to coincide with the intuitive "geometric" dimension of an algebraic variety.

7.1.5 Example Let $R = \mathbb{C}[t^2, t^3] \subset \mathbb{C}[t]$, which is the coordinate ring of a cubic curve $y^2 = x^3$ as $R \simeq \mathbb{C}[x, y]/(x^2 - y^3)$ via $x = t^3, y = t^2$. Let us localize at the prime ideal $\mathfrak{p} = (t^2, t^3)$: we get $R_{\mathfrak{p}}$.

Now $\operatorname{Spec} R$ is singular at the origin. In fact, as a result, $\mathfrak{p}R_{\mathfrak{p}} \subset R_{\mathfrak{p}}$ needs two generators, but the variety it corresponds to is one-dimensional.

So the embedding dimension is the smallest dimension into which you can embed R into a smooth space. But for singular varieties this is not the dimension we want.

So instead of considering simply $\mathfrak{m}/\mathfrak{m}^2$, let us consider the *sequence* of finite-dimensional vector spaces

$$\mathfrak{m}^k/\mathfrak{m}^{k+1}.$$

Computing these dimensions as a function of k gives some invariant that describes the local geometry of $\operatorname{Spec} R$.

We shall eventually prove:

7.1.6 Theorem *Let $(R, \mathfrak{m})$ be a local noetherian ring. Then there exists a polynomial $f \in \mathbb{Q}[t]$ such that*

$$f(n) = \ell(R/\mathfrak{m}^n) = \sum_{i=0}^{n-1} \dim \mathfrak{m}^i/\mathfrak{m}^{i+1} \quad \forall n \gg 0.$$

Moreover, $\deg f \leqslant \dim \mathfrak{m}/\mathfrak{m}^2$.

Note that this polynomial is well-defined, as any two polynomials agreeing for large n coincide. Note also that $R/\mathfrak{m}^n$ is artinian so of finite length, and that we have used the fact that the length is additive for short exact sequences. We would have liked to write $\dim R/\mathfrak{m}^n$, but we can't, in general, so we use the substitute of the length.

Based on this, we define:

7.1.7 Definition The **dimension** of the local ring R is the degree of the polynomial f above. For an arbitrary noetherian ring R, we define $\dim R = \sup_{\mathfrak{p} \in \operatorname{Spec} R} \dim(R_{\mathfrak{p}})$.

Let us now do a few example computations.

7.1.8 Example (The affine line) Consider the local ring $(R, \mathfrak{m}) = \mathbb{C}[t]_{(t)}$. Then $\mathfrak{m} = (t)$ and $\mathfrak{m}^k/\mathfrak{m}^{k+1}$ is one-dimensional, generated by t^k. In particular, the ring has dimension one.

7.1.9 Example (A singular curve) Consider $R = \mathbb{C}[t^2, t^3]_{(t^2,t^3)}$, the local ring of $y^2 = x^3$ at zero. Then $\mathfrak{m}^n$ is generated by $t^{2n}, t^{2n+1}, \dots$. $\mathfrak{m}^{n+1}$ is generated by $t^{2n+2}, t^{2n+3}, \dots$. So the quotients all have dimension two. The dimension of these quotients is a little larger than in Example 7.1.8, but they do not grow. The ring still has dimension one.

7.1.10 Example (The affine plane) Consider $R = \mathbb{C}[x, y]_{(x,y)}$. Then $\mathfrak{m}^k$ is generated by polynomials in x, y that are homogeneous in degree k. So $\mathfrak{m}^k/\mathfrak{m}^{k+1}$ has dimensions that *grow* linearly in k. This is a genuinely two-dimensional example.

It is this difference between constant linear and quadratic growth in $R/\mathfrak{m}^n$ as $n \to \infty$, and not the size of the initial terms, that we want for our definition of dimension.

Let us now generalize Example 7.1.8 and Example 7.1.10 above to affine spaces of arbitrary dimension.

7.1.11 Example (Affine space) Consider $R = \mathbb{C}[x_1, \dots, x_n]_{(x_1,\dots,x_n)}$. This represents the variety $\mathbb{C}^n = \mathbb{A}^n_{\mathbb{C}}$ near the origin geometrically, so it should intuitively have dimension n. Let us check that it does.

Namely, we need to compute the polynomial f above. Here $R/\mathfrak{m}^k$ looks like the set of polynomials of degree $< k$ over $\mathbb{C}$. The dimension as a vector space of this is given by some binomial coefficient $\binom{n+k-1}{n}$. This is a polynomial in k of degree n. In particular, $\ell(R/\mathfrak{m}^k)$ grows like k^n. So R is n-dimensional.

Finally, we offer one more example, showing that DVRs have dimension one. In fact, among noetherian integrally closed local domains, DVRs are *characterized* by this property (**??** of **??**).

7.1.12 Example (The dimension of a DVR) Let R be a DVR. Then $\mathfrak{m}^k/\mathfrak{m}^{k+1}$ is of length one for each k. So $R/\mathfrak{m}^k$ has length k. Thus we can take $f(t) = t$, so R has dimension one.

The Hilbert function is a polynomial

While we have given a definition of dimension and computed various examples, we have yet to check that our definition is well-defined. Namely, we have to prove Theorem 7.1.6.

Proof of Theorem 7.1.6. Fix a noetherian local ring $(R, \mathfrak{m})$. We are to show that $\ell(R/\mathfrak{m}^n)$ is a polynomial for $n \gg 0$. We also have to bound this degree by $\dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2$, the embedding dimension. We will do this by reducing to a general fact about graded modules over a polynomial ring.

Let $S = \bigoplus_n \mathfrak{m}^n/\mathfrak{m}^{n+1}$. Then S has a natural grading, and in fact it is a graded ring in a natural way from the multiplication map

$$\mathfrak{m}^{n_1} \times \mathfrak{m}^{n_2} \to \mathfrak{m}^{n_1+n_2}.$$

In fact, S is the *associated graded ring* of the $\mathfrak{m}$-adic filtration. Note that $S_0 = R/\mathfrak{m}$ is a field, which we will denote by k. So S is a graded k-algebra.

7.1.13 Lemma *S is a finitely generated k-algebra. In fact, S can be generated by at most* $\mathrm{Emdim}(R)$ *elements.*

Proof. Let $x_1, \dots, x_r$ be generators for $\mathfrak{m}$ with $r = \mathrm{Emdim}(R)$. They (or rather, their images) are thus a k-basis for $\mathfrak{m}/\mathfrak{m}^2$. Then their images in $\mathfrak{m}/\mathfrak{m}^2 \subset S$ generate S. This follows because S_1 generates S as an S_0-algebra: the products of the elements in $\mathfrak{m}$ generate the higher powers of $\mathfrak{m}$. ☐

So S is a graded quotient of the polynomial ring $k[t_1, \dots, t_r]$, with t_i mapping to x_i. In particular, S is a finitely generated, graded $k[t_1, \dots, t_r]$-module. Note that also $\ell(R/\mathfrak{m}^n) = \dim_k(S_0) + \cdots + \dim_k(S_{n-1})$ for any n, thanks to the filtration. This is the invariant we are interested in.

It will now suffice to prove the following more general proposition.

7.1.14 Proposition *Let M be any finitely generated graded module over the polynomial ring $k[x_1, \dots, x_r]$. Then there exists a polynomial $f_M^+ \in \mathbb{Q}[t]$ of degree $\leqslant r$, such that*

$$f_M^+(t) = \sum_{s \leqslant t} \dim M_s \quad t \gg 0.$$

Applying this to $M = S$ will give the desired result. We can forget about everything else, and look at this problem over graded polynomial rings.

This function is called the **Hilbert function**.

Proof of Proposition 7.1.14. Note that if we have an exact sequence of graded modules over the polynomial ring,

$$0 \to M' \to M \to M'' \to 0,$$

and polynomials $f_{M'}, f_{M''}$ as in the proposition, then f_M exists and

$$f_M = f_{M'} + f_{M''}.$$

This is obvious from the definitions. Next, we observe that if M is a finitely generated graded module, over two different polynomial rings, but with the same grading, then the existence (and value) of f_M is independent of which polynomial ring one considers. Finally, we observe that it is sufficient to prove that $f_M(t) = \dim M_t$ is a polynomial in t for $t \gg 0$.

We will use these three observations and induct on n.

If $n = 0$, then M is a finite-dimensional graded vector space over k, and the grading must be concentrated in finitely many degrees. Thus the result is evident as $f_M(t)$ will just equal $\dim M$ (which will be the appropriate dimension for $t \gg 0$).

Suppose $n > 0$. Then consider the filtration of M

$$0 \subset \ker(x_1 : M \to M) \subset \ker(x_1^2 : M \to M) \subset \dots \subset M.$$

This must stabilize by noetherianness at some $M' \subset M$. Each of the quotients $\ker(x_1^i)/\ker(x_1^{i+1})$ is a finitely generated module over $k[x_1, \dots, x_n]/(x_1)$, which is a smaller polynomial ring. So each of these quotients $\ker(x_1^{i+1})/\ker(x_1^i)$ has a Hilbert function of degree $\leqslant n-1$ by the inductive hypothesis.

Climbing up the filtration, we see that M' has a Hilbert function which is the sum of the Hilbert functions of these quotients $\ker(x_1^{i+1})/\ker(x_1^i)$. In particular, $f_{M'}$ exists. If we show that $f_{M/M'}$ exists, then f_M necessarily exists. So we might as well show that the Hilbert function f_M exists when x_1 is a non-zero-divisor on M.

So, we have reduced to the case where $M \xrightarrow{x_1} M$ is injective. Now M has a filtration

$$M \supset x_1 M \supset x_1^2 M \supset \dots$$

which is an exhaustive filtration of M in that nothing can be divisible by powers of x_1 over and over, or the degree would not be finite. So it follows that $\bigcap x_1^m M = 0$.

Let $N = M/x_1M$, which is isomorphic to $x_1^m M/x_1^{m+1}M$ since $M \xrightarrow{x_1} M$ is injective. Here N is a finitely generated graded module over $k[x_2, \dots, x_n]$, and by the inductive hypothesis on n, we see that there is a polynomial f_N^+ of degree $\leqslant n-1$ such that

$$f_N^+(t) = \sum_{t' \leqslant t} \dim N_{t'}, \quad t \gg 0.$$

Fix $t \gg 0$ and consider the k-vector space M_t, which has a finite filtration

$$M_t \supset (x_1M)_t \supset (x_1^2M)_t \supset \dots$$

which has successive quotients that are the graded pieces of $N \simeq M/x_1M \simeq x_1M/x_1^2M \simeq \dots$ in dimensions $t, t-1, \dots$. We find that

$$(x_1^2M)_t/(x_1^3M)_t \simeq N_{t-2},$$

for instance. Summing this, we find that

$$\dim M_t = \dim N_t + \dim N_{t-1} + \dots.$$

The sum above is actually finite. In fact, by finite generation, there is $K \gg 0$ such that $\dim N_q = 0$ for $q < -K$. From this, we find that

$$\dim M_t = \sum_{t'=-K}^{t} \dim N_{t'},$$

□

which implies that $\dim M_t$ is a polynomial for $t \gg 0$. This completes the proof. □

Let $(R, \mathfrak{m})$ a noetherian local ring and M a finitely generated R-module.

7.1.15 Proposition *$\ell(M/\mathfrak{m}^m M)$ is a polynomial for $m \gg 0$.*

Proof. This follows from Proposition 7.1.14, and in fact we have essentially seen the argument above. Indeed, we consider the associated graded module

$$N = \bigoplus \mathfrak{m}^k M/\mathfrak{m}^{k+1}M,$$

which is finitely generated over the associated graded ring

$$\bigoplus \mathfrak{m}^k/\mathfrak{m}^{k+1}.$$

□

Consequently, the graded pieces of N have dimensions growing polynomially for large degrees. This implies the result.

7.1.16 Definition We define the **Hilbert function** $H_M(m)$ to be the unique polynomial such that

$$H_M(m) = \ell(M/\mathfrak{m}^m M), \quad m \gg 0.$$

It is clear, incidentally, that H_M is integer-valued, so we see by Proposition 7.1.1 that H_M is a $\mathbb{Z}$-linear combination of binomial coefficients.

The dimension of a module

Let R be a local noetherian ring with maximal ideal $\mathfrak{m}$. We have seen (Proposition 7.1.15) that there is a polynomial $H(t)$ with

$$H(t) = \ell(R/\mathfrak{m}^t), \quad t \gg 0.$$

Earlier, we defined the **dimension** of R is the degree of f_M^+. Since the degree of the Hilbert function is at most the number of generators of the polynomial ring, we saw that

$$\dim R \leqslant \operatorname{Emdim} R.$$

Armed with the machinery of the Hilbert function, we can extend this definition to modules.

7.1.17 Definition If R is local noetherian, and N a finite R-module, then N has a Hilbert polynomial $H_N(t)$ which when evaluated at $t \gg 0$ gives the length $\ell(N/\mathfrak{m}^t N)$. We say that the **dimension of** N is the degree of this Hilbert polynomial.

Clearly, the dimension of the *ring* R is the same thing as that of the *module* R.

We next show that the dimension behaves well with respect to short exact sequences. This is actually slightly subtle since, in general, tensoring with $R/\mathfrak{m}^t$ is not exact; it turns out to be *close* to being exact by the Artin-Rees lemma. On the other hand, the corresponding fact for modules over a *polynomial ring* is very easy, as no tensoring was involved in the definition.

7.1.18 Proposition *Suppose we have an exact sequence*

$$0 \to M' \to M \to M'' \to 0$$

of graded modules over a polynomial ring $k[x_1, \dots, x_n]$*. Then*

$$f_M(t) = f_{M'}(t) + f_{M''}(t), \quad f_M^+(t) = f_{M'}^+(t) + f_{M''}^+(t).$$

As a result, $\deg f_M = \max \deg f_{M'}, \deg f_{M''}$.

Proof. The first part is obvious as the dimension is additive on vector spaces. The second part follows because Hilbert functions have nonnegative leading coefficients. ☐

7.1.19 Proposition *Fix an exact sequence*

$$0 \to N' \to N \to N'' \to 0$$

of finite R-modules. Then $\dim N = \max(\dim N', \dim N'')$.

Proof. We have an exact sequence

$$0 \to K \to N/\mathfrak{m}^t N \to N''/\mathfrak{m}^t N'' \to 0$$

where K is the kernel. Here $K = (N' + \mathfrak{m}^t N)/\mathfrak{m}^t N = N'/(N' \cap \mathfrak{m}^t N)$. This is not quite $N'/\mathfrak{m}^t N'$, but it's pretty close. We have a surjection

$$N'/\mathfrak{m}^t N \twoheadrightarrow N'/(N' \cap \mathfrak{m}^t N) = K.$$

In particular,

$$\ell(K) \leqslant \ell(N'/\mathfrak{m}^t N').$$

On the other hand, we have the Artin-Rees lemma, which gives an inequality in the opposite direction. We have a containment

$$\mathfrak{m}^t N' \subset N' \cap \mathfrak{m}^t N \subset \mathfrak{m}^{t-c} N'$$

for some c. This implies that $\ell(K) \geqslant \ell(N'/\mathfrak{m}^{t-c}N')$.

Define $M = \bigoplus \mathfrak{m}^t N/\mathfrak{m}^{t+1}N$, and define M', M'' similarly in terms of N', N''. Then we have seen that

$$\boxed{f_M^+(t-c) \leqslant \ell(K) \leqslant f_M^+(t).}$$

We also know that the length of K plus the length of $N''/\mathfrak{m}^t N''$ is $f_M^+(t)$, i.e.

$$\ell(K) + f_{M''}^+(t) = f_M^+(t).$$

Now the length of K is a polynomial in t which is pretty similar to $f_{M'}^+$, in that the leading coefficient is the same. So we have an approximate equality $f_{M'}^+(t) + f_{M''}^+(t) \simeq f_M^+(t)$. This implies the result since the degree of f_M^+ is $\dim N$ (and similarly for the others). □

7.1.20 Proposition $\dim R$ *is the same as* $\dim R/\operatorname{Rad} R$.

I.e., the dimension doesn't change when you kill off nilpotent elements, which is what you would expect, as nilpotents don't affect $\operatorname{Spec}(R)$.

Proof. For this, we need a little more information about Hilbert functions. We thus digress substantially.

Finally, let us return to the claim about dimension and nilpotents. Let R be a local noetherian ring and $I = \operatorname{Rad}(R)$. Then I is a finite R-module. In particular, I is nilpotent, so $I^n = 0$ for $n \gg 0$. We will show that

$$\dim R/I = \dim R/I^2 = \dots$$

which will imply the result, as eventually the powers become zero.

In particular, we have to show for each k,

$$\dim R/I^k = \dim R/I^{k+1}.$$

There is an exact sequence

$$0 \to I^k/I^{k+1} \to R/I^{k+1} \to R/I^k \to 0.$$

The dimension of these rings is the same thing as the dimensions as R-modules. So we can use this short exact sequence of modules. By the previous result, we are reduced to showing that

$$\dim I^k/I^{k+1} \leqslant \dim R/I^k.$$

Well, note that I kills I^k/I^{k+1}. In particular, I^k/I^{k+1} is a finitely generated R/I^k-module. There is an exact sequence

$$\bigoplus_N R/I^k \to I^k/I^{k+1} \to 0$$

which implies that $\dim I^k/I^{k+1} \leqslant \dim \bigoplus_N R/I^k = \dim R/I^k$. □

7.1.21 Example Let $\mathfrak{p} \subset \mathbb{C}[x_1, \ldots, x_n]$ and let $R = (\mathbb{C}[x_1, \ldots, x_n]/\mathfrak{p})_{\mathfrak{m}}$ for some maximal ideal $\mathfrak{m}$. What is $\dim R$? What does dimension mean for coordinate rings over $\mathbb{C}$?

Recall by the Noether normalization theorem that there exists a polynomial ring $\mathbb{C}[y_1, \ldots, y_m]$ contained in $S = \mathbb{C}[x_1, \ldots, x_n]/\mathfrak{p}$ and S is a finite integral extension over this polynomial ring. We claim that

$$\dim R = m.$$

There is not sufficient time for that today.

Dimension depends only on the support

Let $(R, \mathfrak{m})$ be a local noetherian ring. Let M be a finitely generated R-module. We defined the **Hilbert polynomial** of M to be the polynomial which evaluates at $t \gg 0$ to $\ell(M/\mathfrak{m}^t M)$. We proved last time that such a polynomial always exists, and called its degree the **dimension of** M. However, we shall now see that $\dim M$ really depends only on the support[1] $\operatorname{supp} M$. In this sense, the dimension is really a statement about the *topological space* $\operatorname{supp} M \subset \operatorname{Spec} R$, not about M itself.

In other words, we will prove:

7.1.22 Proposition $\dim M$ *depends only on* $\operatorname{supp} M$*.*

In fact, we shall show:

7.1.23 Proposition $\dim M = \max_{\mathfrak{p} \in \operatorname{supp} M} \dim R/\mathfrak{p}$.

Proof. By Proposition 2.2.12 in Chapter III.2, there is a finite filtration

$$0 = M_0 \subset M_1 \subset \cdots \subset M_m = M,$$

such that each of the successive quotients is isomorphic to $R/\mathfrak{p}_i \subset R$ for some prime ideal $\mathfrak{p}_i$. Given a short exact sequence of modules, we know that the dimension in the middle is the maximum of the dimensions at the two ends (Proposition 7.1.19). Iterating this, we see that the dimension of M is the maximum of the dimension of the successive quotients M_i/M_{i-1}.

But the $\mathfrak{p}_i$'s that occur are all in $\operatorname{supp} M$, so we find

$$\dim M = \max_{\mathfrak{p}_i} R/\mathfrak{p}_i \leqslant \max_{\mathfrak{p} \in \operatorname{supp} M} \dim R/\mathfrak{p}.$$

We must show the reverse inequality. But fix any prime $\mathfrak{p} \in \operatorname{supp} M$. Then $M_{\mathfrak{p}} \neq 0$, so one of the $R/\mathfrak{p}_i$ localized at $\mathfrak{p}$ must be nonzero, as localization is an exact functor. Thus $\mathfrak{p}$ must contain some $\mathfrak{p}_i$. So $R/\mathfrak{p}$ is a quotient of $R/\mathfrak{p}_i$. In particular,

$$\dim R/\mathfrak{p} \leqslant \dim R/\mathfrak{p}_i.$$

□

Having proved this, we throw out the notation $\dim M$, and henceforth write instead $\dim \operatorname{supp} M$.

[1] Recall that $\operatorname{supp} M = \{\mathfrak{p} : M_{\mathfrak{p}} \neq 0\}$.

7.1.24 Remark (comment) The dimension of an affine ring

Last time, we made a claim. If R is a domain and a finite module over a polynomial ring $k[x_1, \dots, x_n]$, then $R_{\mathfrak{m}}$ for any maximal $\mathfrak{m} \subset R$ has dimension n. This connects the dimension with the transcendence degree.

First, let us talk about finite extensions of rings. Let R be a commutative ring and let $R \to R'$ be a morphism that makes R' a finitely generated R-module (in particular, integral over R). Let $\mathfrak{m}' \subset R'$ be maximal. Let $\mathfrak{m}$ be the pull-back to R, which is also maximal (as $R \to R'$ is integral). Let M be a finitely generated R'-module, hence also a finitely generated R-module.

We can look at $M_{\mathfrak{m}}$ as an $R_{\mathfrak{m}}$-module or $M_{\mathfrak{m}'}$ as an $R'_{\mathfrak{m}'}$-module. Either of these will be finitely generated.

7.1.25 Proposition $\dim \operatorname{supp} M_{\mathfrak{m}} \geqslant \dim \operatorname{supp} M_{\mathfrak{m}'}$.

Here $M_{\mathfrak{m}}$ is an $R_{\mathfrak{m}}$-module, $M_{\mathfrak{m}'}$ is an $R'_{\mathfrak{m}'}$-module.

Proof. Consider $R/\mathfrak{m} \to R'/\mathfrak{m}R' \to R'/\mathfrak{m}'$. Then we see that $R'/\mathfrak{m}R'$ is a finite $R/\mathfrak{m}$-module, so a finite-dimensional $R/\mathfrak{m}$-vector space. In particular, $R'/\mathfrak{m}R'$ is of finite length as an $R/\mathfrak{m}$-module, in particular an artinian ring. It is thus a product of local artinian rings. These artinian rings are the localizations of $R'/\mathfrak{m}R'$ at ideals of R' lying over $\mathfrak{m}$. One of these ideals is $\mathfrak{m}'$. So in particular

$$R'/\mathfrak{m}R \simeq R'/\mathfrak{m}' \times \text{other factors}.$$

The nilradical of an artinian ring being nilpotent, we see that $\mathfrak{m}'^c R'_{\mathfrak{m}'} \subset \mathfrak{m}R'_{\mathfrak{m}}$ for some c.

OK, I'm not following this—too tired. Will pick this up someday. □

7.1.26 Proposition $\dim \operatorname{supp} M_{\mathfrak{m}} = \max_{\mathfrak{m}'|\mathfrak{m}} \dim \operatorname{supp} M_{\mathfrak{m}'}$.

This means $\mathfrak{m}'$ lies over $\mathfrak{m}$.

Proof. Done similarly, using artinian techniques. I'm kind of tired. □

7.1.27 Example Let $R' = \mathbb{C}[x_1, \dots, x_n]/\mathfrak{p}$. Noether normalization says that there exists a finite injective map $\mathbb{C}[y_1, \dots, y_a] \to R'$. The claim is that

$$\dim R'_{\mathfrak{m}} = a$$

for any maximal ideal $\mathfrak{m} \subset R'$. We are set up to prove a slightly weaker definition. In particular (see below for the definition of the dimension of a non-local ring), by the proposition, we find the weaker claim

$$\dim R' = a,$$

as the dimension of a polynomial ring $\mathbb{C}[y_1, \dots, y_a]$ is a. (**I don't think we have proved this yet.**)

7.2. Other definitions and characterizations of dimension

The topological characterization of dimension

We now want a topological characterization of dimension. So, first, we want to study how dimension changes as we do things to a module. Let M be a finitely generated R-module over a local noetherian ring R. Let $x \in \mathfrak{m}$ for $\mathfrak{m}$ as the maximal ideal. You might ask

> What is the relation between $\dim \operatorname{supp} M$ and $\dim \operatorname{supp} M/xM$?

Well, M surjects onto M/xM, so we have the inequality $\geqslant$. But we think of dimension as describing the number of parameters you need to describe something. The number of parameters shouldn't change too much with going from M to M/xM. Indeed, as one can check,

$$\operatorname{supp} M/xM = \operatorname{supp} M \cap V(x)$$

and intersecting $\operatorname{supp} M$ with the "hypersurface" $V(x)$ should shrink the dimension by one.

We thus make:

7.2.1 Remark (Prediction)

$$\dim \operatorname{supp} M/xM = \dim \operatorname{supp} M - 1.$$

Obviously this is not always true, e.g. if x acts by zero on M. But we want to rule that out. Under reasonable cases, in fact, the prediction is correct:

7.2.2 Proposition *Suppose $x \in \mathfrak{m}$ is a non-zero-divisor on M. Then*

$$\dim \operatorname{supp} M/xM = \dim \operatorname{supp} M - 1.$$

Proof. To see this, we look at Hilbert polynomials. Let us consider the exact sequence

$$0 \to xM \to M \to M/xM \to 0$$

which leads to an exact sequence for each t,

$$0 \to xM/(xM \cap \mathfrak{m}^t M) \to M/\mathfrak{m}^t M \to M/(xM + \mathfrak{m}^t M) \to 0.$$

For t large, the lengths of these things are given by Hilbert polynomials, as the thing on the right is $M/xM \otimes_R R/\mathfrak{m}^t$. We have

$$f_M^+(t) = f_{M/xM}^+(t) + \ell(xM/(xM \cap \mathfrak{m}^t M), \quad t \gg 0.$$

In particular, $\ell(xM/(xM \cap \mathfrak{m}^t M))$ is a polynomial in t. What can we say about it? Well, $xM \simeq M$ as x is a non-zero-divisor. In particular

$$xM/(xM \cap \mathfrak{m}^t M) \simeq M/N_t$$

where

$$N_t = \left\{ a \in M : xa \in \mathfrak{m}^t M \right\}.$$

In particular, $N_t \supset \mathfrak{m}^{t-1}M$. This tells us that $\ell(M/N_t) \leqslant \ell(M/\mathfrak{m}^{t-1}M) = f_M^+(t-1)$ for $t \gg 0$. Combining this with the above information, we learn that

$$f_M^+(t) \leqslant f_{M/xM}^+(t) + f_M^+(t-1),$$

which implies that $f_{M/xM}^+(t)$ is at least the successive difference $f_M^+(t) - f_M^+(t-1)$. This last polynomial has degree $\dim \operatorname{supp} M - 1$. In particular, $f_{M/xM}^+(t)$ has degree at least $\dim \operatorname{supp} M - 1$. This gives us one direction, actually the hard one. We showed that intersecting something with codimension one doesn't drive the dimension down too much.

Let us now do the other direction. We essentially did this last time via the Artin-Rees lemma. We know that $N_t = \{a \in M : xa \in \mathfrak{m}^t\}$. The Artin-Rees lemma tells us that there is a constant c such that $N_{t+c} \subset \mathfrak{m}^t M$ for all t. Therefore, $\ell(M/N_{t+c}) \geqslant \ell(M/\mathfrak{m}^t M) = f_M^+(t), t \gg 0$. Now remember the exact sequence $0 \to M/N_t \to M/\mathfrak{m}^t M \to M/(xM + \mathfrak{m}^t M) \to 0$. We see from this that

$$\ell(M/\mathfrak{m}^t M) = \ell(M/N_t) + f_{M/xM}^+(t) \geqslant f_M^+(t-c) + f_{M/xM}^+(t), \quad t \gg 0,$$

which implies that

$$f_{M/xM}^+(t) \leqslant f_M^+(t) - f_M^+(t-c),$$

so the degree must go down. And we find that $\deg f_{M/xM}^+ < \deg f_M^+$. ☐

This gives us an algorithm of computing the dimension of an R-module M. First, it reduces to computing $\dim R/\mathfrak{p}$ for $\mathfrak{p} \subset R$ a prime ideal. We may assume that R is a domain and that we are looking for $\dim R$. Geometrically, this corresponds to taking an irreducible component of $\operatorname{Spec} R$.

Now choose any $x \in R$ such that x is nonzero but noninvertible. If there is no such element, then R is a field and has dimension zero. Then compute $\dim R/x$ (recursively) and add one.

Notice that this algorithm said nothing about Hilbert polynomials, and only talked about the structure of prime ideals.

Recap

Last time, we were talking about dimension theory. Recall that R is a local noetherian ring with maximal ideal $\mathfrak{m}$, M a finitely generated R-module. We can look at the lengths $\ell(M/\mathfrak{m}^t M)$ for varying t; for $t \gg 0$ this is a polynomial function. The degree of this polynomial is called the **dimension** of $\operatorname{supp} M$.

7.2.3 Remark If $M = 0$, then we define $\dim \operatorname{supp} M = -1$ by convention.

Last time, we showed that if $M \neq 0$ and $x \in \mathfrak{m}$ such that x is a non-zero-divisor on M (i.e. $M \xrightarrow{x} M$ injective), then

$$\boxed{\dim \operatorname{supp} M/xM = \dim \operatorname{supp} M - 1.}$$

Using this, we could give a recursion for calculating the dimension. To compute $\dim R = \dim \operatorname{Spec} R$, we note three properties:

1. $\dim R = \sup_{\mathfrak{p} \text{ a minimal prime}} R/\mathfrak{p}$. Intuitively, this says that a variety which is the union of irreducible components has dimension equal to the maximum of these irreducibles.
2. $\dim R = 0$ for R a field. This is obvious from the definitions.
3. If R is a domain, and $x \in \mathfrak{m} - \{0\}$, then $\dim R/(x) + 1 = \dim R$. This is obvious from the boxed formula as x is a non-zero-divisor.

These three properties *uniquely characterize* the dimension invariant.

More precisely, if $d : \{\text{local noetherian rings}\} \to \mathbb{Z}_{\geqslant 0}$ **satisfies the above three properties, then** $d = \dim$**.**

Proof. Induction on $\dim R$. It is clearly sufficient to prove this for R a domain. If R is a field, then it's clear; if $\dim R > 0$, the third condition lets us reduce to a case covered by the inductive hypothesis (i.e. go down). ☐

Let us rephrase 3 above:

3': If R is a domain and not a field, then

$$\dim R = \sup_{x \in \mathfrak{m} - 0} \dim R/(x) + 1.$$

Obviously 3' implies 3, and it is clear by the same argument that 1,2, 3' characterize the notion of dimension.

Krull dimension

We shall now define another notion of dimension, and show that it is equivalent to the older one by showing that it satisfies these axioms.

7.2.4 Definition Let R be a commutative ring. A **chain of prime ideals** in R is a finite sequence

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n.$$

This chain is said to have **length** n**.**

7.2.5 Definition The **Krull dimension** of R is equal to the maximum length of any chain of prime ideals. This might be ∞, but we will soon see this cannot happen for R local and noetherian.

7.2.6 Remark For any maximal chain $\{\mathfrak{p}_i, 0 \leqslant i \leqslant n\}$ of primes (i.e. which can't be expanded), we must have that $\mathfrak{p}_0$ is minimal prime and $\mathfrak{p}_n$ a maximal ideal.

7.2.7 Theorem *For a noetherian local ring R, the Krull dimension of R exists and is equal to the usual* $\dim R$*.*

Proof. We will show that the Krull dimension satisfies the above axioms. For now, write Krdim for Krull dimension.

1. First, note that $\operatorname{Krdim}(R) = \max_{\mathfrak{p} \in R \text{ minimal}} \operatorname{Krdim}(R/\mathfrak{p})$. This is because any chain of prime ideals in R contains a minimal prime. So any chain of prime ideals in R can be viewed as a chain in *some* $R/\mathfrak{p}$, and conversely.

2. Second, we need to check that $\operatorname{Krdim}(R) = 0$ for R a field. This is obvious, as there is precisely one prime ideal.

3. The third condition is interesting. We must check that for $(R, \mathfrak{m})$ a local domain,

$$\operatorname{Krdim}(R) = \max_{x \in \mathfrak{m} - \{0\}} \operatorname{Krdim}(R/(x)) + 1.$$

If we prove this, we will have shown that condition 3' is satisfied by the Krull dimension. It will follow by the inductive argument above that $\operatorname{Krdim}(R) = \dim(R)$ for any R. There are two inequalities to prove. First, we must show

$$\operatorname{Krdim}(R) \geqslant \operatorname{Krdim}(R/x) + 1, \quad \forall x \in \mathfrak{m} - 0.$$

So suppose $k = \operatorname{Krdim}(R/x)$. We want to show that there is a chain of prime ideals of length $k+1$ in R. So say $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_k$ is a chain of length k in $R/(x)$. The inverse images in R give a proper chain of primes in R of length k, all of which contain (x) and thus properly contain 0. Thus adding zero will give a chain of primes in R of length $k+1$.

Conversely, we want to show that if there is a chain of primes in R of length $k+1$, then there is a chain of length k in $R/(x)$ for some $x \in \mathfrak{m} - \{0\}$. Let us write the chain of length $k+1$:

$$\mathfrak{q}_{-1} \subset \mathfrak{q}_0 \subsetneq \cdots \subsetneq \mathfrak{q}_k \subset R.$$

Now evidently $\mathfrak{q}_0$ contains some $x \in \mathfrak{m} - 0$. Then the chain $\mathfrak{q}_0 \subsetneq \cdots \subsetneq \mathfrak{q}_k$ can be identified with a chain in $R/(x)$ for this x. So for this x, we have that $\operatorname{Krdim} R \leqslant \sup \operatorname{Krdim} R/(x) + 1$. $\square$

There is thus a combinatorial definition of definition.

Geometrically, let $X = \operatorname{Spec} R$ for R an affine ring over $\mathbb{C}$ (a polynomial ring mod some ideal). Then R has Krull dimension $\geqslant k$ iff there is a chain of irreducible subvarieties of X,

$$X_0 \supset X_1 \supset \cdots \supset X_k.$$

You will meet justification for this in Section 7.3 below.

7.2.8 Remark (Warning!) Let R be a local noetherian ring of dimension k. This means that there is a chain of prime ideals of length k, and no longer chains. Thus there is a maximal chain whose length is k. However, not all maximal chains in $\operatorname{Spec} R$ have length k.

7.2.9 Example Let $R = (\mathbb{C}[X, Y, Z]/(XY, XZ))_{(X,Y,Z)}$. It is left as an exercise to the reader to see that there are maximal chains of length not two.

There are more complicated local noetherian *domains* which have maximal chains of prime ideals not of the same length. These examples are not what you would encounter in daily experience, and are necessarily complicated. This cannot happen for finitely generated domains over a field.

7.2.10 Example An easier way all maximal chains could fail to be of the same length is if $\operatorname{Spec} R$ has two components (in which case $R = R_0 \times R_1$ for rings R_0, R_1).

Yet another definition

Let's start by thinking about the definition of a module. Recall that if $(R, \mathfrak{m})$ is a local noetherian ring and M a finitely generated R-module, and $x \in \mathfrak{m}$ is a non-zero-divisor on M, then

$$\dim \operatorname{supp} M/xM = \dim \operatorname{supp} M - 1.$$

7.2.11 Remark (Question) What if x is a zero divisor?

This is not necessarily true (e.g. if $x \in \operatorname{Ann}(M)$). Nonetheless, we claim that even in this case:

7.2.12 Proposition *For any $x \in \mathfrak{m}$,*

$$\boxed{\dim \operatorname{supp} M \geqslant \dim \operatorname{supp} M/xM \geqslant \dim \operatorname{supp} M - 1.}$$

The upper bound on $\dim M/xM$ is obvious as M/xM is a quotient of M. The lower bound is trickier.

Proof. Let $N = \{a \in M : x^n a = 0 \text{ for some } n\}$. We can construct an exact sequence

$$0 \to N \to M \to M/N \to 0.$$

Let $M'' = M/N$. Now x is a non-zero-divisor on M/N by construction. We claim that

$$0 \to N/xN \to M/xM \to M''/xM'' \to 0$$

is exact as well. For this we only need to see exactness at the beginning, i.e. injectivity of $N/xN \to M/xM$. So we need to show that if $a \in N$ and $a \in xM$, then $a \in xN$.

To see this, suppose $a = xb$ where $b \in M$. Then if $\phi : M \to M''$, then $\phi(b) \in M''$ is killed by x as $x\phi(b) = \phi(bx) = \phi(a)$. This means that $\phi(b) = 0$ as $M'' \xrightarrow{x} M''$ is injective. Thus $b \in N$ in fact. So $a \in xN$ in fact.

From the exactness, we see that (as x is a non-zero-divisor on M'')

$$\begin{aligned}\dim M/xM &= \max(\dim M''/xM'', \dim N/xN) \geqslant \max(\dim M'' - 1, \dim N) \\ &\geqslant \max(\dim M'', \dim N) - 1.\end{aligned}$$

The reason for the last claim is that $\operatorname{supp} N/xN = \operatorname{supp} N$ as N is x-torsion, and the dimension depends only on the support. But the thing on the right is just $\dim M - 1$. ∎

As a result, we find:

7.2.13 Proposition $\dim \operatorname{supp} M$ *is the minimal integer n such that there exist elements $x_1, \ldots, x_n \in \mathfrak{m}$ with $M/(x_1, \ldots, x_n)M$ has finite length.*

Note that n always exists, since we can look at a bunch of generators of the maximal ideal, and $M/\mathfrak{m}M$ is a finite-dimensional vector space and is thus of finite length.

Proof. Induction on $\dim \operatorname{supp} M$. Note that $\dim \operatorname{supp}(M) = 0$ if and only if the Hilbert polynomial has degree zero, i.e. M has finite length or that $n = 0$ (n being defined as in the statement).

Suppose $\dim \operatorname{supp} M > 0$.

1. We first show that there are $x_1, \dots, x_{\dim M}$ with $M/(x_1, \dots, x_{\dim M})M$ have finite length. Let $M' \subset M$ be the maximal submodule having finite length. There is an exact sequence

$$0 \to M' \to M \to M'' \to 0$$

where $M'' = M/M'$ has no finite length submodules. In this case, we can basically ignore M', and replace M by M''. The reason is that modding out by M' doesn't affect either n or the dimension.

So let us replace M with M'' and thereby assume that M has no finite length submodules. In particular, M does not contain a copy of $R/\mathfrak{m}$, i.e. $\mathfrak{m} \notin \operatorname{Ass}(M)$. By prime avoidance, this means that there is $x_1 \in \mathfrak{m}$ that acts as a non-zero-divisor on M. Thus

$$\dim M/x_1 M = \dim M - 1.$$

The inductive hypothesis says that there are $x_2, \dots, x_{\dim M}$ with

$$(M/x_1M)/(x_2, \dots, x_{\dim M})(M/xM) \simeq M/(x_1, \dots, x_{\dim M})M$$

of finite length. This shows the claim.

2. Conversely, suppose that there $M/(x_1, \dots, x_n)M$ has finite length. Then we claim that $n \geqslant \dim M$. This follows because we had the previous result that modding out by a single element can chop off the dimension by at most 1. Recursively applying this, and using the fact that dim of a finite length module is zero, we find

$$0 = \dim M/(x_1, \dots, x_n)M \geqslant \dim M - n.$$

□

7.2.14 Corollary *Let $(R, \mathfrak{m})$ be a local noetherian ring. Then $\dim R$ is equal to the minimal n such that there exist $x_1, \dots, x_n \in R$ with $R/(x_1, \dots, x_n)R$ is artinian. Or, equivalently, such that $(x_1, \dots, x_n)$ contains a power of $\mathfrak{m}$.*

7.2.15 Remark We manifestly have here that the dimension of R is at most the embedding dimension. Here, we're not worried about generating the maximal ideal, but simply something containing a power of it.

We have been talking about dimension. Let R be a local noetherian ring with maximal ideal $\mathfrak{m}$. Then, as we have said in previous lectures, $\dim R$ can be characterized by:

1. The minimal n such that there is an n-primary ideal generated by n elements $x_1, \dots, x_n \in \mathfrak{m}$. That is, the closed point $\mathfrak{m}$ of $\operatorname{Spec} R$ is cut out *set-theoretically* by the intersection $\bigcap V(x_i)$. This is one way of saying that the closed point can be defined by n parameters.

2. The *maximal* n such that there exists a chain of prime ideals

$$\mathfrak{p}_0 \subset \mathfrak{p}_1 \subset \cdots \subset \mathfrak{p}_n.$$

3. The degree of the Hilbert polynomial $f^+(t)$, which equals $\ell(R/\mathfrak{m}^t)$ for $t \gg 0$.

Krull's Hauptidealsatz

Let R be a local noetherian ring. The following is now clear from what we have shown:

7.2.16 Theorem *R has dimension 1 if and only if there is a non-zero-divisor $x \in \mathfrak{m}$ such that $R/(x)$ is artinian.*

7.2.17 Remark Let R be a domain. We said that a nonzero prime $\mathfrak{p} \subset R$ is **height one** if $\mathfrak{p}$ is minimal among the prime ideals containing some nonzero $x \in R$.

According to Krull's Hauptidealsatz, $\mathfrak{p}$ has height one **if and only if** $\dim R_{\mathfrak{p}} = 1$**.**

We can generalize the notion of $\mathfrak{p}$ as follows.

7.2.18 Definition Let R be a noetherian ring (not necessarily local), and $\mathfrak{p} \in \operatorname{Spec} R$. Then we define the **height** of $\mathfrak{p}$, denoted height($\mathfrak{p}$), as $\dim R_{\mathfrak{p}}$. We know that this is the length of a maximal chain of primes in $R_{\mathfrak{p}}$. This is thus the maximal length of prime ideals of R,

$$\mathfrak{p}_0 \subset \cdots \subset \mathfrak{p}_n = \mathfrak{p}$$

that ends in $\mathfrak{p}$. This is the origin of the term "height."

7.2.19 Remark Sometimes, the height is called the **codimension**. This corresponds to the codimension in $\operatorname{Spec} R$ of the corresponding irreducible closed subset of $\operatorname{Spec} R$.

7.2.20 Theorem (Krull's Hauptidealsatz) *Let R be a noetherian ring, and $x \in R$ a non-zero-divisor. If $\mathfrak{p} \in \operatorname{Spec} R$ is minimal over x, then $\mathfrak{p}$ has height one.*

Proof. Immediate from theorem 7.2.16. □

7.2.21 Theorem (Artin-Tate) *Let A be a noetherian domain. Then the following are equivalent:*

1. *There is $f \in A - \{0\}$ such that A_f is a field.*
2. *A has finitely many maximal ideals and has dimension at most 1.*

Proof. We follow **?**.

Suppose first that there is f with A_f a field. Then all nonzero prime ideals of A contain f. We need to deduce that A has dimension $\leqslant 1$. Without loss of generality, we may assume that A is not a field.

There are finitely many primes $\mathfrak{p}_1, \dots, \mathfrak{p}_k$ which are minimal over f; these are all height one. The claim is that any maximal ideal of A is of this form. Suppose $\mathfrak{m}$ were maximal and not one of the $\mathfrak{p}_i$. Then by prime avoidance, there is $g \in \mathfrak{m}$ which lies in no $\mathfrak{p}_i$. A minimal prime $\mathfrak{P}$ of g has height one, so by our assumptions contains f. However, it is then one of the $\mathfrak{p}_i$; this is a contradiction as $g \in \mathfrak{P}$. □

Further remarks

We can recast earlier notions in terms of dimension.

7.2.22 Remark A noetherian ring has dimension zero if and only if R is artinian. Indeed, R has dimension zero iff all primes are maximal.

7.2.23 Remark A noetherian domain has dimension zero iff it is a field. Indeed, in this case (0) is maximal.

7.2.24 Remark R has dimension $\leqslant 1$ if and only if every non-minimal prime of R is maximal. That is, there are no chains of length $\geqslant 2$.

7.2.25 Remark A (noetherian) domain R has dimension $\leqslant 1$ iff every nonzero prime ideal is maximal.

In particular,

7.2.26 Proposition *R is Dedekind iff it is a noetherian, integrally closed domain of dimension 1.*

7.3. Further topics

Change of rings

Let $f : R \to R'$ be a map of noetherian rings.

7.3.1 Remark (Question) What is the relationship between $\dim R$ and $\dim R'$?

A map f gives a map $\operatorname{Spec} R' \to \operatorname{Spec} R$, where $\operatorname{Spec} R'$ is the union of various fibers over the points of $\operatorname{Spec} R$. You might imagine that the dimension is the dimension of R plus the fiber dimension. This is sometimes true.

Now assume that R, R' are *local* with maximal ideals $\mathfrak{m}, \mathfrak{m}'$. Assume furthermore that f is local, i.e. $f(\mathfrak{m}) \subset \mathfrak{m}'$.

7.3.2 Theorem $\dim R' \leqslant \dim R + \dim R'/\mathfrak{m}R'$. *Equality holds if $f : R \to R'$ is flat.*

Here $R'/\mathfrak{m}R'$ is to be interpreted as the "fiber" of $\operatorname{Spec} R'$ above $\mathfrak{m} \in \operatorname{Spec} R$. The fibers can behave weirdly as the basepoint varies in $\operatorname{Spec} R$, so we can't expect equality in general.

7.3.3 Remark Let us review flatness as it has been a while. An R-module M is *flat* iff the operation of tensoring with M is an exact functor. The map $f : R \to R'$ is *flat* iff R' is a flat R-module. Since the construction of taking fibers is a tensor product (i.e. $R'/\mathfrak{m}R' = R' \otimes_R R/\mathfrak{m}$), perhaps the condition of flatness here is not as surprising as it might be.

Proof. Let us first prove the inequality. Say

$$\dim R = a, \ \dim R'/\mathfrak{m}R' = b.$$

We'd like to see that

$$\dim R' \leqslant a + b.$$

To do this, we need to find $a+b$ elements in the maximal ideal $\mathfrak{m}'$ that generate a $\mathfrak{m}'$-primary ideal of R'.

There are elements $x_1, \dots, x_a \in \mathfrak{m}$ that generate an $\mathfrak{m}$-primary ideal $I = (x_1, \dots, x_a)$ in R. There is a surjection $R'/IR' \twoheadrightarrow R'/\mathfrak{m}R'$. The kernel $\mathfrak{m}R'/IR'$ is nilpotent since I contains a power of $\mathfrak{m}$. We've seen that nilpotents *don't* affect the dimension. In particular,

$$\dim R'/IR' = \dim R'/\mathfrak{m}R' = b.$$

There are thus elements $y_1, \dots, y_b \in \mathfrak{m}'/IR'$ such that the ideal $J = (y_1, \dots, y_b) \subset R'/IR'$ is $\mathfrak{m}'/IR'$-primary. The inverse image of J in R', call it $\overline{J} \subset R'$, is $\mathfrak{m}'$-primary. However, $\overline{J}$ is generated by the $a+b$ elements

$$f(x_1), \dots, f(x_a), \overline{y_1}, \dots, \overline{y_b}$$

if the $\overline{y_i}$ lift y_i.

But we don't always have equality. Nonetheless, if all the fibers are similar, then we should expect that the dimension of the "total space" $\operatorname{Spec} R'$ is the dimension of the "base" $\operatorname{Spec} R$ plus the "fiber" dimension $\operatorname{Spec} R'/\mathfrak{m}R'$. *The precise condition of f flat articulates the condition that the fibers "behave well."* Why this is so is something of a mystery, for now. But for some evidence, take the present result about fiber dimension.

Anyway, let us now prove equality for flat R-algebras. As before, write $a = \dim R, b = \dim R'/\mathfrak{m}R'$. We'd like to show that

$$\dim R' \geqslant a + b.$$

By what has been shown, this will be enough. This is going to be tricky since we now need to give *lower bounds* on the dimension; finding a sequence $x_1, \dots, x_{a+b}$ such that the quotient $R/(x_1, \dots, x_{a+b})$ is artinian would bound *above* the dimension.

So our strategy will be to find a chain of primes of length $a+b$. Well, first we know that there are primes

$$\mathfrak{q}_0 \subset \mathfrak{q}_1 \subset \cdots \subset \mathfrak{q}_b \subset R'/\mathfrak{m}R'.$$

Let $\overline{\mathfrak{q}_i}$ be the inverse images in R'. Then the $\overline{\mathfrak{q}_i}$ are a strictly ascending chain of primes in R' where $\overline{\mathfrak{q}_0}$ contains $\mathfrak{m}R'$. So we have a chain of length b; we need to extend this by additional terms.

Now $f^{-1}(\overline{\mathfrak{q}_0})$ contains $\mathfrak{m}$, hence is $\mathfrak{m}$. Since $\dim R = a$, there is a chain $\{\mathfrak{p}_i\}$ of prime ideals of length a going down from $f^{-1}(\overline{\mathfrak{q}_0}) = \mathfrak{m}$. We are now going to find primes $\mathfrak{p}_i' \subset R'$ forming a chain such that $f^{-1}(\mathfrak{p}_i') = \mathfrak{p}_i$. In other words, we are going to *lift* the chain $\mathfrak{p}_i$ to $\operatorname{Spec} R'$. We can do this at the first stage for $i = a$, where $\mathfrak{p}_a = \mathfrak{m}$ and we can set $\mathfrak{p}_a' = \overline{\mathfrak{q}_0}$. If we can indeed do this lifting, and catenate the chains $\overline{\mathfrak{q}_j}, \mathfrak{p}_i'$, then we will have a chain of the appropriate length.

We will proceed by descending induction. Assume that we have $\mathfrak{p}'_{i+1} \subset R'$ and $f^{-1}(\mathfrak{p}'_{i+1}) = \mathfrak{p}_{i+1} \subset R$. We want to find $\mathfrak{p}'_i \subset \mathfrak{p}'_{i+1}$ such that $f^{-1}(\mathfrak{p}'_i) = \mathfrak{p}_i$. The existence of that prime is a consequence of the following general fact.

7.3.4 Theorem (Going down) *Let $f : R \to R'$ be a flat map of noetherian commutative rings. Suppose $\mathfrak{q} \in \operatorname{Spec} R'$, and let $\mathfrak{p} = f^{-1}(\mathfrak{q})$. Suppose $\mathfrak{p}_0 \subset \mathfrak{p}$ is a prime of R. Then there is a prime $\mathfrak{q}_0 \subset \mathfrak{q}$ with*

$$f^{-1}(\mathfrak{q}_0) = \mathfrak{p}_0.$$

Proof. We may replace R' with $R'_{\mathfrak{q}}$. There is still a map

$$R \to R'_{\mathfrak{q}}$$

which is flat as localization is flat. The maximal ideal in $R'_{\mathfrak{q}}$ has inverse image $\mathfrak{p}$. So the problem now reduces to finding *some* $\mathfrak{p}_0$ in the localization that pulls back appropriately.

Anyhow, throwing out the old R and replacing with the localization, we may assume that R' is local and $\mathfrak{q}$ the maximal ideal. (The condition $\mathfrak{q}_0 \subset \mathfrak{q}$ is now automatic.)

The claim now is that we can replace R with $R/\mathfrak{p}_0$ and R' with $R'/\mathfrak{p}_0 R' = R' \otimes R/\mathfrak{p}_0$. We can do this because base change preserves flatness (see below), and in this case we can reduce to the case of $\mathfrak{p}_0 = (0)$—in particular, R is a domain. Taking these quotients just replaces $\operatorname{Spec} R, \operatorname{Spec} R'$ with closed subsets where all the action happens anyhow.

Under these replacements, we now have:

1. R' is local with maximal ideal $\mathfrak{q}$
2. R is a domain and $\mathfrak{p}_0 = (0)$.

We want a prime of R' that pulls back to (0) in R. I claim that any minimal prime of R' will work. Suppose otherwise. Let $\mathfrak{q}_0 \subset R'$ be a minimal prime, and suppose $x \in R \cap f^{-1}(\mathfrak{q}_0) - \{0\}$. But $\mathfrak{q}_0 \in \operatorname{Ass}(R')$. So $f(x)$ is a zero divisor on R'. Thus multiplication by x on R' is not injective.

But, R is a domain, so $R \xrightarrow{x} R$ is injective. Tensoring with R' must preserve this, implying that $R' \xrightarrow{x} R'$ is injective because R' is flat. This is a contradiction. $\square$

We used:

7.3.5 Lemma *Let $R \to R'$ be a flat map, and S an R-algebra. Then $S \to S \otimes_R R'$ is a flat map.*

Proof. The construction of taking an S-module with $S \otimes_R R'$ is an exact functor, because that's the same thing as taking an S-module, restricting to R, and tensoring with R'. $\square$

The proof of the fiber dimension theorem is now complete.

The dimension of a polynomial ring

Adding an indeterminate variable corresponds geometrically to taking the product with the affine line, and so should increase the dimension by one. We show that this is indeed the case.

7.3.6 Theorem *Let R be a noetherian ring. Then* $\dim R[X] = \dim R + 1$.

Interestingly, this is *false* if R is non-noetherian, cf. . Let R be a ring of dimension n.

7.3.7 Lemma $\dim R[x] \geqslant \dim R + 1$.

Proof. Let $\mathfrak{p}_0 \subset \cdots \subset \mathfrak{p}_n$ be a chain of primes of length $n = \dim R$. Then $\mathfrak{p}_0 R[x] \subset \cdots \subset \mathfrak{p}_n R[x] \subset (x, \mathfrak{p}_n)R[x]$ is a chain of primes in $R[x]$ of length $n+1$ because of the following fact: if $\mathfrak{q} \subset R$ is prime, then so is $\mathfrak{q}R[x] \subset R[x]$.[2] Note also that as $\mathfrak{p}_n \subsetneq R$, we have that $\mathfrak{p}_n R[x] \subsetneq (x, \mathfrak{p}_n)$. So this is indeed a legitimate chain. $\square$

Now we need only show:

7.3.8 Lemma *Let R be noetherian of dimension n. Then* $\dim R[x] \leqslant \dim R + 1$.

Proof. Let $\mathfrak{q}_0 \subset \cdots \subset \mathfrak{q}_m \subset R[x]$ be a chain of primes in $R[x]$. Let $\mathfrak{m} = \mathfrak{q}_m \cap R$. Then if we localize and replace R with $R_{\mathfrak{m}}$, we get a chain of primes of length m in $R_{\mathfrak{m}}[x]$. In fact, we get more. We get a chain of primes of length m in $(R[x])_{\mathfrak{q}_m}$, and a *local* inclusion of noetherian local rings

$$R_{\mathfrak{m}} \hookrightarrow (R[x])_{\mathfrak{q}_m}.$$

To this we can apply the fiber dimension theorem. In particular, this implies that

$$m \leqslant \dim(R[x])_{\mathfrak{q}_m} \leqslant \dim R_{\mathfrak{m}} + \dim(R[x])_{\mathfrak{q}_m}/\mathfrak{m}(R[x])_{\mathfrak{q}_m}.$$

Here $\dim R_{\mathfrak{m}} \leqslant \dim R = n$. So if we show that $\dim(R[x])_{\mathfrak{q}_m}/\mathfrak{m}(R[x])_{\mathfrak{q}_m} \leqslant 1$, we will have seen that $m \leqslant n+1$, and will be done. But this last ring is a localization of $(R_{\mathfrak{m}}/\mathfrak{m}R_{\mathfrak{m}})[x]$, which is a PID by the euclidean algorithm for polynomial rings over a field, and thus of dimension $\leqslant 1$. $\square$

A refined fiber dimension theorem

Let R be a local noetherian domain, and let $R \to S$ be an injection of rings making S into an R-algebra. Suppose S is also a local domain, such that the morphism $R \to S$ is local. This is essentially the setup of section 7.3, but in this section, we make the refining assumption that S is *essentially of finite type* over R; in other words, S is the localization of a finitely generated R-algebra.

Let k be the residue field of R, and k' that of S; because $R \to S$ is local, there is induced a morphism of fields $k \to k'$. We shall prove, following **?**:

7.3.9 Theorem (Dimension formula)

$$\dim S + \text{tr.deg.} S/R \leqslant \dim R + \text{tr.deg.} k'/k. \tag{7.3.1}$$

[2] This is because $R[x]/\mathfrak{q}R[x] = (R/\mathfrak{q})[x]$ is a domain.

Here tr.deg.B/A is more properly the transcendence degree of the quotient field of B over that of A. Geometrically, it corresponds to the dimension of the "generic" fiber.

Proof. Let $\mathfrak{m} \subset R$ be the maximal ideal. We know that S is a localization of an algebra of the form $(R[x_1, \ldots, x_k])/\mathfrak{p}$ where $\mathfrak{p} \subset R[x_1, \ldots, x_n]$ is a prime ideal $\mathfrak{q}$. We induct on k.

Since we can "dévissage" the extension $R \to S$ as the composite

$$R \to (R[x_1, \ldots, x_{k-1}]/(\mathfrak{p} \cap R[x_1, \ldots, x_{k-1}])_{\mathfrak{q}'} \to S,$$

(where $\mathfrak{q}' \in \operatorname{Spec} R[x_1, \ldots, x_{k-1}]/(\mathfrak{p} \cap R[x_1, \ldots, x_{k-1}]$ is the pull-back of $\mathfrak{q}$), we see that it suffices to prove (7.3.1) when $k = 1$, that is S is the localization of a quotient of $R[x]$.

So suppose $k = 1$. Then we have $S = (R[x])_{\mathfrak{q}}/\mathfrak{p}$ where $\mathfrak{q} \subset R[x]$ is another prime ideal lying over $\mathfrak{m}$. Let us start by considering the case where $\mathfrak{q} = 0$.

7.3.10 Lemma *Let $(R, \mathfrak{m})$ be a local noetherian domain as above. Let $S = R[x]_{\mathfrak{q}}$ where $\mathfrak{q} \in \operatorname{Spec} R[x]$ is a prime lying over $\mathfrak{m}$. Then (7.3.1) holds with equality.*

Proof. In this case, tr.deg.$S/R = 1$. Now $\mathfrak{q}$ could be $\mathfrak{m}R[x]$ or a prime ideal containing that, which is then automatically maximal, as we know from the proof of section 7.3. Indeed, primes containing $\mathfrak{m}R[x]$ are in bijection with primes of $R/\mathfrak{m}[x]$, and these come in two forms: zero, and those generated by one element. (Note that in the former case, the residue field is the field of rational functions $k(x)$ and in the second, the residue field is finite over k.)

1. In the first case, $\dim S = \dim R[x]_{\mathfrak{m}R[x]} = \dim R$ and but the residue field extension is $(R[x]_{\mathfrak{m}R[x]})/\mathfrak{m}R[x]_{\mathfrak{m}R[x]} = k(x)$, so tr.deg.$k'/k = 1$ and the formula is satisfied.
2. In the second case, $\mathfrak{q}$ properly contains $\mathfrak{m}R[x]$. Then $\dim R[x]_{\mathfrak{q}} = \dim R + 1$, but the residue field extension is finite. So in this case too, the formula is satisfied. $\square$

Now, finally, we have to consider the case where $\mathfrak{p} \subset R[x]$ is not zero, and we have $S = (R[x]/\mathfrak{p})_{\mathfrak{q}}$ for $\mathfrak{q} \in \operatorname{Spec} R[x]/\mathfrak{p}$ lying over $\mathfrak{m}$. In this case, tr.deg.$S/R = 0$. So we need to prove

$$\dim S \leqslant \dim R + \text{tr.deg.}k'/k.$$

Let us, by abuse of notation, identify $\mathfrak{q}$ with its preimage in $R[x]$. (Recall that $\operatorname{Spec} R[x]/\mathfrak{p}$ is canonically identified as a closed subset of $\operatorname{Spec} R[x]$.) Then we know that $\dim(R[x]/\mathfrak{p})_{\mathfrak{q}}$ is the largest chain of primes in $R[x]$ between $\mathfrak{p}, \mathfrak{q}$. In particular, it is at most $\dim R[x]_{\mathfrak{q}} - \text{height}\mathfrak{p} \leqslant \dim R + 1 - 1 = \dim R$. So the result is clear. $\square$

In **?**, this is used to prove the geometric result that if $\phi : X \to Y$ is a morphism of varieties over an algebraically closed field (or a morphism of finite type between nice schemes), then the local dimension (that is, the dimension at x) of the fiber $\phi^{-1}(\phi(x))$ is an upper semi-continuous function of $x \in X$.

An infinite-dimensional noetherian ring

We shall now present an example, due to Nagata, of an infinite-dimensional noetherian ring. Note that such a ring cannot be *local*.

Consider the ring $R = \mathbb{C}[\{x_{i,j}\}_{0 \leqslant i \leqslant j}]$ of polynomials in infinitely many variables $x_{i,j}$. This is clearly an infinite-dimensional ring, but it is also not noetherian. We will localize it suitably to make it noetherian.

Let $\mathfrak{p}_n \subset R$ be the ideal $(x_{1,n}, x_{2,n}, \dots, x_{n,n})$ for all $i \leqslant n$. Let $S = R - \bigcup \mathfrak{p}_n$; this is a multiplicatively closed set.

7.3.11 Theorem (Nagata) *The ring $S^{-1}R$ is noetherian and has infinite dimension.*

We start with

7.3.12 Proposition *The ring in the statement of the problem is noetherian.*

The proof is slightly messy, so we first prove a few lemmas.

Let $R' = S^{-1}R$ as in the problem statement. We start by proving that every ideal in R' is contained in one of the $\mathfrak{p}_n$ (which, by abuse of notation, we identify with their localizations in $R' = S^{-1}R$). In particular, the $\mathfrak{p}_n$ are the maximal ideals in R'.

7.3.13 Lemma *The $\mathfrak{p}_n$ are the maximal ideals in R'.*

Proof. We start with an observation:

> If $f \neq 0$, then f belongs to only finitely many $\mathfrak{p}_n$.

To see this, let us use the following notation. If M is a monomial, we let $S(M)$ denote the set of subscripts $x_{a,b}$ that occur and $S_2(M)$ the set of second subscripts (i.e. the b's). For $f \in R$, we define $S(f)$ to be the intersection of the $S(M)$ for M a monomial occurring nontrivially in f. Similarly we define $S_2(f)$.

Let us prove:

7.3.14 Lemma *$f \in \mathfrak{p}_n$ iff $n \in S_2(f)$. Moreover, $S(f)$ and $S_2(f)$ are finite for any $f \neq 0$.*

Proof. Indeed, $f \in \mathfrak{p}_n$ iff every monomial in f is divisible by some $x_{i,n}, i \leqslant n$, as $\mathfrak{p}_n = (x_{i,n})_{i \leqslant n}$. From this the first assertion is clear. The second too, because f will contain a nonzero monomial, and that can be divisible by only finitely many $x_{a,b}$. ∎

From this, it is clear how to define $S_2(f)$ for any element in R', not necessarily a polynomial in R. Namely, it is the set of n such that $f \in \mathfrak{p}_n$. It is now clear, from the second statement of the lemma, that any $f \neq 0$ lies in *only finitely many* $\mathfrak{p}_n$. In particular, the observation has been proved.

Let $\mathcal{T} = \{S_2(f), f \in I - 0\}$. *I claim that $\varnothing \in \mathcal{T}$ iff $I = (1)$.* For $\varnothing \in \mathcal{T}$ iff there is a polynomial lying in no $\mathfrak{p}_n$. Since the union $\bigcup \mathfrak{p}_n$ is the set of non-units (by construction), we find that the assertion is clear.

7.3.15 Lemma *$\mathcal{T}$ is closed under finite intersections.*

Proof. Suppose $T_1, T_2 \in \mathcal{T}$. Without loss of generality, there are *polynomials* $F_1, F_2 \in R$ such that $S_2(F_1) = T_1, S_2(F_2) = T_2$. A generic linear combination $aF_1 + bF_2$ will involve no cancellation for $a, b \in \mathbb{C}$, and the monomials in this linear combination will be the union of those in F_1 and those in F_2 (scaled appropriately). So $S_2(aF_1 + bF_2) = S_2(F_1) \cap S_2(F_2)$. ∎

Finally, we can prove the result that the $\mathfrak{p}_n$ are the only maximal ideals. Suppose I was contained in no $\mathfrak{p}_n$, and form the set $\mathcal{T}$ as above. This is a collection of finite sets. Since $I \not\subset \mathfrak{p}_n$ for each n, we find that $n \notin \bigcap_{T \in \mathcal{T}} T$. This intersection is thus empty. It follows that there is a *finite* intersection of sets in $\mathcal{T}$ which is empty as $\mathcal{T}$ consists of finite sets. But $\mathcal{T}$ is closed under intersections. There is thus an element in I whose S_2 is empty, and is thus a unit. Thus $I = (1)$. ∎

We have proved that the $\mathfrak{p}_n$ are the only maximal ideals. This is not enough, though. We need:

7.3.16 Lemma *$R'_{\mathfrak{p}_n}$ is noetherian for each n.*

Proof. Indeed, any polynomial involving the variables $x_{a,b}$ for $b \neq n$ is invertible in this ring. We see that this ring contains the field

$$\mathbb{C}(\{x_{a,b}, b \neq n\}),$$

and over it is contained in the field $\mathbb{C}(\{x_{a,b}, \forall a, b\})$. It is a localization of the algebra $\mathbb{C}(\{x_{a,b}, b \neq n\})[x_{1,n}, \dots, x_{n,n}]$ and is consequently noetherian by Hilbert's basis theorem. ∎

The proof will be completed with:

7.3.17 Lemma *Let R be a ring. Suppose every element $x \neq 0$ in the ring belongs to only finitely many maximal ideals, and suppose that $R_{\mathfrak{m}}$ is noetherian for each $\mathfrak{m} \subset R$ maximal. Then R is noetherian.*

Proof. Let $I \subset R$ be a nonzero ideal. We must show that it is finitely generated. We know that I is contained in only finitely many maximal ideals $\mathfrak{m}_1, \dots, \mathfrak{m}_k$. At each of these maximal ideals, we know that $I_{\mathfrak{m}_i}$ is finitely generated. Clearing denominators, we can choose a finite set of generators in R. So we can collect them together and get a finite set $a_1, \dots, a_N \subset I$ which generate $I_{\mathfrak{m}_i} \subset R_{\mathfrak{m}_i}$ for each i. It is not necessarily true that $J = (a_1, \dots, a_N) = I$, though we do have $\subset$. However, $I_{\mathfrak{m}} = J_{\mathfrak{m}}$ except at finitely many maximal ideals $\mathfrak{n}_1, \dots, \mathfrak{n}_M$ because a nonzero element is a.e. a unit. However, these $\mathfrak{n}_j$ are not among the $\mathfrak{m}_i$. In particular, for each j, there is $b_j \in I - \mathfrak{n}_j$ as $I \not\subset \mathfrak{n}_j$. Then we find that the ideal

$$(a_1, \dots, a_N, b_1, \dots, b_M) \subset I$$ ∎

becomes equal to I in all the localizations. So it is I, and I is finitely generated

We need only see that the ring R' has infinite dimension. But for each n, there is a chain of primes $(x_{1,n}) \subset (x_{1,n}, x_{2,n}) \subset \cdots \subset (x_{1,n}, \dots, x_{n,n})$ of length $n-1$. The supremum of the lengths is thus infinite.

Catenary rings

7.3.18 Definition A ring R is *catenary* if given any two primes $\mathfrak{p} \subsetneq \mathfrak{p}'$, any two maximal prime chains from $\mathfrak{p}$ to $\mathfrak{p}'$ have the same length.

Nagata showed that there are noetherian domains which are not catenary. We shall see that *affine rings*, or rings finitely generated over a field, are always catenary.

7.3.19 Definition If $\mathfrak{p} \in \operatorname{Spec} R$, then $\dim \mathfrak{p} := \dim R/\mathfrak{p}$.

7.3.20 Lemma *Let S be a k-affine domain with $tr.d._k S = n$, and let $\mathfrak{p} \in \operatorname{Spec} S$ be height one. Then $tr.d._k(S/\mathfrak{p}) = n - 1$.*

Proof. Case 1: assume $S = k[x_1, \ldots, x_n]$ is a polynomial algebra. In this case, height 1 primes are principal, so $\mathfrak{p} = (f)$ for some f. Say f has positive degree with respect to x_1, so $f = g_r(x_2, \ldots, x_n)x_1^r + \cdots$. We have that $k[x_2, \ldots, x_n] \cap (f) = (0)$ (just look at degree with respect to x_1). It follows that $k[x_2, \ldots, x_n] \hookrightarrow S/(f)$, so $\bar{x}_2, \ldots, \bar{x}_n$ are algebraically independent in $S/\mathfrak{p}$. By $\bar{x}_1$ is algebraic over $Q(k[\bar{x}_2, \ldots, \bar{x}_n])$ as witnessed by f. This, $tr.d._k S/\mathfrak{p} = n - 1$.

Case 2: reduction to case 1. Let $R = k[x_1, \ldots, x_n]$ be a Noether normalization for S, and let $\mathfrak{p}_0 = \mathfrak{p} \cap R$. Observe that Going Down applies (because S is a domain and R is normal). It follows that $ht_R(\mathfrak{p}_0) = ht_S(\mathfrak{p}) = 1$. By case 1, we get that $tr.d.(R/\mathfrak{p}_0) = n - 1$. By $(*)$, we get that $tr.d.R/\mathfrak{p}_0 = tr.d.(S/\mathfrak{p})$. □

7.3.21 Theorem *Any k-affine algebra S is catenary (even if S is not a domain). In fact, any saturated prime chain from $\mathfrak{p}$ to $\mathfrak{p}'$ has length $\dim \mathfrak{p} - \dim \mathfrak{p}'$. If S is a domain, then all maximal ideals have the same height.*

Proof. Consider any chain $\mathfrak{p} \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_r = \mathfrak{p}'$. Then we get the chain

$$S/\mathfrak{p} \twoheadrightarrow S/\mathfrak{p}_1 \twoheadrightarrow \cdots \twoheadrightarrow S/\mathfrak{p}_r = S/\mathfrak{p}'$$

Here $\mathfrak{p}_i/\mathfrak{p}_{i-1}$ is height 1 in $S/\mathfrak{p}_{i-1}$, so each arrow decreases the transcendence degree by exactly 1. Therefore, $tr.d._k S/\mathfrak{p}' = tr.d._k S/\mathfrak{p} - r$.

$$r = tr.d._k S/\mathfrak{p} - tr.d._k S/\mathfrak{p}' = \dim S/\mathfrak{p} - \dim S/\mathfrak{p}' = \dim \mathfrak{p} - \dim \mathfrak{p}'.$$

To get the last statement, take $\mathfrak{p} = 0$ and $\mathfrak{p}' = \mathfrak{m}$. Then we get that $ht(\mathfrak{m}) = \dim S$. □

Note that the last statement fails in general.

7.3.22 Example Take $S = k \times k[x_1, \ldots, x_n]$. Then $ht(0 \times k[x_1, \ldots, x_n]) = 0$, but $ht\big(k \times (x_1, \ldots, x_n)\big) = n$.

But that example is not connected.

7.3.23 Example $S = k[x, y, z]/(xy, xz)$.

But this example is not a domain. In general, for any prime $\mathfrak{p}$ in any ring S, we have

$$ht(\mathfrak{p}) + \dim \mathfrak{p} \leqslant \dim S.$$

7.3.24 Theorem *Let S be an affine algebra, with minimal primes $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$. Then the following are equivalent.*

1. $\dim \mathfrak{p}_i$ *are all equal.*
2. $ht(\mathfrak{p}) + \dim \mathfrak{p} = \dim S$ *for all primes* $\mathfrak{p} \in \operatorname{Spec} S$*. In particular, if S is a domain, we get this condition.*

Proof. ($1 \Rightarrow 2$) $ht(\mathfrak{p})$ is the length of some saturated prime chain from $\mathfrak{p}$ to some minimal prime $\mathfrak{p}_i$. This length is $\dim \mathfrak{p}_i - \dim \mathfrak{p} = \dim S - \dim \mathfrak{p}$ (by condition 1). Thus, we get (2).

($2 \Rightarrow 1$) Apply (2) to the minimal prime $\mathfrak{p}_i$ to get $\dim \mathfrak{p}_i = \dim S$ for all i. □

We finish with a (non-affine) noetherian domain S with maximal ideals of different heights. We need the following fact.
Fact: If R is a ring with $a \in R$, then there is a canonical R-algebra isomorphism $R[x]/(ax-1) \cong R[a^{-1}]$, $x \leftrightarrow a^{-1}$.

7.3.25 Example Let $\big(R, (\mathfrak{p}i)\big)$ be a DVR with quotient field K. Let $S = R[x]$, and assume for now that we know that $\dim S = 2$. Look at $\mathfrak{m}_2 = (\mathfrak{p}i, x)$ and $\mathfrak{m}_1 = (\mathfrak{p}ix - 1)$. Note that $\mathfrak{m}_1$ is maximal because $S/\mathfrak{m}_1 = K$. It is easy to show that $ht(\mathfrak{m}_1) = 1$. However, $\mathfrak{m}_2 \supsetneq (x) \supsetneq (0)$, so $ht(\mathfrak{m}_2) = 2$.

Dimension theory for topological spaces

The present subsec (which consists mostly of exercises) is a digression that may illuminate the notion of Krull dimension.

7.3.26 Definition Let X be a topological space.[3] Recall that X is **irreducible** if cannot be written as the union of two proper closed subsets $F_1, F_2 \subsetneq X$.

We say that a subset of X is irreducible if it is irreducible with respect to the induced topology.

In general, this notion is not valid from the topological spaces familiar from analysis. For instance:

7.3.27 Remark (exercise) Points are the only irreducible subsets of $\mathbb{R}$.

Nonetheless, irreducible sets behave very nicely with respect to certain operations. As you will now prove, if $U \subset X$ is an open subset, then the irreducible closed subsets of U are in bijection with the irreducible closed subsets of X that intersect U.

[3] We do not include the empty space.

7.3.28 Remark (exercise) A space is irreducible if and only if every open set is dense, or alternatively if every open set is connected.

7.3.29 Remark (exercise) Let X be a space, $Y \subset X$ an irreducible subset. Then $\overline{Y} \subset X$ is irreducible.

7.3.30 Remark (exercise) Let X be a space, $U \subset X$ an open subset. Then the map $Z \to Z \cap U$ gives a bijection between the irreducible closed subsets of X meeting U and the irreducible closed subsets of U. The inverse is given by $Z' \to \overline{Z'}$.

As stated above, the notion of irreducibility is useless for spaces like manifolds. In fact, by Remark 7.3.28, a Hausdorff space cannot be irreducible unless it consists of one point. However, for the highly non-Hausdorff spaces encountered in algebraic geometry, this notion is very useful.

Let R be a commutative ring, and $X = \operatorname{Spec} R$.

7.3.31 Remark (exercise) A closed subset $F \subset \operatorname{Spec} R$ is irreducible if and only if it can be written in the form $F = V(\mathfrak{p})$ for $\mathfrak{p} \subset R$ prime. In particular, $\operatorname{Spec} R$ is irreducible if and only if R has one minimal prime.

In fact, spectra of rings are particularly nice: they are **sober spaces.**

7.3.32 Definition A space X is called **sober** if to every irreducible closed $F \subset X$, there is a unique point $\xi \in F$ such that $F = \overline{\{\xi\}}$. This point is called the **generic point.**

7.3.33 Remark (exercise) Check that if X is any topological space and $\xi \in X$, then the closure $\overline{\{\xi\}}$ of the point ξ is irreducible.

7.3.34 Remark (exercise) Show that $\operatorname{Spec} R$ for R a ring is sober.

7.3.35 Remark (exercise) Let X be a space with a cover $\{X_\alpha\}$ by open subsets, each of which is a sober space. Then X is a sober space. (Hint: any irreducible closed subset must intersect one of the X_α, so is the closure of its intersection with that one.)

We now come to the main motivation of this subsec, and the reason for its inclusion here.

7.3.36 Definition Let X be a topological space. Then the **dimension** (or **combinatorial dimension**) of X is the maximal k such that a chain

$$F_0 \subsetneq F_1 \subsetneq \cdots \subsetneq F_k \subset X$$

with the F_i irreducible, exists. This number is denoted $\dim X$ and may be infinite.

7.3.37 Remark (exercise) What is the Krull dimension of $\mathbb{R}$?

7.3.38 Remark (exercise) Let $X = \bigcup X_i$ be the finite union of subsets $X_i \subset X$.

7.3.39 Remark (exercise) Let R be a ring. Then $\dim \operatorname{Spec} R$ is equal to the Krull dimension of R.

Most of the spaces one wishes to work with in standard algebraic geometry have a strong form of compactness. Actually, compactness is the wrong word, since the spaces of algebraic geometry are not Hausdorff.

7.3.40 Definition A space is **noetherian** if every descending sequence of closed subsets $F_0 \supset F_1 \supset \dots$ stabilizes.

7.3.41 Remark (exercise) If R is noetherian, $\operatorname{Spec} R$ is noetherian as a topological space.

The dimension of a tensor product of fields

The following very clear result gives us the dimension of the tensor product of fields.

7.3.42 Theorem (Grothendieck-Sharp) *Let K, L be field extensions of a field k. Then*

$$\dim K \otimes_k L = \min(\text{tr.deg.}K, \text{tr.deg.}L).$$

This result is stated in the errata of **?**, vol IV (4.2.1.5), but that did not make it well-known; apparently it was independently discovered and published again by R. Y. Sharp, ten years later.[4] Note that in general, this tensor product is *not* noetherian.

Proof. We start by assuming K is a finitely generated, purely transcendental extension of k. Then K is the quotient field of a polynomial ring $k[x_1, \dots, x_n]$. It follows that $K \otimes_k L$ is a localization of $L[x_1, \dots, x_n]$, and consequently of dimension at most $n = \text{tr.deg.}K$.

Now the claim is that if $\text{tr.deg.}L > n$, then we have equality

$$\dim K \otimes_k L = n.$$

To see this, we have to show that $K \otimes_k L$ admits an L-homomorphism to L. For then there will be a maximal ideal $\mathfrak{m}$ of $K \otimes_k L$ which comes from a maximal ideal $\mathfrak{M}$ of $L[x_1, \dots, x_n]$ (corresponding to this homomorphism). Consequently, we will have $(K \otimes_k L)_{\mathfrak{m}} = (L[x_1, \dots, x_n])_{\mathfrak{M}}$, which has dimension n.

So we need to produce this homomorphism $K \otimes_k L \to L$. Since $K = k(x_1, \dots, x_n)$ and L has transcendence degree more than n, we just choose n algebraically independent elements of L, and use that to define a map of k-algebras $K \to L$. By the universal property of the tensor product, we get a section $K \otimes_k L \to L$. This proves the result in the case where K is a finitely generated, purely transcendental extension.

Now we assume that K has finite transcendence degree over k, but is not necessarily purely transcendental. Then K contains a subfield E which is purely transcendental over k and such that E/K is algebraic. Then $K \otimes_k L$ is *integral* over its subring $E \otimes_k L$. The previous analysis applies to $E \otimes_k L$, and by integrality the dimensions of the two objects are the same.

Finally, we need to consider the case when K is allowed to have infinite transcendence degree over k. Again, we may assume that K is the quotient field of the polynomial ring $k[\{x_\alpha\}]$ (by the

[4] Thanks to Georges Elencwajg for a helpful discussion at `http://math.stackexchange.com/questions/56669/a-tensor-product-of-a-power-series/56794`.

integrality argument above). We need to show that if L has *larger* transcendence degree than K, then $\dim K \otimes_k L = \infty$. As before, there is a section $K \otimes_k L \to L$, and $K \otimes_k L$ is a localization of the polynomial ring $L[\{x_\alpha\}]$. If we take the maximal ideal in $L[\{x_\alpha\}]$ corresponding to this section $K \otimes_k L \to L$, it is of the form $(x_\alpha - t_\alpha)_\alpha$ for the $t_\alpha \in L$. It is easy to check that the localization of $L[\{x_\alpha\}]$ at this maximal ideal, which is a localization of $K \otimes_k L$, has infinite dimension. $\square$

III.8. Completions

The algebraic version of completion is essentially analogous to the familiar process of completing a metric space as in analysis, i.e. the process whereby $\mathbb{R}$ is constructed from $\mathbb{Q}$. Here, however, the emphasis will be on how the algebraic properties and structure pass to the completion. For instance, we will see that the dimension is invariant under completion for noetherian local rings.

Completions are used in geometry and number theory in order to give a finer picture of local structure; for example, taking completions of rings allows for the recovery of a topology that looks more like the Euclidean topology as it has more open sets than the Zariski topology. Completions are also used in algebraic number theory to allow for the study of fields around a prime number (or prime ideal).

8.1. Introduction

Motivation

Let R be a commutative ring. Consider a maximal ideal $\mathfrak{m} \in \operatorname{Spec} R$. If one thinks of $\operatorname{Spec} R$ as a space, and R as a collection of functions on that space, then $R_{\mathfrak{m}}$ is to be interpreted as the collection of "germs" of functions defined near the point $\mathfrak{m}$. (In the language of schemes, $R_{\mathfrak{m}}$ is the *stalk* of the structure sheaf.)

However, the Zariski topology is coarse, making it difficult small neighborhoods of $\mathfrak{m}$. Thus the word "near" is to be taken with a grain of salt.

8.1.1 Example Let X be a compact Riemann surface, and let $x \in X$. Let R be the ring of holomorphic functions on $X - \{x\}$ which are meromorphic at x. In this case, $\operatorname{Spec} R$ has the ideal (0) and maximal ideals corresponding to functions vanishing at some point in $X - \{x\}$. So $\operatorname{Spec} R$ is $X - \{x\}$ together with a "generic" point.

Let us just look at the closed points. If we pick $y \in X - \{x\}$, then we can consider the local ring $R_y = \left\{s^{-1}r, s(y) \neq 0\right\}$. This ring is a direct limit of the rings $\mathcal{O}(U)$ of holomorphic functions on open sets U that extend meromorphically to X. Here, however, U ranges only over open subsets of X containing y that are the nonzero loci of elements R. Thus U really ranges over complements of finite subsets. It does not range over open sets in the *complex* topology.

Near y, X looks like $\mathbb{C}$ in the *complex* topology. In the Zariski topology, this is not the case. Each localization R_y actually remembers the whole Riemann surface. Indeed, the quotient field of R_y is the rational function field of X, which determines X. Thus R_y remembers too much, and it fails to give a truly local picture near y.

We would like a variant of localization that would remember much less about the global topology.

Definition

8.1.2 Definition Let R be a commutative ring and $I \subset R$ an ideal. Then we define the **completion of R at I** as

$$\hat{R}_I = \varprojlim R/I^n.$$

By definition, this is the inverse limit of the quotients R/I^n, via the tower of commutative rings

$$\dots \to R/I^3 \to R/I^2 \to R/I$$

where each map is the natural reduction map. Note that $\hat{R}_I$ is naturally an R-algebra. If the map $R \to \hat{R}_I$ is an isomorphism, then R is said to be **I-adically complete.**

In general, though, we can be more general. Suppose R is a commutative ring with a linear topology. Consider a neighborhood basis at the origin consisting of ideals $\{I_\alpha\}$.

8.1.3 Definition The **completion** $\hat{R}$ of the topological ring R is the inverse limit R-algebra

$$\varprojlim R/I_\alpha,$$

where the maps $R/I_\alpha \to R/I_\beta$ for $I_\alpha \subset I_\beta$ are the obvious ones. $\hat{R}$ is given a structure of a topological ring via the inverse limit topology.

If the map $R \to \hat{R}$ is an isomorphism, then R is said to be **complete.**

The collection of ideals $\{I_\alpha\}$ is a directed set, so we can talk about inverse limits over it. When we endow R with the I-adic topology, we see that the above definition is a generalization of Definition 8.1.2.

8.1.4 Remark (exercise) Let R be a linearly topologized ring. Then the map $R \to \hat{R}$ is injective if and only if $\bigcap I_\alpha = 0$ for the I_α open ideals; that is, if and only if R is *Hausdorff.*

8.1.5 Remark (exercise) If R/I_α is finite for each open ideal $I_\alpha \subset R$, then $\hat{R}$ is compact as a topological ring. (Hint: Tychonoff's theorem.)

To be added: Notation needs to be worked out for the completion

The case of a local ring is particularly important. Let R be a local ring and $\mathfrak{m}$ its maximal ideal. Then the completion of R with respect to $\mathfrak{m}$, denoted $\hat{R}$, is the inverse limit $\hat{R} = \lim_{\leftarrow}(R/\mathfrak{m}^n R)$. We then topologize $\hat{R}$ by setting powers of $\mathfrak{m}$ to be basic open sets around 0. The topology formed by these basic open sets is called the "Krull" or "$\mathfrak{m}$-adic topology."

In fact, the case of local rings is the most important one. Usually, we will complete R at *maximal* ideals. If we wanted to study R near a prime $\mathfrak{p} \in \operatorname{Spec} R$, we might first replace R by $R_\mathfrak{p}$, which is a local ring; we might make another approximation to R by completing $R_\mathfrak{p}$. Then we get a *complete* local ring.

8.1.6 Definition Let R be a ring, M an R-module, $I \subset R$ an ideal. We define the **completion of M at I** as

$$\hat{M}_I = \varprojlim M/I^n M.$$

This is an inverse limit of R-modules, so it is an R-module. Furthermore, it is even an $\hat{R}_I$-module, as one easily checks. It is also functorial.

In fact, we get a functor

$$R - \text{modules} \to \hat{R}_I - \text{modules}.$$

Classical examples

Let us give some examples.

8.1.7 Example Recall that in algebraic number theory, a number field is a finite dimensional algebraic extension of $\mathbb{Q}$. Sitting inside of $\mathbb{Q}$ is the ring of integers, $\mathbb{Z}$. For any prime number $p \in \mathbb{Z}$, we can localize $\mathbb{Z}$ to the prime ideal (p) giving us a local ring $\mathbb{Z}_(p)$. If we take the completion of this local ring we get the p-adic numbers $\mathbb{Q}_p$. Notice that since $\mathbb{Z}_(p)/p^n \cong \mathbb{Z}/p$, this is really the same as taking the inverse limit $\lim_{\leftarrow} \mathbb{Z}/p^n$.

8.1.8 Example Let X be a Riemann surface. Let $x \in X$ be as before, and let R be as before: the ring of meromorphic functions on X with poles only at x. We can complete R at the ideal $\mathfrak{m}_y \subset R$ corresponding to $y \in X - \{x\}$. This is always isomorphic to a power series ring

$$\mathbb{C}[[t]]$$

where t is a holomorphic coordinate at y.

The reason is that if one considers $R/\mathfrak{m}_y^n$, one always gets $\mathbb{C}[t]/(t^n)$, where t corresponds to a local coordinate at y. Thus *these* rings don't remember much about the Riemann surface. They're all isomorphic, for instance.

8.1.9 Remark There is always a map $R \to \hat{R}_I$ by taking the limit of the maps R/I^i.

Noetherianness and completions

A priori, one might think this operation of completion gives a big mess. The amazing thing is that for noetherian rings, completion is surprisingly well-behaved.

8.1.10 Proposition *Let R be noetherian, $I \subset R$ an ideal. Then $\hat{R}_I$ is noetherian.*

Proof. Choose generators $x_1, \dots, x_n \in I$. This can be done as I is finitely generated Consider a power series ring

$$R[[t_1, \dots, t_n]];$$

the claim is that there is a map $R[[t_1 \dots t_n]] \to \hat{R}_I$ sending each t_i to $x_i \in \hat{R}_I$. This is not trivial, since we aren't talking about a polynomial ring, but a power series ring.

To build this map, we want a compatible family of maps

$$R[[t_1,\dots,t_n]] \to R[t_1,\dots,t_n]/(t_1,\dots,t_n)^k \to R/I^k.$$

where the second ring is the polynomial ring where homogeneous polynomials of degree $\geqslant k$ are killed. There is a map from $R[[t_1,\dots,t_n]]$ to the second ring that kills monomials of degree $\geqslant k$. The second map $R[t_1,\dots,t_n]/(t_1,\dots,t_n)^k \to R/I^k$ sends $t_i \to x_i$ and is obviously well-defined.

So we get the map

$$\phi : R[[t_1,\dots,t_n]] \to \hat{R}_I,$$

which I claim is surjective. Let us prove this. Suppose $a \in \hat{R}_I$. Then a can be thought of as a collection of elements $(a_k) \in R/I^k$ which are compatible with one another. We can lift each a_k to some $\overline{a_k} \in R$ in a compatible manner, such that

$$\overline{a_{k+1}} = \overline{a_k} + b_k, \quad b_k \in I^k.$$

Since $b_k \in I^k$, we can write it as

$$b_k = f_k(x_1,\dots,x_n)$$

for f_k a polynomial in R of degree k, by definition of the generators in I^k.

I claim now that

$$a = \phi\left(\sum f_k(t_1,\dots,t_n)\right).$$

The proof is just to check modulo I^k for each k. This we do by induction. When one reduces modulo I^k, one gets a_k (as one easily checks).

As we have seen, $\hat{R}_I$ is the quotient of a power series ring. In the homework, it was seen that $R[[t_1,\dots,t_n]]$ is noetherian; this is a variant of the Hilbert basis theorem proved in class. So $\hat{R}_I$ is noetherian. ☐

In fact, following **?**, we shall sometimes find it convenient to note a generalization of the above argument.

8.1.11 Lemma *Suppose A is a filtered ring, M, N filtered A-modules and $\phi : M \to N$ a morphism of filtered modules. Suppose* $\mathrm{gr}(\phi)$ *surjective and M, N complete; then ϕ is surjective.*

Proof. This will be a straightforward "successive approximation" argument. Indeed, let $\{M_n\}, \{N_n\}$ be the filtrations on M, N. Suppose $n \in N$. We know that there is $m_0 \in M$ such that

$$n - \phi(m_0) \in N_1$$

since $M/M_1 \to N/N_1$ is surjective. Similarly, we can choose $m_1 \in M_1$ such that

$$n - \phi(m_0) - \phi(m_1) \in M_2$$

because $n - \phi(m_0) \in N_1$ and $M_1/M_2 \to N_1/N_2$ is surjective. We inductively continue the sequence $m_2, m_3, \dots$ such that it tends to zero rapidly; we then have that $n - \phi\left(\sum m_i\right) \in \bigcap N_i$, so $n = \phi\left(\sum m_i\right)$ as N is complete. ☐

8.1.12 Theorem *Suppose A is a filtered ring. Let M be a filtered A-module, separated with respect to its topology. If $\mathrm{gr}(M)$ is noetherian over $\mathrm{gr}(A)$, then M is a noetherian A-module.*

Proof. If $N \subset M$, then we can obtain an induced filtration on N such that $\mathrm{gr}(N)$ is a submodule of $\mathrm{gr}(M)$. Since noetherianness equates to the finite generation of each submodule, it suffices to show that if $\mathrm{gr}(M)$ is finitely generated, so is M.

Suppose $\mathrm{gr}(M)$ is generated by homogeneous elements $\overline{e}_1, \dots, \overline{e}_n$ of degrees $d_1, \dots, d_n$, represented by elements $e_1, \dots, e_n \in M$. From this we can define a map

$$A^n \to M$$

sending the ith basis vector to e_i. This will induce a surjection $\mathrm{gr}(A^n) \to \mathrm{gr}(M)$. We will have to be careful, though, exactly how we define the filtration on A^n, because the d_i may have large degrees, and if we are not careful, the map on gr's will be zero.

We choose the filtration such that at the mth level, we get the subgroup of A^n such that the ith coordinate is in I_{n-d_i} (for $\{I_n\}$ the filtration of A). It is then clear that the associated map

$$\mathrm{gr}(A^n) \to \mathrm{gr}(M)$$

has image containing each $\overline{e}_i$. Since A^n is complete with respect to this topology, we find that $A^n \to M$ is surjective by lemma 8.1.11. This shows that M is finitely generated and completes the proof. ∎

8.1.13 Corollary *Suppose A is a ring, complete with respect to the I-adic topology. If A/I is noetherian and I/I^2 a finitely generated A-module, then A is noetherian.*

Proof. Indeed, we need to show that $\mathrm{gr}(A)$ is a noetherian ring (by theorem 8.1.12). But this is the ring

$$A/I \oplus I/I^2 \oplus I^2/I^3 \oplus \dots.$$

It is easy to see that this is generated by I/I^2 as an A/I-algebra. By Hilbert's basis theorem, this is noetherian under the conditions of the result. ∎

corollary 8.1.13 gives another means of showing that if a ring A is noetherian, then its completion $\hat{A}$ with respect to an ideal $I \subset A$ is noetherian. For the algebra $\mathrm{gr}(A)$ (where A is given the I-adic topology) is noetherian because it is finitely generated over A/I. Moreover, $\mathrm{gr}(\hat{A}) = \mathrm{gr}(A)$, so $\hat{A}$ is noetherian.

8.2. Exactness properties

The principal result of this section is:

8.2.1 Theorem *If R is noetherian and $I \subset R$ an ideal, then the construction $M \to \hat{M}_I$ is exact when restricted to finitely generated modules.*

Let's be more precise. If M is finitely generated, and $0 \to M' \to M \to M'' \to 0$ is an exact sequence,[1] then

$$0 \to \hat{M}'_I \to \hat{M}_I \to \hat{M}''_I \to 0$$

is also exact.

We shall prove this theorem in several pieces.

Generalities on inverse limits

For a moment, let us step back and think about exact sequences of inverse limits of abelian groups. Say we have a tower of exact sequences of abelian groups

$$\begin{tikzcd}
0 \arrow[r] & \vdots \arrow[r] \arrow[d] & \vdots \arrow[r] \arrow[d] & \vdots \arrow[r] \arrow[d] & 0 . \\
0 \arrow[r] & A_2 \arrow[r] \arrow[d] & B_2 \arrow[r] \arrow[d] & C_2 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A_1 \arrow[r] \arrow[d] & B_1 \arrow[r] \arrow[d] & C_1 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A_0 \arrow[r] & B_0 \arrow[r] & C_0 \arrow[r] & 0
\end{tikzcd}$$

Then we get a sequence

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n \to 0.$$

In general, it is *not* exact. But it is left-exact.

8.2.2 Proposition *Hypotheses as above,* $0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n$ *is exact.*

Proof. It is obvious that $\phi \circ \psi = 0$.

Let us first show that $\phi : \varprojlim A_n \to \varprojlim B_n$ is injective. So suppose a is in the projective limit, represented by a compatible sequence of elements $(a_k) \in A_k$. If ϕ maps to zero, all the a_k go to zero in B_k. Injectivity of $A_k \to B_k$ implies that each a_k is zero. This implies ϕ is injective.

Now let us show exactness at the next step. Let $\psi : \varprojlim B_n \to \varprojlim C_n$ and let $b = (b_k)$ be in $\ker \psi$. This means that each b_k gets killed when it maps to C_k. This means that each b_k comes from something in a_k. These a_k are unique by injectivity of $A_k \to B_k$. It follows that the a_k have no choice but to be compatible. Thus (a_k) maps into (b_k). So b is in the image of ϕ. □

So far, so good. We get some level of exactness. But the map on the end is not necessarily surjective. Nonetheless:

8.2.3 Proposition $\psi : \varprojlim B_n \to \varprojlim C_n$ *is surjective if each* $A_{n+1} \to A_n$ *is surjective.*

[1] The ends are finitely generated by noetherianness.

Proof. Say $c \in \varprojlim C_n$, represented by a compatible family (c_k). We have to show that there is a compatible family $(b_k) \in \varprojlim B_n$ which maps into c. It is easy to choose the b_k *individually* since $B_k \to C_k$ is surjective. The problem is that a priori we may not get something compatible.

We construct b_k by induction on then, therefore. Assume that b_k which lifts c_k has been constructed. We know that c_k receives a map from c_{k+1}.

$$\begin{tikzcd} & c_{k+1} \arrow[d] \\ b_k \arrow[r] & c_k \end{tikzcd} .$$

Choose any $x \in B_{k+1}$ which maps to c_{k+1}. However, x might not map down to b_k, which would screw up the compatibility conditions. Next, we try to adjust x. Consider $x' \in B_k$ to be the image of x under $B_{k+1} \to B_k$. We know that $x' - b_k$ maps to zero in C_k, because c_{k+1} maps to c_k. So $x' - b_k$ comes from something in A_k, call it a.

$$\begin{tikzcd} x \arrow[r] & c_{k+1} \arrow[d] \\ b_k \arrow[r] & c_k \end{tikzcd} .$$

But a comes from some $\overline{a} \in A_{k+1}$. Then we define

$$b_{k+1} = x - \overline{a},$$

which adjustment doesn't change the fact that b_{k+1} maps to c_{k+1}. However, this adjustment makes b_{k+1} compatible with b_k. Then we construct the family b_k by induction. We have seen surjectivity. $\square$

Now, let us study the exactness of completions.

Proof of Theorem 8.2.1. Let us try to apply the general remarks above to studying the sequence

$$0 \to \hat{M}'_I \to \hat{M}_I \to \hat{M}''_I \to 0.$$

Now $\hat{M}_I = \varprojlim M/I^n$. We can construct surjective maps

$$M/I^n \twoheadrightarrow M''/I^n$$

whose inverse limits lead to $\hat{M}_I \to \hat{M}''_I$. The image is $M/(M' + I^n M)$. What is the kernel? Well, it is $M' + I^n M/I^n M$. This is equivalently

$$M'/M' \cap I^n M.$$

So we get an exact sequence

$$0 \to M'/M' \cap I^n M \to M/I^n M \to M''/I^n M'' \to 0.$$

By the above analysis of exactness of inverse limits, we get an exact sequence

$$0 \to \varprojlim M'/(I^n M \cap M') \to \hat{M}_I \to \hat{M}''_I \to 0.$$

We of course have surjective maps $M'/I^nM' \to M'/(I^nM \cap M')$ though these are generally not isomorphisms. Something "divisible by I^n" in M but in M' is generally not divisible by I^n in M'. Anyway, we get a map

$$\varprojlim M'/I^nM' \to \varprojlim M'/I^nM \cap M'$$

where the individual maps are not necessarily isomorphisms. Nonetheless, I claim that the map on inverse limits is an isomorphism. This will imply that completion is indeed an exact functor.

But this follows because the filtrations $\{I^nM'\}, \{I^nM \cap M'\}$ are equivalent in view of the Artin-Rees lemma, theorem 3.3.1. ◻

Last time, we were talking about completions. We showed that if R is noetherian and $I \subset R$ an ideal, an exact sequence

$$0 \to M' \to M \to M \to 0$$

of finitely generated R-modules leads to a sequence

$$0 \to \hat{M}'_I \to \hat{M}_I \to \hat{M};_I \to 0$$

which is also exact. We showed this using the Artin-Rees lemma.

8.2.4 Remark In particular, for finitely generated modules over a noetherian ring, completion is an **exact functor**: if $A \to B \to C$ is exact, so is the sequence of completions. This can be seen by drawing in kernels and cokernels, and using the fact that completions preserve short exact sequences.

Completions and flatness

Suppose that M is a finitely generated R-module. Then there is a surjection $R^n \twoheadrightarrow M$, whose kernel is also finitely generated as R is noetherian. It follows that M is finitely presented. In particular, there is a sequence

$$R^m \to R^n \to M \to 0.$$

We get an exact sequence

$$\hat{R}^m \to \hat{R}^n \to \hat{M} \to 0$$

where the second map is just multiplication by the same m-by-n matrix as in the first case.

8.2.5 Corollary *If M is finitely generated and R noetherian, there is a canonical isomorphism*

$$\hat{M}_I \simeq M \otimes_R \hat{R}_I.$$

Proof. We know that there is a map $M \to \hat{M}_I$, so the canonical morphism $\phi_M : M \otimes_R \hat{R}_I \to \hat{M}_I$ exists (because this induces a map from $M \otimes_R \hat{R}_I$). We need to check that it is an isomorphism.

If there is an exact sequence $M' \to M \to M'' \to 0$, there is a commutative diagram

$$\begin{tikzcd}
M' \otimes_R \hat{R}_I \arrow[r] \arrow[d, "\phi_{M'}"] & M \otimes_R \hat{R}_I \arrow[r] \arrow[d, "\phi_M"] & M'' \otimes_R \hat{R}_I \arrow[r] \arrow[d] & 0 . \\
\hat{M}'_I \arrow[r] & \hat{M}_I \arrow[r] & \hat{M}''_I \arrow[r] & 0
\end{tikzcd}$$

Exactness of completion and right-exactness of $\otimes$ implies that this diagram is exact. It follows that if $\phi_M, \phi_{M'}$ are isomorphisms, so is $\phi_{M''}$.

But any M'' appears at the end of such a sequence with M', M are free by the finite presentation argument above. So it suffices to prove ϕ an isomorphism for finite frees, which reduces to the case of ϕ_R an isomorphism. That is obvious. ☐

8.2.6 Corollary *If R is noetherian, then $\hat{R}_I$ is a flat R-module.*

Proof. Indeed, tensoring with $\hat{R}_I$ is exact (because it is completion, and completion is exact) on the category of finitely generated R-modules. Exactness on the category of all R-modules follows by taking direct limits, since every module is a direct limit of finitely generated modules, and direct limits preserve exactness. ☐

8.2.7 Remark Warning: $\hat{M}_I$ is, in general, not $M \otimes_R \hat{R}_I$ when M is not finitely generated. One example to think about is $M = \mathbb{Z}[t]$, $R = \mathbb{Z}$. The completion of M at $I = (p)$ is the completion of $\mathbb{Z}[t]$ at $p\mathbb{Z}[t]$, which contains elements like

$$1 + pt + p^2t^2 + \dots,$$

which belong to the completion but not to $\hat{R}_I \otimes M = \mathbb{Z}_p[t]$.

8.2.8 Remark By the Krull intersection theorem, if R is a local noetherian ring, then the map from $R \to \hat{R}$ is an injection.

8.3. Hensel's lemma

One thing that you might be interested in doing is solving Diophantine equations. Say $R = \mathbb{Z}$; you want to find solutions to a polynomial $f(X) \in \mathbb{Z}[X]$. Generally, it is very hard to find solutions. However, there are easy tests you can do that will tell you if there are no solutions. For instance, reduce mod a prime. One way you can prove that there are no solutions is to show that there are no solutions mod 2.

But there might be solutions mod 2 and yet you might not be sure about solutions in $\mathbb{Z}$. So you might try mod 4, mod 8, and so on—you get a whole tower of problems to consider. If you manage to solve all these equations, you can solve the equations in the 2-adic integers $\mathbb{Z}_2 = \hat{\mathbb{Z}}_{(2)}$. But the Krull intersection theorem implies that $\mathbb{Z} \to \mathbb{Z}_2$ is injective. So if you expected that there was a unique solution in $\mathbb{Z}$, you might try looking at the solutions in $\mathbb{Z}_2$ to be the solutions in $\mathbb{Z}$.

The moral is that solving an equation over $\mathbb{Z}_2$ is intermediate in difficulty between $\mathbb{Z}/2$ and $\mathbb{Z}$. Nonetheless, it turns out that solving an equation mod $\mathbb{Z}/2$ is very close to solving it over $\mathbb{Z}_2$, thanks to Hensel's lemma.

The result

8.3.1 Theorem (Hensel's Lemma) *Let R be a noetherian ring, $I \subset R$ an ideal. Let $f(X) \in R[X]$ be a polynomial such that the equation $f(X) = 0$ has a solution $a \in R/I$. Suppose, moreover, that $f'(a)$ is invertible in R/I.*

Then a lifts uniquely to a solution of the equation $f(X) = 0$ in $\hat{R}_I$.

8.3.2 Example Let $R = \mathbb{Z}, I = (5)$. Consider the equation $f(x) = x^2 + 1 = 0$ in R. This has a solution modulo five, namely 2. Then $f'(2) = 4$ is invertible in $\mathbb{Z}/5$. So the equation $x^2 + 1 = 0$ has a solution in $\mathbb{Z}_5$. In other words, $\sqrt{-1} \in \mathbb{Z}_5$.

Let's prove Hensel's lemma.

Proof. Now we have $a \in R/I$ such that $f(a) = 0 \in R/I$ and $f'(a)$ is invertible. The claim is going to be that for each $m \geqslant 1$, there is a *unique* element $a_n \in R/I^n$ such that

$$a_n \to a\ (I), \quad f(a_n) = 0 \in R/I^n.$$

Uniqueness implies that this sequence (a_n) is compatible, and thus gives the required element of the completion. It will be a solution of $f(X) = 0$ since it is a solution at each element of the tower.

Let us now prove the claim. For $n = 1$, $a_1 = a$ necessarily. The proof is induction on n. Assume that a_n exists and is unique. We would like to show that a_{n+1} exists and is unique. Well, if it is going to exist, when we reduce a_{n+1} modulo I^n, we must get a_n or uniqueness at the n-th step would fail.

So let $\overline{a}$ be any lifting of a_n to R/I^{n+1}. Then a_{n+1} is going to be that lifting plus some $\epsilon \in I^n/I^{n+1}$. We want

$$f(\overline{a} + \epsilon) = 0 \in R/I^{n+1}.$$

But this is

$$f(\overline{a}) + \epsilon f'(\overline{a})$$

because $\epsilon^2 = 0 \in R/I^{n+1}$. However, this lets us solve for ϵ, because then necessarily $\epsilon = \frac{-f(\overline{a})}{f'(\overline{a})} \in I^n$. Note that $f'(\overline{a}) \in R/I^{n+1}$ is invertible. If you believe this for a moment, then we have seen that ϵ exists and is unique; note that $\epsilon \in I^n$ because $f(\overline{a}) \in I^n$.

8.3.3 Lemma *$f'(\overline{a}) \in R/I^{n+1}$ is invertible.*

Proof. If we reduce this modulo R/I, we get the invertible element $f'(a) \in R/I$. Note also that the I/I^{n+1} is a nilpotent ideal in R/I^{n+1}. So we are reduced to showing, more generally:

8.3.4 Lemma *Let A be a ring,[2] J a nilpotent ideal.[3] Then an element $x \in A$ is invertible if and only if its reduction in A/J is invertible.*

[2] E.g. R/I^{n+1}.

[3] E.g. $J = I/I^{n+1}$.

Proof. One direction is obvious. For the converse, say $x \in A$ has an invertible image. This implies that there is $y \in A$ such that $xy \equiv 1 \mod J$. Say

$$xy = 1 + m,$$

where $m \in J$. But $1 + m$ is invertible because

$$\frac{1}{1+m} = 1 - m + m^2 \pm \dots .$$ □

The expression makes sense as the high powers of m are zero. So this means that $y(1+m)^{-1}$ is the inverse to x. □

This was one of many versions of Hensel's lemma. There are many ways you can improve on a statement. The above version says something about "nondegenerate" cases, where the derivative is invertible. There are better versions which handle degenerate cases.

8.3.5 Example Consider $x^2 - 1$; let's try to solve this in $\mathbb{Z}_2$. Well, $\mathbb{Z}_2$ is a domain, so the only solutions can be ± 1. But these have the same reduction in $\mathbb{Z}/2$. The lifting of the solution is non-unique.

The reason why Hensel's lemma fails is that $f'(\pm 1) = \pm 2$ is not invertible in $\mathbb{Z}/2$. But it is not far off. If you go to $\mathbb{Z}/4$, we do get two solutions, and the derivative is at least nonzero at those places.

One possible extension of Hensel's lemma is to allow the derivative to be noninvertible, but at least to bound the degree to which it is noninvertible. From this you can get interesting information. But then you may have to look at equations R/I^n instead of just R/I, where n depends on the level of noninvertibility.

Let us describe the multivariable Hensel lemma.

8.3.6 Theorem *Let $f_1, \dots, f_n$ be polynomials in n variables over the ring R. Let J be the Jacobian matrix $(\frac{\partial f_i}{\partial x_j})$. Suppose $\Delta = \det J \in R[x_1, \dots, x_n]$.*

If the system $\{f_i(x) = 0\}$ has a solution $a \in (R/I)^n$ in R/I for some ideal I satisfying the condition that $\Delta(a)$ is invertible, then there is a unique solution of $\{f_i(x) = 0\}$ in $\hat{R}_I^n$ which lifts a.

The proof is the same idea: successive approximation, using the invertibility of Δ.

The classification of complete DVRs (characteristic zero)

Let R be a complete DVR with maximal ideal $\mathfrak{m}$ and quotient field F. We let $k := R/\mathfrak{m}$; this is the **residue field** and is, e.g., the integers mod p for the p-adic integers.

The main result that we shall prove is the following:

8.3.7 Theorem *Suppose k is of characteristic zero. Then $R \simeq k[[X]]$, the power series ring in one variable, with respect to the usual discrete valuation on $k[[X]]$.*

The "usual discrete valuation" on the power series ring is the order at zero. Incidentally, this applies to the (non-complete) subring of $\mathbb{C}[[X]]$ consisting of power series that converge in some neighborhood of zero, which is the ring of germs of holomorphic functions at zero; the valuation again measures the zero at $z = 0$.

To prove it (following **?**), we need to introduce another concept. A **system of representatives** is a set $S \subset R$ such that the reduction map $S \to k$ is bijective. A **uniformizer** is a generator of the maximal ideal $\mathfrak{m}$. Then:

8.3.8 Proposition *If S is a system of representatives and π a uniformizer, we can write each $x \in R$ uniquely as*

$$x = \sum_{i=0}^{\infty} s_i \pi^i, \quad \text{where } s_i \in S.$$

Proof. Given x, we can find by the definitions $s_0 \in S$ with $x - s_0 \in \pi R$. Repeating, we can write $x - s_0\ \pi \in R$ as $x - s_0\ \pi - s_1 \in \pi R$, or $x - s_0 - s_1 \pi \in \pi^2 R$. Repeat the process inductively and note that the differences $x - \sum_{i=0}^{n} s_i \pi^i \in \pi^{n+1} R$ tend to zero.

In the p-adic numbers, we can take $\{0, \ldots, p-1\}$ as a system of representatives, so we find each p-adic integer has a unique p-adic expansion $x = \sum_{i=0}^{\infty} x_i p^i$ for $x_i \in \{0, \ldots, p-1\}$. ∎

We now prove the first theorem.

Proof. Note that $\mathbb{Z} - 0 \subset R$ gets sent to nonzero elements in the residue field k, which is of characteristic zero. This means that $\mathbb{Z} - 0 \subset R$ consists of units, so $\mathbb{Q} \subset R$.

Let $L \subset R$ be a subfield. Then $L \simeq \overline{L} \subset k$; if $t \in k - \overline{L}$, I claim that there is $L' \supset R$ containing L with $t \in \overline{L'}$.

If t is transcendental, lift it to $T \in R$; then T is transcendental over L and is invertible in R, so we can take $L' := L(T)$.

If the minimal polynomial of t over $\overline{L}$ is $\overline{f}(X) \in k[X]$, we have $\overline{f}(t) = 0$. Moreover, $\overline{f}'(t) \neq 0$ because these fields are of characteristic zero and all extensions are separable. So lift $\overline{f}(X)$ to $f(X) \in R[X]$; by Hensel lift t to $u \in R$ with $f(u) = 0$. Then f is irreducible in $L[X]$ (otherwise we could reduce a factoring to get one of $\overline{f} \in \overline{L}[X]$), so $L[u] = L[X]/(f(X))$, which is a field L'.

So if $K \subset R$ is the maximal subfield (use Zorn's lemma), this is our system of representatives by the above argument. ∎

8.4. Henselian rings

There is a substitute for completeness that captures the essential properties: Henselianness. A ring is Henselian if it satisfies Hensel's lemma, more or less. We mostly follow **?** in the treatment.

Semilocal rings

To start with, we shall need a few preliminaries on semi-local rings.

Fix a local ring A with maximal ideal $\mathfrak{m} \subset A$. Fix a finite A-algebra B; by definition, B is a finitely generated A-module.

8.4.1 Proposition *Hypotheses as above, the maximal ideals of B are in bijection with the prime ideals of B containing $\mathfrak{m}B$, or equivalently the prime ideals of $\overline{B} = B \otimes_A A/\mathfrak{m}$.*

Proof. We have to show that every maximal ideal of B contains $\mathfrak{m}B$. Suppose $\mathfrak{n} \subset B$ was maximal and was otherwise. Then by Nakayama's lemma, $\mathfrak{n} + \mathfrak{m}B \neq B$ is a proper ideal strictly containing $\mathfrak{n}$; this contradicts maximality.

It is now clear that the maximal ideals of B are in bijection naturally with those of $\overline{B}$. However, $\overline{B}$ is an artinian ring, as it is finite over the field $A/\mathfrak{m}$, so every prime ideal in it is maximal. □

The next thing to observe is that $\overline{B}$, as an artinian ring, decomposes as a product of local artinian rings. In fact, this decomposition is unique. However, this does not mean that B itself is a product of local rings (B is not necessarily artinian). Nonetheless, if such a splitting exists, it is necessarily unique.

8.4.2 Proposition *Suppose $R = \prod R_i$ is a finite product of local rings R_i. Then the R_i are unique.*

Proof. To give a decomposition $R = \prod R_i$ is equivalent to giving idempotents e_i. If we had another decomposition $R = \prod S_j$, then we would have new idempotents f_j. The image of each f_j in each R_i is either zero or one as a local ring has no nontrivial idempotents. From this, one can easily deduce that the f_j's are sums of the e_i's, and if the S_j are local, one sees that the S_j's are just the R_i's permuted. □

In fact, there is a canonical way of determining the factors R_i. A finite product of local rings as above is *semi-local*; the maximal ideals $\mathfrak{m}_i$ are finite in number, and, furthermore, the canonical map

$$R \to \prod R_{\mathfrak{m}_i}$$

is an isomorphism.

In general, this splitting **fails** for semi-local rings, and in particular for rings finite over a local ring. We have seen that this splitting nonetheless works for rings finite over a field.

To recapitulate, we can give a criterion for when a semi-local ring splits as above.

8.4.3 Proposition *Let R be a semilocal ring with maximal ideals $\mathfrak{m}_1, \dots, \mathfrak{m}_k$. Then R splits into local factors if and only if, for each i, there is an idempotent $e_i \in \bigcap_{j \neq i} \mathfrak{m}_j - \mathfrak{m}_i$. Then the rings Re_i are local and $R = \prod Re_i$.*

Proof. If R splits into local factors, then clearly we can find such idempotents. Conversely, suppose given the e_i. Then for each $i \neq j$, $e_i e_j$ is an idempotent e_{ij} that belongs to all the maximal ideals $\mathfrak{m}_k$. So it is in the Jacobson radical. But then $1-e_{ij}$ is invertible, so $e_{ij}(1-e_{ij}) = 0$ implies that $e_{ij} = 0$.

It follows that the $\{e_i\}$ are *orthogonal* idempotents. To see that $R = \prod Re_i$ as rings, we now need only to see that the $\{e_i\}$ form a *complete* set; that is, $\sum e_i = 1$. But the sum $\sum e_i$ is an idempotent itself since the e_i are mutually orthogonal. Moreover, the sum $\sum e_i$ belongs to no $\mathfrak{m}_i$, so it is invertible, thus equal to 1. The claim is now clear, since each Re_i is local by assumption. □

Note that if we can decompose a semilocal ring into a product of local rings, then we can go no further in a sense—it is easy to check that a local ring has no nontrivial idempotents.

Henselian rings

8.4.4 Definition A local ring $(R, \mathfrak{m})$ is **henselian** if every finite R-algebra is a product of local R-algebras.

It is clear from the remarks of the previous section that the decomposition as a product of local algebras is unique. Furthermore, we have already seen:

8.4.5 Proposition *A field is henselian.*

Proof. Indeed, then any finite algebra over a field is artinian (as a finite-dimensional vector space). □

This result was essentially a corollary of basic facts about artinian rings. In general, though, henselian rings are very far from artinian. For instance, we will see that every *complete* local ring is henselian.

We continue with a couple of further easy claims.

8.4.6 Proposition *A local ring that is finite over a henselian ring is henselian.*

Proof. Indeed, if R is a henselian local ring and S a finite R-algebra, then every finite S-algebra is a finite R-algebra, and thus splits into a product of local rings. □

We have seen that henselianness of a local ring $(R, \mathfrak{m})$ with residue field k is equivalent to the condition that every finite R-algebra S splits into a product of local rings. Since $S \otimes_R k$ always splits into a product of local rings, and this splitting is unique, we see that if a splitting of S exists, it necessarily lifts the splitting of $S \otimes_R k$.

Since a "splitting" is the same thing (by proposition 8.4.3) as a complete collection of idempotents, one for each maximal ideal, we are going to characterize henselian rings by the property that one can lift idempotents from the residue ring.

8.4.7 Definition A local ring $(R, \mathfrak{m})$ **satisfies lifting idempotents** if for every finite R-algebra S, the canonical (reduction) map between idempotents of S and those of $S/\mathfrak{m}S$ is surjective.

Recall that there is a functor Idem from rings to sets that sends each ring to its collection of idempotents. So the claim is that the natural map $\mathrm{Idem}(S) \to \mathrm{Idem}(S/\mathfrak{m}S)$ is a surjection.

In fact, in this case, we shall see that the map $\mathrm{Idem}(S) \to \mathrm{Idem}(S/\mathfrak{m}S)$ is even injective.

8.4.8 Proposition *The map from idempotents of S to those of $S/\mathfrak{m}S$ is always injective.*

We shall not even use the fact that S is a finite R-algebra here.

Proof. Suppose $e, e' \in S$ are idempotents whose images in $S/\mathfrak{m}S$ are the same. Then

$$(e-e')^3 = e^3 - 3e^2e' + 3e'^2e - e'^3 = e^3 - e'^3 = e - e.$$

Thus if we let $x = e - e'$, we have $x^3 - x = 0$, and x belongs to $\mathfrak{m}S$. Thus

$$x(1-x^2) = 0,$$

and $1 - x^2$ is invertible in S (because x^2 belongs to the Jacobson radical of S). Thus $x = 0$ and $e = e'$. ☐

With this, we now want a characterization of henselian rings in terms of the lifting idempotents property.

8.4.9 Proposition *Suppose $(R, \mathfrak{m})$ satisfies lifting idempotents, and let S be a finite R-algebra. Then given orthogonal idempotents $\overline{e}_1, \dots, \overline{e}_n$ of $S/\mathfrak{m}S$, there are mutually orthogonal lifts $\{e_i\} \in S$.*

The point is that we can make the lifts mutually orthogonal. (Recall that idempotents are *orthogonal* if their product is zero.)

Proof. Indeed, by assumption we can get lifts $\{e_i\}$ which are idempotent; we need to show that they are mutually orthogonal. But in any case e_ie_j for $i \neq j$ is an idempotent, which lies in $\mathfrak{m}S \subset S$ and thus in the Jacobson radical. It follows that $e_ie_j = 0$, proving the orthogonality. ☐

8.4.10 Proposition *A local ring is henselian if and only if it satisfies lifting idempotents.*

Proof. Suppose first $(R, \mathfrak{m})$ satisfies lifting idempotents. Let S be any finite R-algebra. Then $S/\mathfrak{m}S$ is artinian, so factors as a product of local artinian rings $\prod \overline{S}_i$. This factorization corresponds to idempotents $\overline{e}_i \in S/\mathfrak{m}S$. We can lift these to orthogonal idempotents $e_i \in S$ by proposition 8.4.9. These idempotents correspond to a decomposition

$$S = \prod S_i$$

which lifts the decomposition $\overline{S} = \prod \overline{S}_i$. Since the $\overline{S_i}$ are local, so are the S_i. Thus R is henselian.

Conversely, suppose R henselian. Let S be a finite R-algebra and let $\overline{e} \in \overline{S} = S/\mathfrak{m}S$ be idempotent. Since $\overline{S}$ is a product of local rings, $\overline{e}$ is a finite sum of the primitive idempotents in $\overline{S}$. By henselianness, each of these primitive idempotents lifts to S, so $\overline{e}$ does too. ☐

8.4.11 Proposition *Let R be a local ring and $I \subset R$ an ideal consisting of nilpotent elements. Then R is henselian if and only if R/I is.*

Proof. One direction is clear by proposition 8.4.6. For the other, suppose R/I is henselian. Let $\mathfrak{m} \subset R$ be the maximal ideal. Let S be any finite R-algebra; we have to show surjectivity of

$$\mathrm{Idem}(S) \to \mathrm{Idem}(S/\mathfrak{m}S).$$

However, we are given that, by henselianness of S/I,

$$\mathrm{Idem}(S/IS) \to \mathrm{Idem}(S/\mathfrak{m}S)$$

is a surjection. Now we need only observe that $\mathrm{Idem}(S) \to \mathrm{Idem}(S/IS)$ is a bijection. This follows because idempotents in S (resp. S/IS) correspond to disconnections of $\operatorname{Spec} S$ (resp. $\operatorname{Spec} S/IS$) by **??**. However, as I consists of nilpotents, $\operatorname{Spec} S$ and $\operatorname{Spec} S/IS$ are homeomorphic naturally. □

Hensel's lemma

We now want to show that Hensel's lemma is essentially what characterizes henselian rings, which explains the name. Throughout, we use the symbol to denote reduction mod an ideal (usually $\mathfrak{m}$ or $\mathfrak{m}$ times another ring).

8.4.12 Proposition *Let $(R, \mathfrak{m})$ be a local ring with residue field k. Then R is henselian if and only if, whenever a monic polynomial $P \in R[X]$ satisfies*

$$\overline{P} = \overline{QR} \in k[X],$$

for some relatively prime polynomials $\overline{Q}, \overline{R} \in k[X]$, then the factorization lifts to a factorization

$$P = QR \in R[X].$$

This notation should be improved.

Proof. Suppose R henselian and suppose P is a polynomial whose reduction admits such a factorization. Consider the finite R-algebra $S = R[X]/(P)$; since $\overline{S} = S/\mathfrak{m}S$ can be represented as $k[X]/(\overline{P})$, it admits a splitting into components

$$\overline{S} = k[X]/(\overline{Q}) \times k[X]/(\overline{R}).$$

Since R is henselian, this splitting lifts to S, and we get a splitting

$$S = S_1 \times S_2.$$

Here $S_1 \otimes k \simeq k[X]/(\overline{Q})$ and $S_2 \otimes k \simeq k[X]/(\overline{R})$. The image of X in $S_1 \otimes k$ is annihilated by $\overline{Q}$, and the image of X in $S_2 \otimes k$ is annihilated by $\overline{R}$.

8.4.13 Lemma *Suppose R is a local ring, S a finite R-algebra generated by an element $x \in S$. Suppose the image $\overline{x} \in \overline{S} = S \otimes_R k$ satisfies a monic polynomial equation $u(\overline{x}) = 0$. Then there is a monic polynomial U lifting u such that $U(x) = 0$ (in S).*

Proof. Let $\overline{x} \in \overline{S}$ be the generating element that satisfies $u(\overline{x}) = 0$, and let $x \in S$ be a lift of it. Suppose u has rank n. Then $1, x, \dots, x^{n-1}$ spans S by Nakayama's lemma. Thus there is a monic polynomial U of degree n that annihilates x; the reduction must be a multiple of u, hence u.

Returning to the proposition, we see that the image of the generator X in S_1, S_2 must satisfy polynomial equations Q, R that lift $\overline{Q}, \overline{R}$. Thus X satisfies QR in $S[X]/(P)$; in other words, QR is a multiple of P, hence equal to P. Thus we have lifted the factorization $\overline{P} = \overline{QR}$. This proves that factorizations can be lifted.

Now, let us suppose that factorizations can always be lifted for finite R-algebras. We are now going to show that R satisfies lifting idempotents. Suppose S is a finite R-algebra, $\overline{e}$ a primitive idempotent in $\overline{S}$. We can lift $\overline{e}$ to some element $e' \in S$. Since e' is contained in a finite R-algebra that contains R, we know that e' is *integral* over R, so that we can find a map $R[X]/(P) \to S$ sending the generator $X \mapsto e'$, for some polynomial P. We are going to use the fact that $R[X]/(P)$ splits to lift the idempotent $\overline{e}$.

Let $\mathfrak{m}_1, \dots, \mathfrak{m}_k$ be the maximal ideals of S. These equivalently correspond to the points of $\operatorname{Spec} \overline{S}$. We know that e' belongs precisely to one of the $\mathfrak{m}_i$ (because a primitive idempotent in $\overline{S}$ is one on one maximal ideal and zero elsewhere). Call this $\mathfrak{m}_1$, say.

We have a map $\operatorname{Spec} S \to \operatorname{Spec} R[X]/(P)$ coming from the map $\phi : R[X]/(P) \to S$. We claim that the image of $\mathfrak{m}_1$ is different from the images of the $\mathfrak{m}_j, j > 1$. Indeed, $b \in \mathfrak{m}_j$ precisely for $j > 1$, so the image of $\mathfrak{m}_1$ does not contain X. However, the image of $\mathfrak{m}_j, j > 1$ does contain X.

Consider a primitive idempotent for $R[X]/(P)$ corresponding to $\phi^{-1}(\mathfrak{m}_1)$, say f. Then f belongs to every other maximal ideal of $R[X]/(P)$ but not to $\phi^{-1}(\mathfrak{m}_1)$. Thus $\phi(f)$, which is idempotent, belongs to $\mathfrak{m}_1$ but not to any other maximal ideal of S. It follows that $\phi(f)$ must lift $\overline{e}$, and we have completed the proof. ∎

8.4.14 Corollary *If every monogenic,[4] finitely presented and finite R-algebra is a product of local rings, then R is henselian.*

Proof. Indeed, the proof of the above result shows that if $R[X]/(P)$ splits for every monic P, then R is henselian. ∎

From the above result, we can get a quick example of a non-complete henselian ring:

8.4.15 Example The integral closure of the localization $\mathbb{Z}_{(p)}$ in the ring $\mathbb{Z}_p$ of p-adic integers is a henselian ring. Indeed, it is first of all a discrete valuation ring (as we can restrict the valuation on $\mathbb{Z}_p$; note that an element of $\mathbb{Q}_p$ which is algebraic over $\mathbb{Q}$ and has norm at most one is *integral* over $\mathbb{Z}_{(p)}$). This follows from the criterion of proposition 8.4.12. If a monic polynomial P factors in the residue field, then it factors in $\mathbb{Z}_p$, and if P has coefficients integral over $\mathbb{Z}_{(p)}$, so does any factor.

[4] That is, generated by one element.

8.4.16 Example If k is a complete field with a nontrivial absolute value and X is any topological space, we can consider for each open subset $U \subset X$ the ring $\mathcal{A}(U)$ of continuous maps $U \to k$. As U ranges over the open subsets containing an element x, the colimit $\varinjlim \mathcal{A}(U)$ (the "local ring" at x) is a local henselian ring. See **?**.

8.4.17 Proposition *Let $(R_i, \mathfrak{m}_i)$ be an inductive system of local rings and local homomorphisms. If each R_i is henselian, then the colimit $\varinjlim R_i$ is henselian too.*

Proof. We already know (**??**) that the colimit is a local ring, and that the maximal ideal of $\varinjlim R_i$ is the colimit $\varinjlim \mathfrak{m}_i$. Finally, given any monic polynomial in $\varinjlim R_i$ with a factoring in the residue field, the polynomial and the factoring come from some finite R_i; the henselianness of R_i allows us to lift the factoring. ☐

Example: Puiseux's theorem

Using the machinery developed here, we are going to prove:

8.4.18 Theorem *Let K be an algebraically closed field of characteristic zero. Then any finite extension of the field of meromorphic power series*[5] *$K((T))$ is of the form $K((T^{1/n}))$ for some n.*

In particular, we see that any finite extension of $K((T))$ is abelian, even cyclic. The idea is going to be to look at the integral closure of $K[[T]]$ in the finite extension, argue that it itself is a DVR, and then refine an "approximate" root in this DVR of the equation $\alpha^n = T$ to an exact one.

Proof. Let $R = K[[T]]$ be the power series ring; it is a complete, and thus henselian, DVR. Let L be a finite extension of $K((T))$ of degree n and S the integral closure of R in S, which we know to be a DVR. This is a finite R-algebra (cf. **??**), so S is a product of local domains. Since S is a domain, it is itself local. It is easy to see that if $\mathfrak{n} \subset S$ is the maximal ideal, then S is $\mathfrak{n}$-adically complete (for instance because the maximal ideal of R is a power of $\mathfrak{n}$, and S is a free R-module).

Let $\mathfrak{m} \subset R$ be the maximal ideal. We have the formula $ef = n$, because there is only one prime of S lying above $\mathfrak{m}$. But $f = 1$ as the residue field of R is algebraically closed. Hence $e = n$, and the extension is *totally* ramified.

Let $\alpha \in S$ be a uniformizer; we know that α is congruent, modulo $\mathfrak{n}^2$, to something in R as the residue extension is trivial. Then α^n is congruent to something in R, which must be a uniformizer by looking at the valuation. By rescaling, we may assume

$$\alpha^n \equiv T \mod \mathfrak{n}^2.$$

Since the polynomial $X^n - T$ is separable in the residue field, we can (using Hensel's lemma) refine α to a new $\alpha' \equiv \alpha \mod \mathfrak{n}^2$ with

$$\alpha'^n = T.$$

Then α' is also a uniformizer at $\mathfrak{n}$ (as $\alpha' \equiv \alpha \mod \mathfrak{n}^2$). It follows that $R[\alpha']$ must in fact be equal to S,[6] and thus L is equal to $K((T))(\alpha') = K((T^{1/n}))$. ☐

[5] That is, the quotient field of $K[[T]]$.
[6] **??**; a citation here is needed.

III.9. Regularity, differentials, and smoothness

In this chapter, we shall introduce two notions. First, we shall discuss *regular* local rings. On varieties over an algebraically closed field, regularity corresponds to nonsingularity of the variety at that point. (Over non-algebraically closed fields, the connection is more subtle.) This will be a continuation of the local algebra done earlier in the chapter chapter III.7 on dimension theory.

We shall next introduce the module of *Kähler differentials* of a morphism of rings $A \to B$, which itself can measure smoothness (though this connection will not be fully elucidated until a later chapter). The module of Kähler differentials is the algebraic analog of the *cotangent bundle* to a manifold, and we will show that for an affine ring, it can be computed very explicitly. For a smooth variety, we will see that this module is *projective*, and hence a good candidate of a vector bundle.

Despite the title, we shall actually wait a few chapters before introducing the general theory of smooth morphisms.

9.1. Regular local rings

We shall start by introducing the concept of a *regular local* ring, which is one where the embedding dimension and Krull dimension coincide.

Regular local rings

Let A be a local noetherian ring with maximal ideal $\mathfrak{m} \subset A$ and residue field $k = A/\mathfrak{m}$. Endow A with the $\mathfrak{m}$-adic topology, so that there is a natural graded k-algebra $\mathrm{gr}(A) = \bigoplus \mathfrak{m}^i/\mathfrak{m}^{i+1}$. This is a finitely generated k-algebra; indeed, a system of generators for the ideal $\mathfrak{m}$ (considered as elements of $\mathfrak{m}\mathfrak{m}^2$) generates $\mathrm{gr}(A)$ over k. As a result, we have a natural surjective map of *graded* k-algebras.

$$\mathsf{S}_k \mathfrak{m}/\mathfrak{m}^2 \to \mathrm{gr}(A). \tag{9.1.1}$$

Here S denotes the *symmetric algebra.*

9.1.1 Definition The local ring $(A, \mathfrak{m})$ is called **regular** if the above map is an isomorphism, or equivalently if the embedding dimension of A is equal to the Krull dimension.

We want to show the "equivalently" in the definition is justified. One direction is easy: if (9.1.1) is an isomorphism, then $\mathrm{gr}(A)$ is a polynomial ring with $\dim_k \mathfrak{m}/\mathfrak{m}^2$ generators. But the dimension of A was defined in terms of the growth of $\dim_k \mathfrak{m}^i/\mathfrak{m}^{i+1} = (\mathrm{gr}\, A)_i$. For a polynomial ring on r generators, however, the ith graded piece has dimension a degree-r polynomial in i (easy verification). As a result, we get the claim in one direction.

However, we still have to show that if the embedding dimension equals the Krull dimension, then (9.1.1) is an isomorphism. This will follow from the next lemma.

9.1.2 Lemma *If the inequality*

$$\dim(A) \leqslant \dim_k(\mathfrak{m}/\mathfrak{m}^2)$$

is an equality, then (9.1.1) *is an isomorphism.*

Proof. Suppose (9.1.1) is not an isomorphism. Then there is an element $f \in \mathsf{S}_k\mathfrak{m}/\mathfrak{m}^2$ which is not zero and which maps to zero in $\mathrm{gr}(A)$; we can assume f homogeneous, since the map of graded rings is graded.

Now the claim is that if $k[x_1, \dots, x_n]$ is a polynomial ring and $f \neq 0$ a homogeneous element, then the Hilbert polynomial of $k[x_1, \dots, x_n]/(f)$ is of degree less than n. This will easily imply the lemma, since (9.1.1) is always a surjection, and because $\mathsf{S}_k\mathfrak{m}/\mathfrak{m}^2$'s Hilbert polynomial is of degree $\dim_k \mathfrak{m}/\mathfrak{m}^2$. Now if $\deg f = d$, then the dimension of $(k[x_1, \dots, x_n]/f)_i$ (where i is a degree) is $\dim(k[x_1, \dots, x_n])_i = \dim(k[x_1, \dots, x_n])_{i-d}$. It follows that if P is the Hilbert polynomial of the polynomial ring, that of the quotient is $P(\cdot) - P(\cdot - d)$, which has a strictly smaller degree. ∎

We now would like to establish a few properties of regular local rings.

Let A be a local ring and $\hat{A}$ its completion. Then $\dim(A) = \dim(\hat{A})$, because $A/\mathfrak{m}^n = \hat{A}/\hat{\mathfrak{m}}^n$, so the Hilbert functions are the same. Similarly, $\mathrm{gr}(A) = \mathrm{gr}(\hat{A})$. However, by $\hat{A}$ is also a local ring. So applying the above lemma, we see:

9.1.3 Proposition *A noetherian local ring A is regular if and only if its completion $\hat{A}$ is regular.*

Regular local rings are well-behaved. We are eventually going to show that any regular local ring is in fact a unique factorization domain. Right now, we start with a much simpler claim:

9.1.4 Proposition *A regular local ring is a domain.*

This is a formal consequence of the fact that $\mathrm{gr}(A)$ is a domain and the filtration on A is Hausdorff.

Proof. Let $a, b \neq 0$. Note that $\bigcap \mathfrak{m}^n = 0$ by the Krull intersection theorem (theorem 3.3.4), so there are k_1 and k_2 such that $a \in \mathfrak{m}^{k_1} - \mathfrak{m}^{k_1+1}$ and $b \in \mathfrak{m}^{k_2} - \mathfrak{m}^{k_2+1}$. Let $\overline{a}, \overline{b}$ be the images of a, b in $\mathrm{gr}(A)$ (in degrees k_1, k_2); neither is zero. But then $\bar{a}\bar{b} \neq 0 \in \mathrm{gr}(A)$, because $\mathrm{gr}(A) = \mathsf{S}(\mathfrak{m}/\mathfrak{m}^2)$ is a domain. So $ab \neq 0$, as desired. ∎

9.1.5 Remark (exercise) Prove more generally that if A is a filtered ring with a descending filtration of ideals $I_1 \supset I_2 \supset \dots$ such that $\bigcap I_k = 0$, and such that the associated graded algebra $\mathrm{gr}(A)$ is a domain, then A is itself a domain.

Later we will prove the aforementioned fact that a regular local ring is a factorial ring. One consequence of that will be the following algebro-geometric fact. Let $X = \operatorname{Spec} \mathbb{C}[X_1, \dots, X_n]/I$ for some ideal I; so X is basically a subset of $\mathbb{C}^n$ plus some nonclosed points. Then if X is smooth, we find that $\mathbb{C}[X_1, \dots, X_n]/I$ is locally factorial. Indeed, smoothness implies regularity, hence local factoriality. The whole apparatus of Weil and Cartier divisors now kicks in.

9.1.6 Remark (exercise) Nevertheless, it is possible to prove directly that a regular local ring $(A, \mathfrak{m})$ is *integrally closed.* To do this, we shall use the fact that the associated graded $\mathrm{gr}(A)$ is integrally closed (as a polynomial ring). Here is the argument:

1. Let C be a noetherian domain with quotient field K. Then C is integrally closed if and only if for every $x \in K$ such that there exists $d \in A$ with $dx^n \in A$ for all n, we have $x \in A$. (In general, this fails for C non-noetherian; then this condition is called being *completely integrally closed.*)
2. Let C be a noetherian domain. Suppose on C there is an exhaustive filtration $\{C_v\}$ (i.e. such that $\bigcap C_v = 0$) and such that $\mathrm{gr}(C)$ is a *completely* integrally closed domain. Suppose further that every principal ideal is closed in the topology on C (i.e., for each principal ideal I, we have $I = \bigcap I + C_v$.) Then C is integrally closed too. Indeed:
   a) Suppose $b/a, a, b \in C$ is such that $(b/a)^n$ is contained in a finitely generated submodule of K, say $d^{-1}A$ for some $d \in A$. We need to show that $b \in Ca + C_v$ for all v. Write $b = xa + r$ for $r \in C_w - C_{w+1}$. We will show that "w" can be improved to $w + 1$ (by changing x). To do this, it suffices to write $r \in Ca + C_{w+1}$.
   b) By hypothesis, $db^n \in Ca^n$ for all n. Consequently $dr^n \in Ca^n$ for all n.
   c) Let $\overline{r}$ be the image of r in $\mathrm{gr}(C)$ (in some possibly positive homogeneous degree; choose the unique one such that the image of r is defined and not zero). Choosing $\overline{d}, \overline{a}$ similarly, we get $\overline{d}\overline{r}^n$ lies in the ideal of $\overline{a}^n$ for all n. This implies $\overline{r}$ is a multiple of $\overline{a}$. Deduce that $r \in Ca + C_{w+1}$.
3. The hypotheses of the previous part apply to a regular local ring, which is thus integrally closed.

The essential part of this argument is explained in **?**, ch. 5, §1.4. The application to regular local rings is mentioned in **?**, vol. IV, §16.

We now give a couple of easy examples. More interesting examples will come in the future. Let R be a noetherian local ring with maximal ideal $\mathfrak{m}$ and residue field k.

9.1.7 Example If $\dim(R) = 0$, i.e. R is artinian, then R is regular iff the maximal ideal is zero, i.e. if R is a field. Indeed, the requirement for regularity is that $\dim_k \mathfrak{m}/\mathfrak{m}^2 = 0$, or $\mathfrak{m} = 0$ (by Nakayama). This implies that R is a field.

Recall that $\dim_k \mathfrak{m}/\mathfrak{m}^2$ is the size of the minimal set of generators of the ideal $\mathfrak{m}$, by Nakayama's lemma. As a result, a local ring is regular if and only if the maximal ideal has a set of generators of the appropriate size. This is a refinement of the above remarks.

9.1.8 Example If $\dim(R) = 1$, then R is regular iff the maximal ideal $\mathfrak{m}$ is principal (by the preceding observation). The claim is that this happens if and only if R is a DVR. Certainly a DVR is regular, so only the other direction is interesting. But it is easy to see that a local domain whose maximal ideal is principal is a DVR (i.e. define the valuation of x in terms of the minimal i such that $x \notin \mathfrak{m}^i$).

We find:

9.1.9 Proposition *A one-dimensional regular local ring is the same thing as a DVR.*

Finally, we extend the notion to general noetherian rings:

9.1.10 Definition A general noetherian ring is called **regular** if every localization at a maximal ideal is a regular local ring.

In fact, it turns out that if a noetherian ring is regular, then so are *all* its localizations. This fact relies on a fact, to be proved in the distant future, that the localization of a regular local ring at a prime ideal is regular.

Quotients of regular local rings

We now study quotients of regular local rings. In general, if $(A, \mathfrak{m})$ is a regular local ring and $f_1, \ldots, f_k \in \mathfrak{m}$, the quotient $A/(f_1, \ldots, f_k)$ is far from being regular. For instance, if k is a field and A is $k[x]_{(x)}$ (geometrically, this is the local ring of the affine line at the origin), then $A/x^2 = k[\epsilon]/\epsilon^2$ is not a regular local ring; it is not even a domain. In fact, the local ring of *any* variety at a point is a *quotient* of a regular local ring, and this is because any variety locally sits inside affine space.[1]

9.1.11 Proposition *If $(A, \mathfrak{m}_A)$ is a regular local ring, and $f \in \mathfrak{m}$ is such that $f \in \mathfrak{m}_A - \mathfrak{m}_A^2$. Then $A' = A/fA$ is also regular of dimension* $\dim(A) - 1$.

Proof. First let us show the dimension part of the statement. We know from proposition 7.2.2 that the dimension has to drop precisely by one (since f is a non-zero-divisor on A by proposition 9.1.4).

Now we want to show that $A' = A/fA$ is regular. Let $\mathfrak{m}_{A'} = \mathfrak{m}/fA$ be the maximal ideal of A'. Then we should show that $\dim_{A'/\mathfrak{m}_{A'}}(\mathfrak{m}_{A'}/\mathfrak{m}_{A'}^2) = \dim(A')$, and it suffices to see that

$$\dim_{A'/\mathfrak{m}_{A'}}(\mathfrak{m}_{A'}/\mathfrak{m}_{A'}^2) \leqslant \dim_{A/\mathfrak{m}_A}(\mathfrak{m}_A/\mathfrak{m}_A^2) - 1. \tag{9.1.2}$$

In other words, we have to show that the embedding dimension drops by one.

Note that the residue fields $k = A/\mathfrak{m}_A, A'/\mathfrak{m}_{A'}$ are naturally isomorphic. To see (9.1.2), we use the natural surjection of k-vector spaces

$$\mathfrak{m}_A/\mathfrak{m}_A^2 \to \mathfrak{m}_{A'}/\mathfrak{m}_{A'}^2.$$

Since there is a nontrivial kernel (the class of f is in the kernel), we obtain the inequality (9.1.2). □

[1] Incidentally, the condition that a noetherian local ring $(A, \mathfrak{m})$ is a quotient of a regular local ring $(B, \mathfrak{n})$ imposes conditions on A: for instance, it has to be *catenary.* As a result, one can obtain examples of local rings which cannot be expressed as quotients in this way.

9.1.12 Corollary *Consider elements $f_1, \dots f_m$ in $\mathfrak{m}$ such that $\bar{f_1}, \dots \bar{f_m} \in \mathfrak{m}/\mathfrak{m}^2$ are linearly independent. Then $A/(f_1, \dots f_m)$ is regular with* $\dim(A/(f_1, \dots f_m)) = \dim(A) - m$

Proof. This follows from proposition 9.1.11 by induction. One just needs to check that in $A_1 = A/(f_1)$, $\mathfrak{m}_1 = \mathfrak{m}/(f_1)$, we have that $f_2, \dots f_m$ are still linearly independent in $\mathfrak{m}_1/\mathfrak{m}_1^2$. This is easy to check. ☐

9.1.13 Remark In fact, note in the above result that each f_i is a *non-zero-divisor* on $A/(f_1, \dots, f_{i-1})$, because a regular local ring is a domain. We will later say that the $\{f_i\}$ form a *regular sequence.*

We can now obtain a full characterization of when a quotient of a regular local ring is still regular; it essentially states that the above situation is the only possible case. Geometrically, the intuition is that we are analyzing when a subvariety of a smooth variety is smooth; the answer is when the subvariety is cut out by functions with linearly independent images in the maximal ideal mod its square.

This corresponds to the following fact: if M is a smooth manifold and $f_1, \dots, f_m$ smooth functions such that the gradients $\{df_i\}$ are everywhere independent, then the common zero locus of the $\{f_i\}$ is a smooth submanifold of M, and conversely every smooth submanifold of M locally looks like that.

9.1.14 Theorem *Let A_0 be a regular local ring of dimension n, and let $I \subset A_0$ be a proper ideal. Let $A = A_0/I$. Then the following are equivalent*

1. *A is regular.*
2. *There are elements $f_1, \dots f_m \in I$ such that $\bar{f_1}, \dots \bar{f_m}$ are linearly independent in $\mathfrak{m}_{A_0}/\mathfrak{m}_{A_0}^2$ where $m = n - \dim(A)$ such that $(f_1, \dots f_m) = I$.*

Proof. **(2) ⇒ (1)** This is exactly the statement of corollary 9.1.12.

(1) ⇒ (2) Let k be the residue field of A (or A_0, since I is contained in the maximal ideal). We see that there is an exact sequence

$$I \otimes_{A_0} k \to \mathfrak{m}_{A_0}/\mathfrak{m}_{A_0}^2 \to \mathfrak{m}_A/\mathfrak{m}_A^2 \to 0.$$

We can obtain this from the exact sequence $I \to A_0 \to A \to 0$ by tensoring with k.

By assumption A_0 and A are regular local, so

$$\dim_{A_0/\mathfrak{m}_{A_0}}(\mathfrak{m}_{A_0}/\mathfrak{m}_{A_0}^2) = \dim(A_0) = n$$

and

$$\dim_{A_0/\mathfrak{m}_{A_0}}(\mathfrak{m}_A/\mathfrak{m}_A^2) = \dim(A)$$

so the image of $I \otimes_{A_0} k$ in $\mathfrak{m}_{A_0}/\mathfrak{m}_{A_0}^2$ has dimension $m = n - \dim(A)$. Let $\bar{f_1}, \dots \bar{f_m}$ be a set of linearly independent generators of the image of $I \otimes_{A_0} k$ in $\mathfrak{m}_{A_0}/\mathfrak{m}_{A_0}^2$, and let $f_1, \dots f_m$ be liftings to I. The claim is that the $\{f_i\}$ generate I.

Let $I' \subset A_0$ be the ideal generated by $f_1, \dots f_m$ and consider $A' = A_0/I'$. Then by corollary 9.1.12, we know that A' is a regular local ring with dimension $n - m = \dim(A)$. Also $I' \subset I$ so we have an exact sequence

$$0 \to I/I' \to A' \to A \to 0$$

But proposition 9.1.4, this means that A' is a domain, and we have just seen that it has the same dimension as A. Now if $I/I' \neq 0$, then A would be a proper quotient of A', and hence of a *smaller* dimension (because quotienting by a non-zero-divisor drops the dimension). This contradiction shows that $I = I'$, which means that I is generated by the sequence $\{f_i\}$ as claimed. ☐

So the reason that $k[x]_{(x)}/(x^2)$ was not regular is that x^2 vanishes to too high an order: it lies in the square of the maximal ideal.

We can motivate the results above further with:

9.1.15 Definition In a regular local ring $(R, \mathfrak{m})$, a **regular system of parameters** is a minimal system of generators for $\mathfrak{m}$, i.e. elements of $\mathfrak{m}$ that project to a basis of $\mathfrak{m}/\mathfrak{m}^2$.

So a quotient of a regular local ring is regular if and only if the ideal is generated by a portion of a system of parameters.

Regularity and smoothness

We now want to connect the intuition (described in the past) that, in the algebro-geometric context, regularity of a local ring corresponds to smoothness of the associated variety (at that point).

Namely, let R be be the (reduced) coordinate ring $\mathbb{C}[x_1, \dots, x_n]/I$ of an algebraic variety. Let $\mathfrak{m}$ be a maximal ideal corresponding to the origin, so generated by $(x_1, \dots, x_n)$. Suppose $I \subset \mathfrak{m}$, which is to say the origin belongs to the corresponding variety. Then $\mathrm{MaxSpec} R \subset \operatorname{Spec} R$ is the corresponding subvariety of $\mathbb{C}^n$, which is what we apply the intuition to. Note that 0 is in this subvariety.

Then we claim:

9.1.16 Proposition *$R_{\mathfrak{m}}$ is regular iff $\mathrm{MaxSpec} R$ is a smooth submanifold near 0.*

Proof. We will show that regularity implies smoothness. The other direction is omitted for now.

Note that $S = \mathbb{C}[x_1, \dots, x_n]_{\mathfrak{m}}$ is clearly a regular local ring of dimension n ($\mathbb{C}^n$ is smooth, intuitively), and $R_{\mathfrak{m}}$ is the quotient S/I. By theorem 9.1.14, we have a good criterion for when $R_{\mathfrak{m}}$ is regular. Namely, it is regular if and only if I is generated by elements (without loss of generality, polynomials) $f_1, \dots, f_k$ whose images in the quotient $\mathfrak{m}_S/\mathfrak{m}_S^2$ (where we write $\mathfrak{m}_S$ to emphasize that this is the maximal ideal of S).

But we know that this "cotangent space" corresponds to cotangent vectors in $\mathbb{C}^n$, and in particular, we can say the following. There are elements $\epsilon_1, \dots, \epsilon_n \in \mathfrak{m}_S/\mathfrak{m}_S^2$ that form a basis for this space (namely, the images of $x_1, \dots, x_n \in \mathfrak{m}_S$). If f is a polynomial vanishing at the origin, then the image of f in $\mathfrak{m}_S/\mathfrak{m}_S^2$ takes only the linear terms—that is, it can be identified with

$$\sum \frac{\partial f}{\partial x_i}|_0 \epsilon_i,$$

which is essentially the gradient of f.

It follows that $R_\mathfrak{m}$ is regular if and only if I is generated (in $R_\mathfrak{m}$, so we should really say $IR_\mathfrak{m}$) by a family of polynomials vanishing at zero with linearly independent gradients, or if the variety is cut out by the vanishing of such a family of polynomials. However, we know that this implies that the variety is locally a smooth manifold (by the inverse function theorem). ☐

The other direction is a bit trickier, and will require a bit of "descent." For now, we omit it. But we have shown *something* in both directions: the ring $R_\mathfrak{m}$ is regular if and only if I is generated locally (i.e., in $R_\mathfrak{m}$ by a family of polynomials with linearly independent gradients). Hartshorne uses this as the definition of smoothness in **?**, and thus obtains the result that a variety over an algebraically closed field (not necessarily $\mathbb{C}$!) is smooth if and only if its local rings are regular.

9.1.17 Remark (Warning) This argument proves that if $R \simeq K[x_1, \dots, x_n]/I$ for K algebraically closed, then $R_\mathfrak{m}$ is regular local for some maximal ideal $\mathfrak{m}$ if the corresponding algebraic variety is smooth at the corresponding point. We proved this in the special case $K = \mathbb{C}$ and $\mathfrak{m}$ the ideal of the origin.

If K is not algebraically closed, we **can't assume** that any maximal ideal corresponds to a point in the usual sense. Moreover, if K is not perfect, regularity does **not** imply smoothness. We have not quite defined smoothness, but here's a definition: smoothness means that the local ring you get by base-changing K to the algebraic closure is regular. So what this means is that regularity of affine rings over a field K is not preserved under base-change from K to $\overline{K}$.

9.1.18 Example Let K be non-perfect of characteristic p. Let a not have a pth root. Consider $K[x]/(x^p - a)$. This is a regular local ring of dimension zero, i.e. is a field. If K is replaced by its algebraic closure, then we get $\overline{K}[x]/(x^p - a)$, which is $\overline{K}[x]/(x - a^{1/p})^p$. This is still zero-dimensional but is not a field. Over the algebraic closure, the ring fails to be regular.

Regular local rings look alike

So, as we've seen, regularity corresponds to smoothness. Complex analytically, all smooth points are the same though—they're locally $\mathbb{C}^n$. Manifolds have no local invariants. We'd like an algebraic version of this. The vague claim is that all regular local rings of the same dimension "look alike." We have already seen one instance of this phenomenon: a regular local ring's associated graded is uniquely determined by its dimension (as a polynomial ring). This was in fact how we defined the notion, in part. Now we would like to transfer this to statements about things closer to R.

Let $(R, \mathfrak{m})$ be a regular local ring. **Assume now for simplicity that the residue field of $k = R/\mathfrak{m}$ maps back into R.** In other words, R contains a copy of its residue field, or there is a section of $R \to k$. This is always true in the case we use for geometric intuition—complex algebraic geometry—as the residue field at any maximal ideal is just $\mathbb{C}$ (by the Nullstellensatz), and one works with $\mathbb{C}$-algebras.

Choose generators $y_1, \dots, y_n \in \mathfrak{m}$ where $n = \dim_k \mathfrak{m}/\mathfrak{m}^2$ is the embedding dimension. We get a map in the other direction

$$\phi : k[Y_1, \dots, Y_n] \to R, \quad Y_i \mapsto y_i,$$

thanks to the section $k \to R$. This map from the polynomial ring is not an isomorphism (the polynomial ring is not local), but if we let $\mathfrak{m} \subset R$ be the maximal ideal and $\mathfrak{n} = (y_1, \dots, y_n)$, then the map on associated gradeds is an isomorphism (by definition). That is, $\phi : \mathfrak{n}^t/\mathfrak{n}^{t+1} \to \mathfrak{m}^t/\mathfrak{m}^{t+1}$ is an isomorphism for each $t \in \mathbb{Z}_{\geqslant 0}$.

Consequently, ϕ induces an isomorphism

$$k[Y_1, \dots, Y_n]/\mathfrak{n}^t \simeq R/\mathfrak{m}^t$$

for all t, because it is an isomorphism on the associated graded level. So this in turn is equivalent, upon taking inverse limits, to the statement that ϕ induces an isomorphism

$$k[[Y_1, \dots, Y_n]] \to \hat{R}$$

at the level of completions.

We can now conclude:

9.1.19 Theorem *Let R be a regular local ring of dimension n. Suppose R contains a copy of its residue field k. Then, as k-algebras, $\hat{R} \simeq k[[Y_1, \dots, Y_m]]$.*

Finally:

9.1.20 Corollary *A complete noetherian regular local ring that contains a copy of its residue field k is a power series ring over k.*

It now makes sense to say:

> **All *complete* regular local rings of the same dimension look alike.** (More precisely, this is true when R is assumed to contain a copy of its residue field, but this is not a strong assumption in practice. One can show that this will be satisfied if R contains *any* field.[2])

We won't get into the precise statement of the general structure theorem, when the ring is not assumed to contain its residue field, but a safe intuition to take away from this is the above bolded statement. Note that "looking alike" requires the completeness, because completions are intuitively like taking analytically local invariants (while localization corresponds to working *Zariski* locally, which is much weaker).

[2] This is not always satisfied—take the p-adic integers, for instance.

9.2. Kähler differentials

Derivations and Kähler differentials

Let R be a ring with the maximal ideal $\mathfrak{m}$. Then there is a $R/\mathfrak{m}$-vector space $\mathfrak{m}/\mathfrak{m}^2$. This is what we would like to think of as the "cotangent space" of $\operatorname{Spec} R$ at $\mathfrak{m}$. Intuitively, the cotangent space is what you get by differentiating functions which vanish at the point, but differentiating functions that vanish twice should give zero. This is the moral justification. (Recall that on a smooth manifold M, if $\mathcal{O}_p$ is the local ring of smooth functions defined in a neighborhood of $p \in M$, and $\mathfrak{m}_p \subset \mathcal{O}_p$ is the maximal ideal consisting of "germs" vanishing at p, then the cotangent space T_p^*M is naturally $\mathfrak{m}_p/\mathfrak{m}_p^2$.)

A goal might be to generalize this. What if you wanted to think about all points at once? We'd like to describe the "cotangent bundle" to $\operatorname{Spec} R$ in an analogous way. Let's try and describe what would be a section to this cotangent bundle. A section of $\Omega^*_{\operatorname{Spec} R}$ should be the same thing as a "1-form" on $\operatorname{Spec} R$. We don't know what a 1-form is yet, but at least we can give some examples. If $f \in R$, then f is a "function" on $\operatorname{Spec} R$, and its "differential" should be a 1-form. So there should be a "df" which should be a 1-form. This is analogous to the fact that if g is a real-valued function on the smooth manifold M, then there is a 1-form dg.

We should expect the rules $d(fg) = df + dg$ and $d(fg) = f(dg) + g(df)$ as the usual rules of differentiation. For this to make sense, 1-forms should be an R-module. Before defining the appropriate object, we start with:

9.2.1 Definition Let R be a commutative ring, M an R-module. A **derivation** from R to M is a map $D : R \to M$ such that the two identities below hold:

$$D(fg) = Df + Dg \tag{9.2.1}$$
$$D(fg) = f(Dg) + g(Df). \tag{9.2.2}$$

These equations make sense as M is an R-module.

Whatever a 1-form on $\operatorname{Spec} R$ might be, there should be a derivation

$$d : R \to \{1\text{–forms}\}.$$

An idea would be to *define* the 1-forms or the "cotangent bundle" Ω_R by a universal property. It should be universal among R-modules with a derivation.

To make this precise:

9.2.2 Proposition *There is an R-module Ω_R and a derivation $d_{\text{univ}} : R \to \Omega_R$ satisfying the following universal property. For all R-modules M, there is a canonical isomorphism*

$$\hom_R(\Omega_R, M) \simeq \operatorname{Der}(R, M)$$

given by composing the universal d_{univ} with a map $\Omega_R \to M$.

That is, any derivation $d : R \to M$ factors through this universal derivation in a unique way. Given the derivation $d : R \to M$, we can make the following diagram commutative in a unique way such that $\Omega_R \to M$ is a morphism of R-modules:

$$\begin{tikzcd} R \arrow[r, "d"] \arrow[d, "d_{\mathrm{univ}}"] & M \\ \Omega_R \arrow[ur] & \end{tikzcd}$$

9.2.3 Definition Ω_R is called the module of **Kähler differentials** of R.

Let us now verify this proposition.

Proof. This is like the verification of the tensor product. Namely, build a free gadget and quotient out to enforce the desired relations.

Let Ω_R be the quotient of the free R-module generated by elements da for $a \in R$ by enforcing the relations

1. $d(a + b) = da + db$.
2. $d(ab) = adb + bda$.

By construction, the map $a \to da$ is a derivation $R \to \Omega_R$. It is easy to see that it is universal. Given a derivation $d' : R \to M$, we get a map $\Omega_R \to M$ sending $da \to d'(a), a \in R$. $\square$

The philosophy of Grothendieck says that we should do this, as with everything, in a relative context. Indeed, we are going to need a slight variant, for the case of a *morphism* of rings.

Relative differentials

On a smooth manifold M, the derivation d from smooth functions to 1-forms satisfies an additional property: it maps the constant functions to zero. This is the motivation for the next definition:

9.2.4 Definition Let $f : R \to R'$ be a ring-homomorphism. Let M be an R'-module. A derivation $d : R' \to M$ is **R-linear if $d(f(a)) = 0, a \in R$.** This is equivalent to saying that d is an R-homomorphism by the Leibniz rule.

Now we want to construct an analog of the "cotangent bundle" taking into account linearity.

9.2.5 Proposition *Let R' be an R-algebra. Then there is a universal R-linear derivation $R' \overset{d_{\mathrm{univ}}}{\to} \Omega_{R'/R}$.*

Proof. Use the same construction as in the absolute case. We get a map $R' \to \Omega_{R'}$ as before. This is not generally R-linear, so one has to quotient out by the images of $d(f(r)), r \in R$. In other words, $\Omega_{R'/R}$ is the quotient of the free R'-module on symbols $\{dr', r' \in R'\}$ with the relations:

1. $d(r'_1 r'_2) = r'_1 d(r'_2) + d(r'_1) r'_2$.

2. $d(r'_1 + r'_2) = dr'_1 + dr'_2$.
3. $dr = 0$ for $r \in R$ (where we identify r with its image $f(r)$ in R', by abuse of notation). $\square$

9.2.6 Definition $\Omega_{R'/R}$ is called the module of **relative Kähler differentials,** or simply Kähler differentials.

Here $\Omega_{R'/R}$ also corepresents a simple functor on the category of R'-modules: given an R'-module M, we have

$$\hom_{R'}(\Omega_{R'/R}, M) = \mathrm{Der}_R(R', M),$$

where Der_R denotes R-derivations. This is a *subfunctor* of the functor $\mathrm{Der}_R(R', \cdot)$, and so by Yoneda's lemma there is a natural map $\Omega_{R'} \to \Omega_{R'/R}$. We shall expand on this in the future.

The case of a polynomial ring

Let us do a simple example to make this more concrete.

9.2.7 Example Let $R' = \mathbb{C}[x_1, \ldots, x_n], R = \mathbb{C}$. In this case, the claim is that there is an isomorphism

$$\Omega_{R'/R} \simeq R'^n.$$

More precisely, $\Omega_{R'/R}$ is free on $dx_1, \ldots, dx_n$. So the cotangent bundle is "free." In general, the module $\Omega_{R'/R}$ will not be free, or even projective, so the intuition that it is a vector bundle will be rather loose. (The projectivity will be connected to *smoothness* of R'/R.)

Proof. The construction $f \to \left(\frac{\partial f}{\partial x_i}\right)$ gives a map $R' \to R'^n$. By elementary calculus, this is a derivation, even an R-linear derivation. We get a map

$$\phi : \Omega_{R'/R} \to R'^n$$

by the universal property of the Kähler differentials. The claim is that this map is an isomorphism. The map is characterized by sending *df* to $\left(\frac{\partial f}{\partial x_i}\right)$. Note that $dx_1, \ldots, dx_n$ map to a basis of R'^n because differentiating x_i gives 1 at i and zero at $j \neq i$. So we see that ϕ is surjective.

There is a map $\psi : R'^n \to \Omega_{R'/R}$ sending (a_i) to $\sum a_i dx_i$. It is easy to check that $\phi \circ \psi = 1$ from the definition of ϕ. What we still need to show is that $\psi \circ \phi = 1$. Namely, for any f, we want to show that $\psi \circ \phi(df) = df$ for $f \in R'$. This is precisely the claim that $df = \sum \frac{\partial f}{\partial x_i} dx_i$. Both sides are additive in f, indeed are derivations, and coincide on monomials of degree one, so they are equal. $\square$

By the same reasoning, one can show more generally:

9.2.8 Proposition *If R is any ring, then there is a canonical isomorphism*

$$\Omega_{R[x_1,\ldots,x_n]/R} \simeq \bigoplus_{i=1}^{n} R[x_1, \ldots, x_n] dx_i,$$

i.e. it is $R[x_1, \ldots, x_n]$-free on the dx_i.

This is essentially the claim that, given an $R[x_1, \ldots, x_n]$-module M and elements $m_1, \ldots, m_n \in M$, there is a *unique* R-derivation from the polynomial ring into M sending $x_i \mapsto m_i$.

Exact sequences of Kähler differentials

We now want to prove a few basic properties of Kähler differentials, which can be seen either from the explicit construction or in terms of the functors they represent, by formal nonsense. These results will be useful in computation.

Recall that if $\phi : A \to B$ is a map of rings, we can define a B-module $\Omega_{B/A}$ which is generated by formal symbols $dx|_{x \in B}$ and subject to the relations $d(x+y) = dx + dy$, $d(a) = 0, a \in A$, and $d(xy) = xdy + ydx$. By construction, $\Omega_{B/A}$ is the receptacle from the universal A-linear derivation into a B-module.

Let $A \to B \to C$ be a triple of maps of rings. There is an obvious map $dx \to dx$

$$\Omega_{C/A} \to \Omega_{C/B}$$

where both sides have the same generators, except with a few additional relations on $\Omega_{C/B}$. We have to quotient by $db, b \in B$. In particular, there is a map $\Omega_{B/A} \to \Omega_{C/A}$, $dx \to dx$, whose images generate the kernel. This induces a map

$$C \otimes_B \Omega_{B/A} \to \Omega_{C/A}.$$

The image is the C-module generated by $db|_{b \in B}$, and this is the kernel of the previous map. We have proved:

9.2.9 Proposition (First exact sequence) *Given a sequence $A \to B \to C$ of rings, there is an exact sequence*

$$C \otimes_B \Omega_{B/A} \to \Omega_{C/A} \to \Omega_{C/B} \to 0.$$

Second proof. There is, however, a more functorial means of seeing this sequence, which we now describe. Namely, let us consider the category of C-modules, and the functors corepresented by these three objects. We have, for a C-module M:

$$\begin{aligned} \hom_C(\Omega_{C/B}, M) &= \mathrm{Der}_B(C, M) \\ \hom_C(\Omega_{C/A}, M) &= \mathrm{Der}_A(C, M) \\ \hom_C(C \otimes_B \Omega_{B/A}, M) = \hom_B(\Omega_{B/A}, M) &= \mathrm{Der}_A(B, M). \end{aligned}$$

By Yoneda's lemma, we know that a map of modules is the same thing as a natural transformation between the corresponding corepresentable functors, in the reverse direction. It is easy to see that there are natural transformations

$$\mathrm{Der}_B(C, M) \to \mathrm{Der}_A(C, M), \quad \mathrm{Der}_A(C, M) \to \mathrm{Der}_A(B, M)$$

obtained by restriction in the second case, and by doing nothing in the first case (a B-derivation is automatically an A-derivation). The induced maps on the modules of differentials are precisely those described before; this is easy to check (and we could have defined the maps by these functors if we wished). Now to say that the sequence is right exact is to say that for each M, there is an exact sequence of abelian groups

$$0 \to \mathrm{Der}_B(C, M) \to \mathrm{Der}_A(C, M) \to \mathrm{Der}_A(B, M).$$

But this is obvious from the definitions: an A-derivation is a B-derivation if and only if the restriction to B is trivial. This establishes the claim. ∎

Next, we are interested in a second exact sequence. In the past (example 9.2.7), we computed the module of Kähler differentials of a *polynomial* algebra. While this was a special case, any algebra is a quotient of a polynomial algebra. As a result, it will be useful to know how $\Omega_{B/A}$ behaves with respect to quotienting B.

Let $A \to B$ be a homomorphism of rings and $I \subset B$ an ideal. We would like to describe $\Omega_{B/I/A}$. There is a map

$$\Omega_{B/A} \to \Omega_{B/I/A}$$

sending dx to $d\overline{x}$ for $\overline{x}$ the reduction of x in B/I. This is surjective on generators, so it is surjective. It is not injective, though. In $\Omega_{B/I/A}$, the generators dx, dx' are identified if $x \equiv x'$ mod I. Moreover, $\Omega_{B/I/A}$ is a B/I-module. This means that there will be additional relations for that. To remedy this, we can tensor and consider the morphism

$$\Omega_{B/A} \otimes_B B/I \to \Omega_{B/I/A} \to 0.$$

Let us now define a map

$$\phi : I/I^2 \to \Omega_{B/A} \otimes_B B/I,$$

which we claim will generate the kernel. Given $x \in I$, we define $\phi(x) = dx$. If $x \in I^2$, then $dx \in I\Omega_{B/A}$ so ϕ is indeed a map of abelian groups $I/I^2 \to \Omega_{B/A} \otimes_B B/I$. Let us check that this is a B/I-module homorphism. We would like to check that $\phi(xy) = y\phi(x)$ for $x \in I$ in $\Omega_{B/A}/I\Omega_{B/A}$. This follows from the Leibnitz rule, $\phi(xy) = y\phi(x) + xdy \equiv x\phi(x) \mod I\Omega_{B/A}$. So ϕ is also defined. Its image is the submodule of $\Omega_{B/A}/I\Omega_{B/A}$ generated by $dx, x \in I$. This is precisely what one has to quotient out by to get $\Omega_{B/I/A}$. In particular:

9.2.10 Proposition (Second exact sequence) *Let B be an A-algebra and $I \subset B$ an ideal. There is an exact sequence*

$$I/I^2 \to \Omega_{B/A} \otimes_B B/I \to \Omega_{B/I/A} \to 0.$$

These results will let us compute the module of Kähler differentials in cases we want.

9.2.11 Example Let $B = A[x_1, \dots, x_n]/I$ for I an ideal. We will compute $\Omega_{B/A}$.

First, $\Omega_{A[x_1,\dots,x_n]/A} \otimes B \simeq B^n$ generated by symbols dx_i. There is a surjection of

$$B^n \to \Omega_{B/A} \to 0$$

whose kernel is generated by $dx, x \in I$, by the second exact sequence above. If $I = (f_1, \dots, f_m)$, then the kernel is generated by $\{df_i\}$. It follows that $\Omega_{B/A}$ is the cokernel of the map

$$B^m \to B^n$$

that sends the ith generator of B^m to df_i thought of as an element in the free B-module B^n on generators $dx_1, \dots, dx_n$. Here, thanks to the Leibnitz rule, df_i is given by formally differentiating the polynomial, i.e.

$$df_i = \sum_j \frac{\partial f_i}{\partial x_j} dx_j.$$

We have thus explicitly represented $\Omega_{B/A}$ as the cokernel of the matrix $\left(\frac{\partial f_i}{\partial x_j}\right)$.

In particular, the above example shows:

9.2.12 Proposition *If B is a finitely generated A-algebra, then $\Omega_{B/A}$ is a finitely generated B-module.*

Given how Ω behaves with respect to localization, we can extend this to the case where B is *essentially* of finite type over A (recall that this means B is a localization of a finitely generated A-algebra).

Let $R = \mathbb{C}[x_1, \dots, x_n]/I$ be the coordinate ring of an algebraic variety. Let $\mathfrak{m} \subset R$ be the maximal ideal. Then $\Omega_{R/\mathbb{C}}$ is what one should think of as containing information of the cotangent bundle of $\operatorname{Spec} R$. One might ask what the *fibe*r over a point $\mathfrak{m} \in \operatorname{Spec} R$ is, though. That is, we might ask what $\Omega_{R/\mathbb{C}} \otimes_R R/\mathfrak{m}$ is. To see this, we note that there are maps

$$\mathbb{C} \to R \to R/\mathfrak{m} \simeq \mathbb{C}.$$

There is now an exact sequence by proposition 9.2.9

$$\mathfrak{m}/\mathfrak{m}^2 \to \Omega_{R/\mathbb{C}} \otimes_R R/\mathfrak{m} \to \Omega_{\mathbb{R}/\mathfrak{m}/\mathbb{C}} \to 0,$$

where the last thing is zero as $R/\mathfrak{m} \simeq \mathbb{C}$ by the Nullstellensatz. The upshot is that $\Omega_{R/\mathbb{C}} \otimes_R R/\mathfrak{m}$ is a quotient of $\mathfrak{m}/\mathfrak{m}^2$.

In fact, the natural map $\mathfrak{m}/\mathfrak{m}^2 \to \Omega_{R/\mathbb{C}} \otimes_R \mathbb{C}$ (given by d) is an *isomorphism* of $\mathbb{C}$-vector spaces. We have seen that it is surjective, so we need to see that it is injective. That is, if V is a $\mathbb{C}$-vector space, then we need to show that the map

$$\hom_{\mathbb{C}}(\Omega_{R/\mathbb{C}} \otimes_R \mathbb{C}, V) \to \hom_{\mathbb{C}}(\mathfrak{m}/\mathfrak{m}^2, V)$$

is surjective. This means that given any $\mathbb{C}$-linear map $\lambda : \mathfrak{m}/\mathfrak{m}^2 \to V$, we can extend this to a derivation $R \to V$ (where V becomes an R-module by $R/\mathfrak{m} \simeq \mathbb{C}$, as usual). But this is easy: given $f \in R$, we write $f = f_0 + c$ for $c \in \mathbb{C}$ and $f_0 \in \mathfrak{m}$, and have the derivation send f to $\lambda(f_0)$. (Checking that this is a well-defined derivation is straightforward.)

This goes through if $\mathbb{C}$ is replaced by any algebraically closed field. We have found:

9.2.13 Proposition *Let $(R, \mathfrak{m})$ be the localization of a finitely generated algebra over an algebraically closed field k at a maximal ideal $\mathfrak{m}$. Then there is a natural isomorphism:*

$$\Omega_{R/k} \otimes_R k \simeq \mathfrak{m}/\mathfrak{m}^2.$$

This result connects the Kähler differentials to the cotangent bundle: the fiber of the cotangent bundle at a point in a manifold is, similarly, the maximal ideal modulo its square (where the "maximal ideal" is the maximal ideal in the ring of germs of functions at that point).

Kähler differentials and base change

We now want to show that the formation of Ω is compatible with base change. Namely, let B be an A-algebra, visualized by a morphism $A \to B$. If $A \to A'$ is any morphism of rings, we can think of the *base-change* $A' \to A' \otimes_A B$; we often write $B' = A' \otimes_A B$.

9.2.14 Proposition *With the above notation, there is a canonical isomorphism of B'-modules:*

$$\Omega_{B/A} \otimes_A A' \simeq \Omega_{B'/A'}.$$

Note that, for a B-module, the functors $\otimes_A A'$ and $\otimes_B B'$ are the same. So we could have as well written $\Omega_{B/A} \otimes_B B' \simeq \Omega_{B'/A'}$.

Proof. We will use the functorial approach. Namely, for a B'-module M, we will show that there is a canonical isomorphism

$$\hom_{B'}(\Omega_{B/A} \otimes_A A', M) \simeq \hom_{B'}(\Omega_{B'/A'}, M).$$

The right side represents A'-derivations $B' \to M$, or $\mathrm{Der}_{A'}(B', M)$. The left side represents $\hom_B(\Omega_{B/A}, M)$, or $\mathrm{Der}_A(B, M)$. Here the natural map of modules corresponds by Yoneda's lemma to the restriction

$$\mathrm{Der}_{A'}(B', M) \to \mathrm{Der}_A(B, M).$$

We need to see that this restriction map is an isomorphism. But given an A-derivation $B \to M$, this is to say that extends in a *unique* way to an A'-linear derivation $B' \to M$. This is easy to verify directly. $\square$

We next describe how Ω behaves with respect to forming tensor products.

9.2.15 Proposition *Let B, B' be A-algebras. Then there is a natural isomorphism*

$$\Omega_{B \otimes_A B'/A} \simeq \Omega_{B/A} \otimes_A B' \oplus B \otimes_A \Omega_{B'/A}.$$

Since Ω is a linearization process, it is somewhat natural that it should turn tensor products into direct sums.

Proof. The "natural map" can be described in the leftward direction. For instance, there is a natural map $\Omega_{B/A} \otimes_A B' \to \Omega_{B \otimes_A B'/A}$. We just need to show that it is an isomorphism. For this, we essentially have to show that to give an A-derivation of $B \otimes_A B'$ is the same as giving a derivation of B and one of B'. This is easy to check. $\square$

Differentials and localization

We now show that localization behaves *extremely* nicely with respect to the formation of Kähler differentials. This is important in algebraic geometry for knowing that the "cotangent bundle" can be defined locally.

9.2.16 Proposition *Let $f : A \to B$ be a map of rings. Let $S \subset B$ be multiplicatively closed. Then the natural map*

$$S^{-1}\Omega_{B/A} \to \Omega_{S^{-1}B/A}$$

is an isomorphism.

So the formation of Kähler differentials commutes with localization.

Proof. We could prove this by the calculational definition, but perhaps it is better to prove it via the universal property. If M is any $S^{-1}B$-module, then we can look at

$$\hom_{S^{-1}B}(\Omega_{S^{-1}B/A}, M)$$

which is given by the group of A-linear derivations $S^{-1}B \to M$, by the universal property.

On the other hand,

$$\hom_{S^{-1}B}(S^{-1}\Omega_{B/A}, M)$$

is the same thing as the set of B-linear maps $\Omega_{B/A} \to M$, i.e. the set of A-linear derivations $B \to M$.

We want to show that these two are the same thing. Given an A-derivation $S^{-1}B \to M$, we get an A-derivation $B \to M$ by pulling back. We want to show that any A-linear derivation $B \to M$ arises in this way. So we need to show that any A-linear derivation $d : B \to M$ extends uniquely to an A-linear $\overline{d} : S^{-1}B \to M$. Here are two proofs:

1. (Lowbrow proof.) For $x/s \in S^{-1}B$, with $x \in B, s \in S$, we define $\overline{d}(x/s) = dx/s - xds/s^2$ as in calculus. The claim is that this works, and is the only thing that works. One should check this—**remark**.

2. (Highbrow proof.) We start with a digression. Let B be a commutative ring, M a B-module. Consider $B \oplus M$, which is a B-module. We can make it into a ring (via **square zero multiplication**) by multiplying

   $$(b, x)(b', x') = (bb', bx' + b'x).$$

   This is compatible with the B-module structure on $M \subset B \oplus M$. Note that M is an ideal in this ring with square zero. Then the projection $\pi : B \oplus M \to B$ is a ring-homomorphism as well. There is also a ring-homomorphism in the other direction $b \to (b, 0)$, which is a section of π. There may be other homomorphisms $B \to B \oplus M$.

   You might ask what all the right inverses to π are, i.e. ring-homomorphisms $\phi : B \to B \oplus M$ such that $\pi \circ \phi = 1_B$. This must be of the form $\phi : b \to (b, db)$ where $d : B \to M$ is some map. It is easy to check that ϕ is a homomorphism precisely when d is a derivation.

   Suppose now $A \to B$ is a morphism of rings making B an A-algebra. Then $B \oplus M$ is an A-algebra via the inclusion $a \to (a, 0)$. Then you might ask when $\phi : b \to (b, db), B \to B \oplus M$ is an A-homomorphism. The answer is clear: when d is an A-derivation.

Recall that we were in the situation of $f : A \to B$ a morphism of rings, $S \subset B$ a multiplicatively closed subset, and M an $S^{-1}B$-module. The claim was that any A-linear derivation $d : B \to M$ extends uniquely to $\overline{d} : S^{-1}B \to M$. We can draw a diagram

$$\begin{tikzcd} & B \oplus M \arrow[r] \arrow[d] & S^{-1}B \oplus M \arrow[d] \\ A \arrow[r] & B \arrow[r] & S^{-1}B \end{tikzcd}.$$

This is a cartesian diagram. So given a section of A-algebras $B \to B \oplus M$, we have to construct a section of A-algebras $S^{-1}B \to S^{-1}B \oplus M$. We can do this by the universal property of localization, since S acts by invertible elements on $S^{-1}B \oplus M$. (To see this, note that S acts by invertible elements on $S^{-1}B$, and M is a nilpotent ideal.) ☐

Finally, we note that there is an even slicker argument. (We learned this from **?**.) Namely, it suffices to show that $\Omega_{S^{-1}B/B} = 0$, by the exact sequences. But this is a $S^{-1}B$-module, so we have

$$\Omega_{S^{-1}B/B} = \Omega_{S^{-1}B/B} \otimes_B S^{-1}B,$$

because tensoring with $S^{-1}B$ localizes at S, but this does nothing to a $S^{-1}B$-module! By the base change formula (proposition 9.2.14), we have

$$\Omega_{S^{-1}B/B} \otimes_B S^{-1}B = \Omega_{S^{-1}B/S^{-1}B} = 0,$$

where we again use the fact that $S^{-1}B \otimes_B S^{-1}B \simeq S^{-1}B$.

Another construction of $\Omega_{B/A}$

Let B be an A-algebra. We have constructed $\Omega_{B/A}$ by quotienting generators by relations. There is also a simple and elegant "global" construction one sometimes finds useful in generalizing the procedure to schemes.

Consider the algebra $B \otimes_A B$ and the map $B \otimes_A B \to B$ given by multiplication. Note that B acts on $B \otimes_A B$ by multiplication on the first factor: this is how the latter is a B-module, and then the multiplication map is a B-homomorphism. Let $I \subset B \otimes_A B$ be the kernel.

9.2.17 Proposition *There is an isomorphism of B-modules*

$$\Omega_{B/A} \simeq I/I^2$$

given by the derivation $b \mapsto 1 \otimes b - b \otimes 1$, from B to I/I^2.

Proof. It is clear that the maps

$$b \to 1 \otimes b, b \to b \otimes 1 : \quad B \to B \otimes_A B$$

are A-linear, so their difference is too. The quotient $d : B \to I/I^2$ is thus A-linear too.

First, note that if $c, c' \in B$, then $1 \otimes c - c \otimes 1, 1 \otimes c' - c' \otimes 1 \in I$. Their product is thus zero in I/I^2:

$$(1 \otimes c - c \otimes 1)(1 \otimes c' - c' \otimes 1) = 1 \otimes cc' + cc' \otimes 1 - c \otimes c' - c' \otimes c \in I^2.$$

Next we must check that $d : B \to I/I^2$ is a derivation. So fix $b, b' \in B$; we have

$$d(bb') = 1 \otimes bb' - bb' \otimes 1$$

and

$$bdb' = b(1 \otimes b' - b' \otimes 1), \quad b'db = b'(1 \otimes b - b \otimes 1).$$

The second relation shows that

$$bdb' + b'db = b \otimes b' - bb' \otimes 1 + b' \otimes b - bb' \otimes 1.$$

Modulo I^2, we have as above $b \otimes b' + b' \otimes b \equiv 1 \otimes bb' + bb' \otimes 1$, so

$$bdb' + b'db \equiv 1 \otimes bb' - bb' \otimes 1 \mod I^2,$$

and this last is equal to $d(bb')$ by definition. So we have an A-linear derivation $d : B \to I/I^2$. It remains to be checked that this is *universal*. In particular, we must check that the induced

$$\phi : \Omega_{B/A} \to I/I^2$$

sending $db \to 1 \otimes b - b \otimes 1$. is an isomorphism. We can define the inverse $\psi : I/I^2 \to \Omega_{B/A}$ by sending $\sum b_i \otimes b'_i \in I$ to $\sum b_i db'_i$. This is clearly a B-module homomorphism, and it is well-defined mod I^2.

It is clear that $\psi(\phi(db)) = db$ from the definitions, since this is

$$\psi(1 \otimes b - b \otimes 1) = 1(db) - bd1 = db,$$

as $d1 = 0$. So $\psi \circ \phi = 1_{\Omega_{B/A}}$. It follows that ϕ is injective. We will check now that it is surjective. Then we will be done.

9.2.18 Lemma *Any element in I is a B-linear combination of elements of the form $1 \otimes b - b \otimes 1$.*

Every such element is the image of db under ϕ by definition of the derivation $B \to I/I^2$. So this lemma will complete the proof.

Proof. Let $Q = \sum c_i \otimes d_i \in I$. By assumption, $\sum c_i d_i = 0 \in B$. We have by this last identity

$$Q = \sum ((c_i \otimes d_i) - (c_i d_i \otimes 1)) = \sum c_i(1 \otimes d_i - d_i \otimes 1).$$ □

So Q is in the submodule spanned by the $\{1 \otimes b - b \otimes 1\}_{b \in B}$. □

9.3. Introduction to smoothness

Kähler differentials for fields

Let us start with the simplest examples—fields.

9.3.1 Example Let k be a field, k'/k an extension.

9.3.2 Remark (Question) What does $\Omega_{k'/k}$ look like? When does it vanish?

$\Omega_{k'/k}$ is a k'-vector space.

9.3.3 Proposition *Let k'/k be a separable algebraic extension of fields. Then $\Omega_{k'/k} = 0$.*

Proof. We will need a formal property of Kähler differentials that is easy to check, namely that they are compatible with filtered colimits. If $B = \varinjlim B_\alpha$ for A-algebras B_α, then there is a canonical isomorphism

$$\Omega_{B/A} \simeq \varinjlim \Omega_{B_\alpha/A}.$$

One can check this on generators and relations, for instance.

Given this, we can reduce to the case of k'/k finite and separable.

9.3.4 Remark Given a sequence of fields and morphisms $k \to k' \to k''$, then there is an exact sequence

$$\Omega_{k'/k} \otimes k'' \to \Omega_{k''/k} \to \Omega_{k''/k'} \to 0.$$

In particular, if $\Omega_{k'/k} = \Omega_{k''/k'} = 0$, then $\Omega_{k''/k} = 0$. This is a kind of dévissage argument.

Anyway, recall that we have a finite separable extension k'/k where $k' = k(x_1, \dots, x_n)$.[3] We will show that

$$\Omega_{k(x_1,\dots,x_i)/k(x_1,\dots,x_{i-1})} = 0 \quad \forall i,$$

which will imply by the devissage argument that $\Omega_{k'/k} = 0$. In particular, we are reduced to showing the proposition when k' is generated over k by a *single element* x. Then we have that

$$k' \simeq k[X]/(f(X))$$

for $f(X)$ an irreducible polynomial. Set $I = (f(X))$. We have an exact sequence

$$I/I^2 \to \Omega_{k[X]/k} \otimes_{k[X]} k' \to \Omega_{k'/k} \to 0$$

The middle term is a copy of k' and the first term is isomorphic to $k[X]/I \simeq k'$. So there is an exact sequence

$$k' \to k' \to \Omega_{k'/k} \to 0.$$

The first term is, as we have computed, multiplication by $f'(x)$; however this is nonzero by separability. Thus we find that $\Omega_{k'/k} = 0$. ∎

[3] We can take $n = 1$ by the primitive element theorem, but shall not need this.

9.3.5 Remark The above result is **not true** for inseparable extensions in general.

9.3.6 Example Let k be an imperfect field of characteristic $p > 0$. There is $x \in k$ such that $x^{1/p} \notin k$, by definition. Let $k' = k(x^{1/p})$. As a ring, this looks like $k[t]/(t^p - x)$. In writing the exact sequence, we find that $\Omega_{k'/k} = k'$ as this is the cokernel of the map $k' \to k'$ given by multiplication $\frac{d}{dt}|_{x^{1/p}}(t^p - x)$. That polynomial has identically vanishing derivative, though. We find that a generator of $\Omega_{k'/k}$ is dt where t is a pth root of x, and $\Omega_{k'/k} \simeq k$.

Now let us consider transcendental extensions. Let $k' = k(x_1, \dots, x_n)$ be a purely transcendental extension, i.e. the field of rational functions of $x_1, \dots, x_n$.

9.3.7 Proposition *If $k' = k(x_1, \dots, x_n)$, then $\Omega_{k'/k}$ is a free k'-module on the generators dx_i.*

This extends to an *infinitely generated* purely transcendental extension, because Kähler differentials commute with filtered colimits.

Proof. We already know this for the polynomial ring $k[x_1, \dots, x_n]$. However, the rational function field is just a localization of the polynomial ring at the zero ideal. So the result will follow from proposition 9.2.16. □

We have shown that separable algebraic extensions have no Kähler differentials, but that purely transcendental extensions have a free module of rank equal to the transcendence degree.

We can deduce from this:

9.3.8 Corollary *Let L/K be a field extension of fields of char 0. Then*

$$\dim_L \Omega_{L/K} = \operatorname{trdeg}(L/K).$$

Partial proof. Put the above two facts together. Choose a transcendence basis $\{x_\alpha\}$ for L/K. This means that L is algebraic over $K(\{x_\alpha\})$ and the $\{x_\alpha\}$ are algebraically independent. Moreover $L/K(\{x_\alpha\})$ is *separable* algebraic. Now let us use a few things about these cotangent complexes. There is an exact sequence:

$$\Omega_{K(\{x_\alpha\})} \otimes_{K(\{x_\alpha\})} L \to \Omega_{L/K} \to \Omega_{L/K(\{x_\alpha\})} \to 0$$

The last thing is zero, and we know what the first thing is; it's free on the dx_α. So we find that $\Omega_{L/K}$ is generated by the elements dx_α. If we knew that the dx_α were linearly independent, then we would be done. But we don't, yet. □

This is **not true** in characteristic p. If $L = K(\alpha^{1/p})$ for $\alpha \in K$ and $\alpha^{1/p} \notin K$, then $\Omega_{L/K} \neq 0$.

Regularity, smoothness, and Kähler differentials

From this, let us revisit a statement made last time. Let K be an algebraically closed field, let $R = k[x_1, \dots, x_n]/I$ and let $\mathfrak{m} \subset R$ be a maximal ideal. Recall that the Nullstellensatz implies that $R/\mathfrak{m} \simeq k$. We were studying

$$\Omega_{R/k}.$$

This is an R-module, so $\Omega_{R/k} \otimes_R k$ makes sense. There is a surjection

$$\mathfrak{m}/\mathfrak{m}^2 \to \Omega_{R/k} \otimes_R k \to 0,$$

that sends $x \to dx$.

9.3.9 Proposition *This map is an isomorphism.*

Proof. We construct a map going the other way. Call the map $\mathfrak{m}/\mathfrak{m}^2 \to \Omega_{R/k} \otimes_R k$ as ϕ. We want to construct

$$\psi : \Omega_{R/k} \otimes_R k \to \mathfrak{m}/\mathfrak{m}^2.$$

This is equivalent to giving an R-module map

$$\Omega_{R/k} \to \mathfrak{m}/\mathfrak{m}^2,$$

that is a derivation $\partial : R \to \mathfrak{m}/\mathfrak{m}^2$. This acts via $\partial(\lambda + x) = x$ for $\lambda \in k, x \in \mathfrak{m}$. Since $k + \mathfrak{m} = R$, this is indeed well-defined. We must check that ∂ is a derivation. That is, we have to compute $\partial((\lambda + x)(\lambda' + x'))$. But this is

$$\partial(\lambda\lambda' + (\lambda x' + \lambda' x) + xx').$$

The definition of ∂ is to ignore the constant term and look at the nonconstant term mod $\mathfrak{m}^2$. So this becomes

$$\lambda x' + \lambda' x = (\partial(\lambda + x))(x' + \lambda') + (\partial(\lambda' + x'))(x + \lambda)$$

because $xx' \in \mathfrak{m}^2$, and because $\mathfrak{m}$ acts trivially on $\mathfrak{m}/\mathfrak{m}^2$. Thus we get the map ψ in the inverse direction, and one checks that ϕ, ψ are inverses. This is because ϕ sends $x \to dx$ and ψ sends $dx \to x$. ∎

9.3.10 Corollary *Let R be as before. Then $R_\mathfrak{m}$ is regular iff $\dim R_\mathfrak{m} = \dim_k \Omega_{R/k} \otimes_R R/\mathfrak{m}$.*

In particular, the modules of Kähler differentials detect regularity for certain rings.

9.3.11 Definition Let R be a noetherian ring. We say that R is **regular** if $R_\mathfrak{m}$ is regular for every maximal ideal $\mathfrak{m}$. (This actually implies that $R_\mathfrak{p}$ is regular for all primes $\mathfrak{p}$, though we are not ready to see this. It will follow from the fact that the localization of a regular local ring at a prime ideal is regular.)

Let $R = k[x_1, \dots, x_n]/I$ be an affine ring over an algebraically closed field k. Then:

9.3.12 Proposition *TFAE:*

1. *R is regular.*

2. *"R is smooth over k" (to be defined)*
3. $\Omega_{R/k}$ *is a projective module over R of rank* $\dim R$.

A finitely generated projective module is locally free. So the last statement is that $(\Omega_{R/k})_{\mathfrak{p}}$ is free of rank $\dim R$ for each prime $\mathfrak{p}$.

9.3.13 Remark A projective module does not necessarily have a well-defined rank as an integer. For instance, if $R = R_1 \times R_2$ and $M = R_1 \times 0$, then M is a summand of R, hence is projective. But there are two candidates for what the rank should be. The problem is that $\operatorname{Spec} R$ is disconnected into two pieces, and M is of rank one on one piece, and of rank zero on the other. But in this case, it does not happen.

9.3.14 Remark The smoothness condition states that locally on $\operatorname{Spec} R$, we have an isomorphism with $k[y_1, \dots, y_n]/(f_1, \dots, f_m)$ with the gradients ∇f_i linearly independent. Equivalently, if $R_{\mathfrak{m}}$ is the localization of R at a maximal ideal $\mathfrak{m}$, then $R_{\mathfrak{m}}$ is a regular local ring, as we have seen.

Proof. We have already seen that 1 and 2 are equivalent. The new thing is that they are equivalent to 3. First, assume 1 (or 2). First, note that $\Omega_{R/k}$ is a finitely generated R-module; that's a general observation:

9.3.15 Proposition *If $f : A \to B$ is a map of rings that makes B a finitely generated A-algebra, then $\Omega_{B/A}$ is a finitely generated B-module.*

Proof. We've seen this is true for polynomial rings, and we can use the exact sequence. If B is a quotient of a polynomial ring, then $\Omega_{B/A}$ is a quotient of the Kähler differentials of the polynomial ring. □

Return to the main proof. In particular, $\Omega_{R/k}$ is projective if and only if $(\Omega_{R/k})_{\mathfrak{m}}$ is projective for every maximal ideal $\mathfrak{m}$. According to the second assertion, we have that $R_{\mathfrak{m}}$ looks like $(k[y_1, \dots, y_n]/(f_1, \dots, f_m))_{\mathfrak{n}}$ for some maximal ideal $\mathfrak{n}$, with the gradients ∇f_i linearly independent. Thus $(\Omega_{R/k})_{\mathfrak{m}} = \Omega_{R_{\mathfrak{m}}/k}$ looks like the cokernel of

$$R_{\mathfrak{m}}^m \to R_{\mathfrak{m}}^n$$

where the map is multiplication by the Jacobian matrix $\left(\frac{\partial f_i}{\partial y_j}\right)$. By assumption this matrix has full rank. We see that there is a left inverse of the reduced map $k^m \to k^n$. We can lift this to a map $R_{\mathfrak{m}}^n \to R_{\mathfrak{m}}^m$. Since this is a left inverse mod $\mathfrak{m}$, the composite is at least an isomorphism (looking at determinants). Anyway, we see that $\Omega_{R/k}$ is given by the cokernel of a map of free module that splits, hence is projective. The rank is $n - m = \dim R_{\mathfrak{m}}$.

Finally, let us prove that 3 implies 1. Suppose $\Omega_{R/k}$ is projective of rank $\dim R$. So this means that $\Omega_{R_{\mathfrak{m}}/k}$ is free of dimension $\dim R_{\mathfrak{m}}$. But this implies that $(\Omega_{R/k}) \otimes_R R/\mathfrak{m}$ is free of the appropriate rank, and that is—as we have seen already—the embedding dimension $\mathfrak{m}/\mathfrak{m}^2$. So if 3 holds, the embedding dimension equals the usual dimension, and we get regularity. □

9.3.16 Corollary *Let $R = \mathbb{C}[x_1, \dots, x_n]/\mathfrak{p}$ for $\mathfrak{p}$ a prime. Then there is a nonzero $f \in R$ such that $R[f^{-1}]$ is regular.*

Geometrically, this says the following. $\operatorname{Spec} R$ is some algebraic variety, and $\operatorname{Spec} R[f^{-1}]$ is a Zariski open subset. What we are saying is that, in characteristic zero, any algebraic variety has a nonempty open smooth locus. The singular locus is always smaller than the entire variety.

Proof. $\Omega_{R/\mathbb{C}}$ is a finitely generated R-module. Let $K(R)$ be the fraction field of R. Now

$$\Omega_{R/\mathbb{C}} \otimes_R K(R) = \Omega_{K(R)/\mathbb{C}}$$

is a finite $K(R)$-vector space. The dimension is $\operatorname{trdeg}(K(R)/\mathbb{C})$. That is also $d = \dim R$, as we have seen. Choose elements $x_1, \dots, x_d \in \Omega_{R/\mathbb{C}}$ which form a basis for $\Omega_{K(R)/\mathbb{C}}$. There is a map

$$R^d \to \Omega_{R/\mathbb{C}}$$

which is an isomorphism after localization at (0). This implies that there is $f \in R$ such that the map is an isomorphism after localization at f.[4] We find that $\Omega_{R[f^{-1}]/\mathbb{C}}$ is free of rank d for some f, which is what we wanted. ∎

This argument works over any algebraically closed field of characteristic zero, or really any field of characteristic zero.

9.3.17 Remark (Warning) Over imperfect fields in characteristic p, two things can happen:

1. Varieties need not be generically smooth
2. $\Omega_{R/k}$ can be projective with the wrong rank

(Nothing goes wrong for **algebraically closed fields** of characteristic p.)

9.3.18 Example Here is a silly example. Say $R = k[y]/(y^p - x)$ where $x \in K$ has no pth root. We know that $\Omega_{R/k}$ is free of rank one. However, the rank is wrong: the variety has dimension zero.

Last time, were trying to show that $\Omega_{L/K}$ is free on a transcendence basis if L/K is an extension in characteristic zero. So we had a tower of fields

$$K \to K' \to L,$$

where L/K' was separable algebraic. We claim in this case that

$$\Omega_{L/K} \simeq \Omega_{K'/K} \otimes_{K'} L.$$

This will prove the result. But we had not done this yesterday.

Proof. This doesn't follow directly from the previous calculations. Without loss of generality, L is finite over K', and in particular, $L = K'[x]/(f(x))$ for f separable. The claim is that

$$\Omega_{L/K} \simeq (\Omega_{K'/K} \otimes_{K'} L \oplus K'dx)/f'(x)dx + \dots$$

When we kill the vector $f'(x)dx + \dots$, we kill the second component. ∎

[4] There is an inverse defined over the fraction field, so it is defined over some localization.

Part IV.

Homological Algebra

IV.1. Homological algebra à la Cartan–Eilenberg

1.1. Introduction

Homological algebra begins with the notion of a *differential object*, that is, an object with an endomorphism $C \xrightarrow{\partial} C$ such that $\partial^2 = 0$. This equation leads to the obvious inclusion $\mathrm{Im}(\partial) \subset \mathrm{Ker}(\partial)$, but the inclusion generally is not equality. We will find that the difference between $\mathrm{Ker}(\partial)$ and $\mathrm{Im}(\partial)$, called the *homology*, is a highly useful variant of a differential object: its first basic property is that if an exact sequence

$$0 \longrightarrow C' \longrightarrow C \longrightarrow C'' \longrightarrow 0$$

of differential graded objects is given, the homology of C is related to that of C' and C'' through a long exact sequence. The basic example, and the one we shall focus on, is where C is a *chain complex* $(C_k)_{k\in\mathbb{Z}}$, and ∂ is the differential induced by the boundary operators $\partial_k : C_k \to C_{k-1}$. In this case, homology simply measures the failure of a complex to be exact.

After introducing these preliminaries, we develop the theory of *derived functors*. Given a functor that is only left or right-exact, derived functors allow for an extension of a partially exact sequence to a long exact sequence. The most important examples to us, Tor and Ext, provide characterizations of flatness, projectivity, and injectivity.

The classic reference for this part of homological algebra is Cartan & Eilenberg (1999).

1.2. (Co)Chain complexes and their (co) homology

Chain complexes

The chain complex is the most fundamental construction in homological algebra.

1.2.1 Definition Let R be a ring. A *chain complex* (*over* R) is a family of (left) R-modules $(C_k)_{k\in\mathbb{Z}}$ together with so-called *boundary operators* $\partial_k : C_k \to C_{k-1}$, $k \in \mathbb{Z}$, such that $\partial_{k-1}\partial_k = 0$ for all $k \in \mathbb{Z}$. The boundary map ∂ is also called the *differential*. Often, notation is abused and the indices for the boundary map are dropped. A chain complex is often simply denoted by $(C_\bullet, \partial)$ or even only by $C_\bullet$.

One calls a chain complex $C_\bullet$ *bounded below* (respectively *bounded above*) if there exists an $n \in \mathbb{Z}$ such that $C_k = 0$ for all $k \leqslant n$ (respectively $C_k = 0$ for all $k \geqslant n$). If one has $C_k = 0$ for all

$k < 0$ (respectively $C_k = 0$ for all $k > 0$), the chain complex $C_\bullet$ is called *positive* (respectively *negative*). A chain complex $C_\bullet$ is called *bounded* if it is both bounded below and bounded above.

1.2.2 Example Any family of R-modules $(C_k)_{k\in\mathbb{Z}}$ with the boundary operators identically zero forms a chain complex.

We will see plenty of more examples in due time.

1.2.3 Proposition *If $(C_\bullet, \partial)$ is a chain complex, then* $\operatorname{Im}\partial_{k+1} \subset \operatorname{Ker}\partial_k$ *for each $k \in \mathbb{Z}$.*

Proof. The claim is an immediate consequence of the relation $\partial_k\partial_{k+1} = 0$. □

The observation from the proposition leads us to the following definition.

1.2.4 Definition Let $(C_\bullet, \partial)$ be a chain complex. For each $k \in \mathbb{Z}$ one calls the module C_k the module of *k-chains*. The submodule of *k-cycles* $Z_k \subset C_k$ is the kernel $\operatorname{Ker}(\partial_k)$. The submodule of *$k$-boundaries* $B_k \subset C_k$ is the image $\operatorname{Im}(\partial_{k+1})$. The k-th *homology group* of the complex $(C_\bullet, \partial)$ is now defined as the R-module $H_k(C_\bullet) := H_k(C_\bullet, \partial) := Z_k/B_k$. The family $H_\bullet(C_\bullet) = \big(H_k(C_\bullet)\big)_{k\in\mathbb{Z}}$ is usually referred to as the *homology* of $(C_\bullet, \partial)$.

A chain complex $(C_\bullet, \partial)$ for which $Z_k = B_k$ or equivalently $H_k(C_\bullet) = 0$ for every $k \in \mathbb{Z}$ is called *exact*.

1.2.5 Remark In general, a chain complex need not be exact, and this failure of exactness is measured by its homology.

1.2.6 Examples (a) In a chain complex $(C_\bullet, \partial)$ where all the boundary maps are trivial, i.e. where $\partial = 0$, one has $H_k(C_\bullet) = C_k$ for all $k \in \mathbb{Z}$.

(b) The homology $H_\bullet(C_\bullet)$ of a chain complex $C_\bullet$ can and will be understood as a chain complex again with boundary maps being trivial. This interpretation will be very useful when studying formality in rational or real homotopy theory, see **??**.

We have defined chain complexes now, but we have no notion of a morphism between chain complexes yet. We do this next; it turns out that chain complexes form a category when morphisms are appropriately defined.

1.2.7 Definition A *morphism of chain complexes* (*over the ring R*) from $(C_\bullet, \partial)$ to $(D_\bullet, \delta)$ or a *chain map* is a family of R-module maps $f_k : C_k \to D_k$, $k \in \mathbb{Z}$, such that $f_{k-1}\partial_k = \delta_k f_k$ for all $k \in \mathbb{Z}$. In other words this means that the diagram

$$\begin{tikzcd} \cdots \arrow[r, dashed] & C_{k+1} \arrow[r, "\partial_{k+1}"] \arrow[d, "f_{k+1}"] & C_k \arrow[r, "\partial_k"] \arrow[d, "f_k"] & C_{k-1} \arrow[r, dashed] \arrow[d, "f_{k-1}"] & \cdots \\ \cdots \arrow[r, dashed] & D_{k+1} \arrow[r, "\delta_{k+1}"'] & D_k \arrow[r, "\delta_k"'] & D_{k-1} \arrow[r, dashed] & \cdots \end{tikzcd}$$

commutes. We will denote such a morphism of chain complexes by $f : (C_\bullet, \partial) \to (D_\bullet, \delta)$.

1.2.8 Remark To further simplify notation, often all differentials regardless of what chain complex they are part of are denoted ∂, thus the commutativity relation on chain maps is simply $f\partial = \partial f$ with indices and distinction between the boundary operators dropped. Sometimes, though, when a distinction is really necessary, one writes ∂^C or ∂^D to denote the boundery map of $C_\bullet$ respectively $D_\bullet$. We will make sure in this book that the context or the notation will always make clear what is meant.

1.2.9 Proposition and Definition *Chain complexes over a ring R together with their chain maps as morphisms become a category which we denote by $\mathsf{Ch}_\bullet(R\text{-}\mathsf{Mod})$ or just $\mathsf{Ch}_\bullet$ when the ground ring R is clear. The chain complexes bounded below (respectively bounded above, bounded, positive, or negative) form a full subcategory of $\mathsf{Ch}_\bullet(R\text{-}\mathsf{Mod})$. The resulting subcategories are denoted by $\mathsf{Ch}^+_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^-_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^{\mathrm{b}}_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^{\geqslant 0}_\bullet(R\text{-}\mathsf{Mod})$, and $\mathsf{Ch}^{\leqslant 0}_\bullet(R\text{-}\mathsf{Mod})$, respectively.*

Proof. If $(C_\bullet, \partial)$ is a chain complex, then the family of identity maps $\mathrm{id}_{C_k} : C_k \to C_k$ is clearly a chain map which we denote by $\mathrm{id}_{C_\bullet}$. If $f : (C_\bullet, \partial) \to (D_\bullet, \delta)$ and $g : (D_\bullet, \delta) \to (E_\bullet, \varrho)$ are chain maps, then $g \circ f : (C_\bullet, \partial) \to (E_\bullet, \varrho)$ with components $(g \circ f)_k := g_k \circ f_k : C_k \to E_k$ is a chain map as well, since for all $k \in \mathbb{Z}$

$$(g \circ f)_{k-1}\partial_k = g_{k-1} \circ f_{k-1} \circ \partial_k = g_{k-1} \circ \delta_k \circ f_k = \varrho_k \circ g_k \circ f_k = \varrho_k(g \circ f)_k \ .$$

Hence the chain complexes over the ring R together with the chain maps form a category indeed. The rest of the claim is obvious. $\square$

1.2.10 Proposition *A chain map $f : C_\bullet \to D_\bullet$ between chain complexes over a ring R induces for each $k \in \mathbb{Z}$ a map in homology $H_k(f) : H_k(C_\bullet) \to H_k(D_\bullet)$. More precisely, each H_k is a functor from chain complexes to R-modules, and homology becomes a covariant functor from the category of chain complexes to the category of chain complexes with zero differential.*

Proof. Let $f : C_\bullet \to D_\bullet$ be a chain map. Let ∂ and δ be the differentials for $C_\bullet$ and $D_\bullet$ respectively. Then we have a commutative diagram:

$$\begin{tikzcd}
\cdots \arrow[r, dashed] & C_{k+1} \arrow[r, "\partial_{k+1}"] \arrow[d, "f_{k+1}"] & C_k \arrow[r, "\partial_k"] \arrow[d, "f_k"] & C_{k-1} \arrow[r, dashed] \arrow[d, "f_{k-1}"] & \cdots \\
\cdots \arrow[r, dashed] & D_{k+1} \arrow[r, "\delta_{k+1}"'] & D_k \arrow[r, "\delta_k"'] & D_{k-1} \arrow[r, dashed] & \cdots
\end{tikzcd} \quad .$$

Now, in order to check that the chain map f induces a map $H_k(f)$ on homology, we need to check that $f(\mathrm{Im}(\partial)) \subset \mathrm{Im}(\delta)$ and $f(\mathrm{Ker}(\partial)) \subset \mathrm{Ker}(\delta)$. We first check the condition on images: we want to look at $f_k(\mathrm{Im}(\partial_{k+1}))$. By commutativity of f and the boundary maps, this is equal to $\delta_{k+1}(\mathrm{Im}(f_{k+1})$. Hence we have $f_k(\mathrm{Im}(\partial_{k+1})) \subset \mathrm{Im}(\delta_{k+1})$. For the condition on kernels, let $c \in \mathrm{Ker}(\partial_k)$. Then by commutativity, $\delta_k(f_k(c)) = f_{k-1}\partial_k(c) = 0$. Thus we have that f induces for each $k \in \mathbb{Z}$ an R-module map $H_k(f) : H_k(C_\bullet) \to H_k(D_\bullet)$. Hence it induces a morphism on homology as a chain complex with zero differential. $\square$

Long exact sequences

add: OMG! We have all this and not the most basic theorem of them all.

1.2.11 Definition If M is a complex then for any integer k, we define a new complex $M[k]$ by shifting indices, i.e. $(M[k])^i := M^{i+k}$.

1.2.12 Definition If $f : M \to N$ is a map of complexes, we define a complex $\mathrm{Cone}(f) := \{N^i \oplus M^{i+1}\}$ with differential

$$d(n^i, m^{i+1}) := (d_N^i(n_i) + (-1)^i \cdot f(m^{i+1}, d_M^{i+1}(m^{i+1}))$$

Remark: This is a special case of the total complex construction to be seen later.

1.2.13 Proposition *A map $f : M \to N$ is a quasi-isomorphism if and only if* $\mathrm{Cone}(f)$ *is acyclic.*

1.2.14 Proposition *Note that by definition we have a short exact sequence of complexes*

$$0 \to N \to \mathrm{Cone}(f) \to M[1] \to 0$$

so by Proposition 2.1, we have a long exact sequence

$$\cdots \to H^{i-1}(\mathrm{Cone}(f)) \to H^i(M) \to H^i(N) \to H^i(\mathrm{Cone}(f)) \to \ldots$$

so by exactness, we see that $H^i(M) \simeq H^i(N)$ if and only if $H^{i-1}(\mathrm{Cone}(f)) = 0$ and $H^i(\mathrm{Cone}(f)) = 0$. Since this is the case for all i, the claim follows. ■

Cochain complexes

Cochain complexes are much like chain complexes except the arrows point in the opposite direction. Like before, R denotes a fixed ring.

1.2.15 Definition A *cochain complex* is a sequence of R-modules $(C^k)_{k\in\mathbb{Z}}$ with *coboundary operators*, also called *differentials*, $d^k : C^k \to C^{k+1}$, $k \in \mathbb{Z}$, such that $d^{k+1}d^k = 0$. A cochain copmplex is usually denoted by $(C^\bullet, d)$ or shortly by $C^\bullet$.

One calls a cochain complex $C^\bullet$ *bounded below* (respectively *bounded above*) if there exists an $n \in \mathbb{Z}$ such that $C^k = 0$ for all $k \leqslant n$ (respectively $C^k = 0$ for all $k \geqslant n$). If one has $C^k = 0$ for all $k < 0$ (respectively $C^k = 0$ for all $k > 0$), the cochain complex $C^\bullet$ is called *positive* (respectively *negative*). A cochain complex $C^\bullet$ which is both bounded below and bounded above is said to be *bounded*.

Let $(C^\bullet, d)$ and $(D^\bullet, \delta)$ denote cochain complexes. By a *morphism of cochain complexes* or a *cochain map* from $(C^\bullet, d)$ to $(D^\bullet, \delta)$ we understand a family of R-module maps $g^k : C^k \to D^k$, $k \in \mathbb{Z}$, such that $g^{k+1}d^k = \delta^k g^k$ for all $k \in \mathbb{Z}$. In other words this means we have a commutative diagram:

$$
\begin{tikzcd}
\cdots \arrow[r, dashed] & C^{k-1} \arrow[r, "d^{k-1}"] \arrow[d, "g^{k-1}"] & C^k \arrow[r, "d^k"] \arrow[d, "g^k"] & C^{k+1} \arrow[r, dashed] \arrow[d, "g^{k+1}"] & \cdots \\
\cdots \arrow[r, dashed] & D^{k-1} \arrow[r, "\delta^{k-1}"'] & D^k \arrow[r, "\delta^k"'] & D^{k-1} \arrow[r, dashed] & \cdots
\end{tikzcd}
.
$$

We will denote such a morphism of cochain complexes usually by $g : (C^\bullet, d) \to (D^\bullet, \delta)$.

1.2.16 Proposition and Definition *Cochain complexes over a ring R together with their cochain maps as morphisms become a category which we denote by $\mathsf{Ch}^\bullet(R\text{-}\mathsf{Mod})$ or just $\mathsf{Ch}^\bullet$ when the ground ring R is clear. The cochain complexes bounded below (respectively bounded above, bounded, positive, or negative) form a full subcategory of $\mathsf{Ch}^\bullet(R\text{-}\mathsf{Mod})$. The corresponding subcategories are denoted by $\mathsf{Ch}^\bullet_+(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_-(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_{\mathrm{b}}(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_{\geqslant 0}(R\text{-}\mathsf{Mod})$, and $\mathsf{Ch}^\bullet_{\leqslant 0}(R\text{-}\mathsf{Mod})$, respectively.*

Proof. The proof is completely dual to the proof of Proposition 1.2.16. □

The theory of cochain complexes is entirely dual to that of chain complexes, and we often shall not spell it out in detail.

For instance, we can form a category of cochain complexes and **chain maps** (families of morphisms commuting with the differential). Moreover, given a cochain complex $C^\bullet$, we define the **cohomology objects** to be $h^i(C^*) = \ker(\partial^i)/Im(\partial^{i-1})$; one obtains cohomology functors.

It should be noted that the long exact sequence in cohomology runs in the opposite direction. If $0 \to C'_* \to C_* \to C''_* \to 0$ is a short exact sequence of cochain complexes, we get a long exact sequence

$$\cdots \to H^i(C') \to H^i(C) \to H^i(C'') \to H^{i+1}(C') \to H^{i+1}(C) \to \ldots.$$

Similarly, we can also turn cochain complexes and cohomology modules into a graded module.

Let us now give a standard example of a cochain complex.

1.2.17 Example (The de Rham complex) Readers unfamiliar with differential forms may omit this example. Let M be a smooth manifold. For each p, let $C^p(M)$ be the $\mathbb{R}$-vector space of smooth p-forms on M. We can make the $\{C^p(M)\}$ into a complex by defining the maps

$$C^p(M) \to C^{p+1}(M)$$

via $\omega \to d\omega$, for d the exterior derivative. (Note that $d^2 = 0$.) This complex is called the **de Rham complex** of M, and its cohomology is called the **de Rham cohomology.** It is known that the de Rham cohomology is isomorphic to singular cohomology with real coefficients, cf. **?** and Hatcher (2002).

1.3. Chain Homotopies

1.3.1 In general, two maps of complexes $C_\bullet \rightrightarrows D_\bullet$ need not be equal to induce the same morphisms in homology. It is thus of interest to determine conditions when they do. One important condition is given by chain homotopy: chain homotopic maps are indistinguishable in homology. In algebraic topology, this fact is used to show that singular homology is a homotopy invariant. We will find it useful in showing that the construction (to be given later) of a projective resolution is essentially unique.

As before, we will understand all of the following constructions to be performed within the category R-Mod of left modules over a fixed ring R, unless stated differently.

1.3.2 Definition Let $C_\bullet, D_\bullet$ be chain complexes with differentials ∂^C and ∂^D, respectively. A chain homotopy between two chain maps $f, g : C_\bullet \to D_\bullet$ is a sequence of homomorphism $h_k : C_k \to D_{k+1}$, $k \in \mathbb{Z}$ satisfying

$$f_k - g_k = \partial^D_{k+1} h_k + h_{k-1}\partial^C_k \quad \text{for all } k \in \mathbb{Z} \; .$$

Again, often notation is abused and the condition is written $f - g = \partial h + h\partial$.

Dually, if $C^\bullet$ and $D^\bullet$ are two cochain complexes with respective differentials d_C and d_D, then a chain homotopy between two morphisms of cochain complexes $f, g : C^b ullet \to D^b ullet$ is a sequence of homomorphisms $h^k : C^k \to D^{k-1}$, $k \in \mathbb{Z}$ satisfying

$$f^k - g^k = d^{k-1}_D h^k + h^{k+1} d^k_C \quad \text{for all } k \in \mathbb{Z} \; .$$

or shortly $f - g = dh + hd$.

1.3.3 Proposition *If two morphisms of chain complexes $f, g : C_\bullet \to D_\bullet$ are chain homotopic, they are taken to the same induced map after applying the homology functor. Likewise, two chain homotopic morphisms of cochain complexes $f, g : C_\bullet \to D_\bullet$ induce the same map in cohomology.*

Proof. Write $\{d_i\}$ for the various differentials (in both complexes). Let $m \in Z_i(C)$, the group of i-cycles. Suppose there is a chain homotopy h between f, g (that is, a set of morphisms $C_i \to D_{i-1}$). Then

$$f^i(m) - g^i(m) = h^{i+1} \circ d^i(m) + d^{i-1} \circ h^i(m) = d^{i-1} \circ H^i(m) \in \mathfrak{Im}(d^{i-1})$$

which is zero in the cohomology $H^i(D)$. □

1.3.4 Corollary *If two chain complexes are chain homotopically equivalent (there are maps $f : C_* \to D_*$ and $g : D_* \to C_*$ such that both fg and gf are chain homotopic to the identity), they have isomorphic homology.*

Proof. Clear. □

1.3.5 Example Not every quasi-isomorphism is a homotopy equivalence. Consider the complex

$$\cdots \to 0 \to \mathbb{Z}/\!\cdot 2 \to \mathbb{Z} \to 0 \to 0 \to \ldots$$

so $H^0 = \mathbb{Z}/2\mathbb{Z}$ and all cohomologies are 0. We have a quasi-isomorphism from the above complex to the complex

$$\dots \to 0 \to 0 \to \mathbb{Z}/2\mathbb{Z} \to 0 \to 0 \to \dots$$

but no inverse can be defined (no map from $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$).

1.3.6 Proposition *Additive functors preserve chain homotopies*

Proof. Since an additive functor F is a homomorphism on $Hom(-,-)$, the chain homotopy condition will be preserved; in particular, if t is a chain homotopy, then $F(t)$ is a chain homotopy.□

In more sophisticated homological theory, one often makes the definition of the "homotopy category of chain complexes."

1.3.7 Definition The homotopy category of chain complexes is the category $hKom(R)$ where objects are chain complexes of R-modules and morphisms are chain maps modulo chain homotopy.

1.4. Differential modules

Often we will bundle all the modules C_k of a chain complex $C_\bullet$ together to form a graded module $\bigoplus_k C_k$. In this case, the boundary operator is an endomorphism that takes elements from degree k to degree $k-1$. Similarly, we often bundle together all the homology modules to give a graded homology module $\bigoplus_k H_k(C_\bullet)$.

1.4.1 Definition A *differential module* over a ring R is a (left) R-module M together with a morphism $d : M \to M$ such that $d^2 = 0$.

Thus, given a chain complex $C_\bullet$, the module $\bigoplus_{k\in\mathbb{Z}} C_k$ is a differential module with the direct sum of all the differentials ∂_k. A chain complex is just a special kind of differential module, one where the objects are graded and the differential drops the grading by one.

As we have sen, there is a category of chain complexes where the morphisms are chain maps. One can make a similar definition for differential modules.

1.4.2 Definition If (M, d) and (N, d') are differential modules, then a *morphism of differential modules* $(M, d) \to (N, d')$ is a morphism of modules $M \to N$ such that the diagram

$$\begin{tikzcd} M \arrow[r, "d"] \arrow[d] & M \arrow[d] \\ N \arrow[r, "d'"] & N \end{tikzcd}$$

commutes.

There is therefore a category of differential modules, and the map $C_* \to \bigoplus C_i$ gives a functor from the category of chain complexes to that of differential modules.

1.4.3 Remark Define the *homology* $H(M)$ of a differential module (M, d) via $\ker d / \operatorname{im} d$. Show that $M \mapsto H(M)$ is a functor from differential modules to modules.

1.5. Derived functors

Projective resolutions

Fix a ring R. Let us recall (4.2.7) that an R-module P is called *projective* if the functor $N \to \hom_R(P, N)$ (which is always left-exact) is exact.

Projective objects are useful in defining chain exact sequences known as "projective resolutions." In the theory of derived functors, the projective resolution of a module M is in some sense a replacement for M: thus, we want it to satisfy some uniqueness and existence properties. The uniqueness is not quite true, but it is true modulo chain equivalence.

1.5.1 Definition Let M be an arbitrary module, a projective resolution of M is an exact sequence

$$\cdots \to P_i \to P_{i-1} \to P_{i-2} \cdots \to P_1 \to P_0 \to M \tag{1.5.1}$$

where the P_i are projective modules.

1.5.2 Proposition *Any module admits a projective resolution.*

The proof will even show that we can take a *free* resolution.

Proof. We construct the resolution inductively. First, we take a projective module P_0 with $P_0 \twoheadrightarrow N$ surjective by the previous part. Given a portion of the resolution

$$P_n \to P_{n-1} \to \cdots \to P_0 \twoheadrightarrow N \to 0$$

for $n \geqslant 0$, which is exact at each step, we consider $K = \ker(P_n \to P_{n-1})$. The sequence

$$0 \to K \to P_n \to P_{n-1} \to \cdots \to P_0 \twoheadrightarrow N \to 0$$

is exact. So if P_{n+1} is chosen such that it is projective and there is an epimorphism $P_{n+1} \twoheadrightarrow K$, (which we can construct by 2.8.6), then

$$P_{n+1} \to P_n \to \dots$$

is exact at every new step by construction. We can repeat this inductively and get a full projective resolution. $\square$

Here is a useful observation:

1.5.3 Proposition *If R is noetherian, and M is finitely generated, then we can choose a projective resolution where each P_i is finitely generated.*

We can even take a resolution consisting of finitely generated free modules.

Proof. To say that M is finitely generated is to say that it is a quotient of a free module on finitely many generators, so we can take P_0 free and finitely generated. The kernel of $P_0 \to M$ is finitely generated by noetherianness, and we can proceed as before, at each step choosing a finitely generated object. $\square$

1.5.4 Example The abelian group $\mathbb{Z}/2$ has the free resolution $0 \to \cdots 0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2$. Similarly, since any finitely generated abelian group can be decomposed into the direct sum of torsion subgroups and free subgroups, all finitely generated abelian groups admit a free resolution of length two.

Actually, over a principal ideal domain R (e.g. $R = \mathbb{Z}$), *every* module admits a free resolution of length two. The reason is that if $F \twoheadrightarrow M$ is a surjection with F free, then the kernel $F' \subset F$ is free by a general fact (**add: citation needed**) that a submodule of a free module is free (if one works over a PID). So we get a free resolution of the type

$$0 \to F' \to F \to M \to 0.$$

In general, projective resolutions are not at all unique. Nonetheless, they *are* unique up to chain homotopy. Thus a projective resolution is a rather good "replacement" for the initial module.

1.5.5 Proposition *Let M, N be modules and let $P_* \to M, P'_* \to N$ be projective resolutions. Let $f : M \to N$ be a morphism. Then there is a morphism*

$$P_* \to P'_*$$

such that the following diagram commutes:

$$\begin{tikzcd}
\dots \arrow[r] & P_1 \arrow[r] \arrow[d] & P_0 \arrow[r] \arrow[d] & M \arrow[d, "f"] \\
\dots \arrow[r] & P'_1 \arrow[r] & P'_0 \arrow[r] & N
\end{tikzcd}$$

This morphism is unique up to chain homotopy.

Proof. Let $P_* \to M$ and $P'_* \to N$ be projective resolutions. We will define a morphism of complexes $P_* \to P'_*$ such that the diagram commutes. Let the boundary maps in P_*, P'_* be denoted d (by abuse of notation). We have an exact diagram

$$\begin{tikzcd}
\dots \arrow[r] & P_n \arrow[r, "d"] & P_{n-1} \arrow[r, "d"] & \dots \arrow[r, "d"] & P_0 \arrow[r] & M \arrow[r] \arrow[d, "f"] & 0 \\
\dots \arrow[r] & P'_n \arrow[r, "d"] & P'_{n-1} \arrow[r] & \dots \arrow[r, "d"] & P'_0 \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

Since $P'_0 \twoheadrightarrow N$ is an epimorphism, the map $P_0 \to M \to N$ lifts to a map $P_0 \to P'_0$ making the diagram

$$\begin{tikzcd}
P_0 \arrow[r] \arrow[d] & M \arrow[d, "f"] \\
P'_0 \arrow[r] & N
\end{tikzcd}$$

commute. Suppose we have defined maps $P_i \to P'_i$ for $i \leqslant n$ such that the following diagram commutes:

$$\begin{tikzcd}
P_n \arrow[r, "d"] \arrow[d] & P_{n-1} \arrow[r, "d"] \arrow[d] & \dots \arrow[r, "d"] & P_0 \arrow[r] \arrow[d] & M \arrow[r] \arrow[d, "f"] & 0 \\
P'_n \arrow[r, "d"] & P'_{n-1} \arrow[r] & \dots \arrow[r, "d"] & P'_0 \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

Then we will define $P_{n+1} \to P'_{n+1}$, after which induction will prove the existence of a map. To do this, note that the map

$$P_{n+1} \to P_n \to P'_n \to P'_{n-1}$$

is zero, because this is the same as $P_{n+1} \to P_n \to P_{n-1} \to P'_{n-1}$ (by induction, the diagrams before n commute), and this is zero because two P-differentials were composed one after another. In particular, in the diagram

$$\begin{tikzcd} P_{n+1} \arrow[r] & P_n \arrow[d] \\ P'_{n+1} \arrow[r] & P'_n \end{tikzcd},$$

the image in P'_n of P_{n+1} lies in the kernel of $P'_n \to P'_{n-1}$, i.e. in the image I of P'_{n+1}. The exact diagram

$$\begin{tikzcd} & P_{n+1} \arrow[d] & \\ P'_{n+1} \arrow[r] & I \arrow[r] & 0 \end{tikzcd}$$

shows that we can lift $P_{n+1} \to I$ to $P_{n+1} \to P'_{n+1}$ (by projectivity). This implies that we can continue the diagram further and get a morphism $P_* \to P'_*$ of complexes.

Suppose $f, g : P_* \to P'_*$ are two morphisms of the projective resolutions making

$$\begin{tikzcd} P_0 \arrow[r] \arrow[d] & M \arrow[d] \\ P'_0 \arrow[r] & N \end{tikzcd}$$

commute. We will show that f, g are chain homotopic.

For this, we start by defining $D_0 : P_0 \to P'_1$ such that $dD_0 = f - g : P_0 \to P'_0$. This we can do because $f - g$ sends P_0 into $\ker(P'_0 \to N)$, i.e. into the image of $P'_1 \to P'_0$, and P_0 is projective. Suppose we have defined chain-homotopies $D_i : P_i \to P_{i+1}$ for $i \leqslant n$ such that $dD_i + D_{i-1}d = f - g$ for $i \leqslant n$. We will define D_{n+1}. There is a diagram

$$\begin{tikzcd} & P_{n+1} \arrow[r] \arrow[d] & P_n \arrow[r] \arrow[d] \arrow[dl, "D_n"] & P_{n-1} \arrow[d] \arrow[dl, "D_{n-1}"] \\ P'_{n+2} \arrow[r] & P'_{n+1} \arrow[r] & P'_n \arrow[r] & P'_{n-1} \end{tikzcd}$$

where the squares commute regardless of whether you take the vertical maps to be f or g (provided that the choice is consistent).

We would like to define $D_{n+1} : P_n \to P'_{n+1}$. The key condition we need satisfied is that

$$dD_{n+1} = f - g - D_n d.$$

However, we know that, by the inductive hypothesis on the D's

$$d(f - g - D_n d) = fd - gd - dD_n d = fd - gd - (f - g)d + D_n dd = 0. \qquad \square$$

In particular, $f - g - D_n d$ lies in the image of $P'_{n+1} \to P'_n$. The projectivity of P_n ensures that we can define D_{n+1} satisfying the necessary condition.

1.5.6 Corollary *Let $P_* \to M, P'_* \to M$ be projective resolutions of M. Then there are maps $P_* \to P'_*, P'_* \to P_*$ under M such that the compositions are chain homotopic to the identity.*

Proof. Immediate. □

Injective resolutions

One can dualize all this to injective resolutions. **add: do this**

Definition

Often in homological algebra, we see that "short exact sequences induce long exact sequences." Using the theory of derived functors, we can make this formal.

Let us work in the category of modules over a ring R. Fix two such categories. Recall that a right-exact functor F (from the category of modules over a ring to the category of modules over another ring) is an additive functor such that for every short exact sequence $0 \to A \to B \to C \to 0$, we get a exact sequence $F(A) \to F(B) \to F(C) \to 0$.

We want a natural way to continue this exact sequence to the left; one way of doing this is to define the left derived functors.

1.5.7 Definition Let F be a right-exact functor and $P_* \to M$ are projective resolution. We can form a chain complex $F(P_*)$ whose object in degree i is $F(P_i)$ with boundary maps $F(\partial)$. The homology of this chain complex denoted L_iF are the left derived functors.

For this definition to be useful, it is important to verify that deriving a functor yields functors independent on choice of resolution. This is clear by **??**.

1.5.8 Theorem *The following properties characterize derived functors:*

1. $L_0F(-) = F(-)$
2. *Suppose $0 \to A \to B \to C \to 0$ is an exact sequence and F a right-exact functor; the left derived functors fit into the following exact sequence:*

$$\cdots L_iF(A) \to L_iF(B) \to L_iF(C) \to L_{i-1}F(A) \cdots \to L_1(C) \to L_0F(A) \to L_0F(B) \to L_0F(C) \to 0 \tag{1.5.2}$$

Proof. The second property is the hardest to prove, but it is by far the most useful; it is essentially an application of the snake lemma. □

One can define right derived functors analogously; if one has a left exact functor (an additive functor that takes an exact sequence $0 \to A \to B \to C \to 0$ to $0 \to F(A) \to F(B) \to F(C)$), we can pick an injective resolution instead (the injective criterion is simply the projective criterion with arrows reversed). If $M \to I^*$ is a injective resolution then the cohomology of the chain complex $F(I^*)$ gives the right derived functors. However, variance must also be taken into consideration so the choice of whether or not to use a projective or injective resolution is of importance (in all of the above, functors were assumed to be covariant). In the following, we see an example of when right derived functors can be computed using projective resolutions.

Ext **functors**

1.5.9 Definition The right derived functors of $Hom(-, N)$ are called the Ext-modules denoted $Ext^i_R(-, N)$.

We now look at the specific construction:

Let M, M' be R-modules. Choose a projective resolution

$$\cdots \to P_2 \to P_1 \to P_0 \to M \to 0$$

and consider what happens when you hom this resolution into N. Namely, we can consider $\hom_R(M, N)$, which is the kernel of $\hom(P_0, M) \to \hom(P_1, M)$ by exactness of the sequence

$$0 \to \hom_R(M, N) \to \hom_R(P_0, N) \to \hom_R(P_1, N).$$

You might try to continue this with the sequence

$$0 \to \hom_R(M, N) \to \hom_R(P_0, N) \to \hom_R(P_1, N) \to \hom_R(P_2, N) \to \dots.$$

In general, it won't be exact, because $\hom_R$ is only left-exact. But it is a chain complex. You can thus consider the homologies.

1.5.10 Definition The homology of the complex $\{\hom_R(P_i, N)\}$ is denoted $\mathrm{Ext}^i_R(M, N)$. By definition, this is $\ker(\hom(P_i, N) \to \hom(P_{i+1}, N))/\operatorname{im}(\hom(P_{i-1}, N) \to \hom(P_i, N))$. This is an R-module, and is called the ith ext group.

Let us list some properties (some of these properties are just case-specific examples of general properties of derived functors)

1.5.11 Proposition $\mathrm{Ext}^0_R(M, N) = \hom_R(M, N)$.

Proof. This is obvious from the left-exactness of $\hom(-, N)$. (We discussed this.) □

1.5.12 Proposition $\mathrm{Ext}^i(M, N)$ *is a functor of* N.

Proof. Obvious from the definition. □

Here is a harder statement.

1.5.13 Proposition $\mathrm{Ext}^i(M,N)$ *is well-defined, independent of the projective resolution* $P_* \to M$*, and is in fact a contravariant additive functor of* M*.*[1]

Proof. Omitted. We won't really need this, though; it requires more theory about chain complexes. ☐

1.5.14 Proposition *If* M *is annihilated by some ideal* $I \subset R$*, then so is* $\mathrm{Ext}^i(M,N)$ *for each* i*.*

Proof. This is a consequence of the functoriality in M. If $x \in I$,then $x : M \to M$ is the zero map, so it induces the zero map on $\mathrm{Ext}^i(M,N)$.

1.5.15 Proposition $\mathrm{Ext}^i(M,N) = 0$ *if* M *projective and* $i > 0$*.*

Proof. In that case, one can use the projective resolution

$$0 \to M \to M \to 0.$$

Computing Ext via this gives the result. ☐

1.5.16 Proposition *If there is an exact sequence*

$$0 \to N' \to N \to N'' \to 0,$$

there is a long exact sequence of Ext *groups*

$$0 \to \hom(M,N') \to \hom(M,N) \to \hom(M,N'') \to \mathrm{Ext}^1(M,N') \to \mathrm{Ext}^1(M,N) \to \dots$$

Proof. This proof will assume a little homological algebra. Choose a projective resolution $P_* \to M$. (The notation P_* means the chain complex $\dots \to P_2 \to P_1 \to P_0$.) In general, homming out of M is not exact, but homming out of a projective module is exact. For each i, we get an exact sequence

$$0 \to \hom_R(P_i,N') \to \hom_R(P_i,N) \to \hom_R(P_i,N'') \to 0,$$

which leads to an exact sequence of *chain complexes*

$$0 \to \hom_R(P_*,N') \to \hom_R(P_*,N) \to \hom_R(P_*,N'') \to 0.$$

Taking the long exact sequence in homology gives the result. ☐

Much less obvious is:

1.5.17 Proposition *There is a long exact sequence in the* M *variable. That is, a short exact sequence*

$$0 \to M' \to M \to M'' \to 0$$

leads a long exact sequence

$$0 \to \hom_R(M'',N) \to \hom_R(M,N) \to \hom_R(M',N) \to \mathrm{Ext}^1(M'',N) \to \mathrm{Ext}^1(M,N) \to \dots.$$

[1] I.e. a map $M \to M'$ induces $\mathrm{Ext}^i(M',N) \to \mathrm{Ext}^i(M,N)$.

Proof. Omitted. □

We now can characterize projectivity:

1.5.18 Corollary *TFAE:*

1. *M is projective.*
2. $\mathrm{Ext}^i(M, N) = 0$ *for all R-modules N and $i > 0$.*
3. $\mathrm{Ext}^1(M, N) = 0$ *for all N.*

Proof. We have seen that 1 implies 2 because projective modules have simple projective resolutions. 2 obviously implies 3. Let's show that 3 implies 1.Choose a projective module P and a surjection $P \twoheadrightarrow M$ with kernel K. There is a short exact sequence $0 \to K \to P \to M \to 0$. The sequence

$$0 \to \hom(M, K) \to \hom(P, K) \to \hom(K, K) \to \mathrm{Ext}^1(M, K) = 0$$

shows that there is a map $P \to K$ which restricts to the identity $K \to K$. The sequence $0 \to K \to P \to M \to 0$ thus splits, so M is a direct summand in a projective module, so is projective. □

Finally, we note that there is another way of constructing Ext. We constructed them by choosing a projective resolution of M. But you can also do this by resolving N by *injective* modules.

1.5.19 Definition An R-module Q is **injective** if $\hom_R(-, Q)$ is an exact (or, equivalently, right-exact) functor. That is, if $M_0 \subset M$ is an inclusion of R-modules, then any map $M_0 \to Q$ can be extended to $M \to Q$.

If we are given M, N, and an injective resolution $N \to Q_*$, we can look at the chain complex $\{\hom(M, Q_i)\}$, i.e. the chain complex

$$0 \to \hom(M, Q^0) \to \hom(M, Q^1) \to \dots$$

and we can consider the cohomologies.

1.5.20 Definition We call these cohomologies

$$\mathrm{Ext}^i_R(M, N)' = \ker(\hom(M, Q^i) \to \hom(M, Q^{i+1}))/\operatorname{im}(\hom(M, Q^{i-1}) \to \hom(M, Q^i)).$$

This is dual to the previous definitions, and it is easy to check that the properties that we couldn't verify for the previous Exts are true for the Ext$'$'s.

Nonetheless:

1.5.21 Theorem *There are canonical isomorphisms:*

$$\mathrm{Ext}^i(M, N)' \simeq \mathrm{Ext}^i(M, N).$$

In particular, to compute Ext groups, you are free either to take a projective resolution of M, or an injective resolution of N.

Idea of proof. In general, it might be a good idea to construct a third more complex construction that resembles both. Given M, N construct a projective resolution $P_* \to M$ and an injective resolution $N \to Q^*$. Having made these choices, we get a *double complex*

$$\hom_R(P_i, Q^j)$$

of a whole lot of R-modules. The claim is that in such a situation, where you have a double complex C_{ij}, you can form an ordinary chain complex C' by adding along the diagonals. Namely, the nth term is $C'_n = \bigoplus_{i+j=n} C_{ij}$. This *total complex* will receive a map from the chain complex used to compute the Ext groups and a chain complex used to compute the Ext' groups. There are maps on cohomology,

$$\mathrm{Ext}^i(M, N) \to H^i(C'_*), \quad \mathrm{Ext}^i(M, N)' \to H^i(C'_*).$$

The claim is that isomorphisms on cohomology will be induced in each case. That will prove the result, but we shall not prove the claim. $\square$

Last time we were talking about Ext groups over commutative rings. For R a commutative ring and M, N R-modules, we defined an R-module $\mathrm{Ext}^i(M, N)$ for each i, and proved various properties. We forgot to mention one.

1.5.22 Proposition *If R noetherian, and M, N are finitely generated,* $\mathrm{Ext}^i(M, N)$ *is also finitely generated.*

Proof. We can take a projective resolution P_* of M by finitely generated free modules, R being noetherian. Consequently the complex $\hom(P_*, N)$ consists of finitely generated modules. Thus the cohomology is finitely generated, and this cohomology consists of the Ext groups. $\square$

IV.2. Homological algebra à la Grothendieck

2.1. Exact categories

In this section we will introduce exact categories which allow us to define the concepts of kernels, cokernels, images and coimages crucial for doing homological algebra in a more abstract sense. The main goal will to prove the snake lemma and the five lemma in exact categories.

2.1.1 Definition Let C be a category. An object of C which is is both an initial and a terminal object is termed a *zero object* of C. It is necessarily uniquely defined up to isomorphism and will be denoted by $\mathsf{0}_{\mathsf{C}}$ or briefly by $\mathsf{0}$.

For every object A of C, the unique morphism $\mathsf{0} \to A$ will be denoted by $\mathsf{0}_{\mathsf{0}\to A}$. The unique morphism $A \to \mathsf{0}$ will be denoted by $\mathsf{0}_{A\to\mathsf{0}}$. If B is another object, we denote the composition $\mathsf{0}_{\mathsf{0}\to B}\mathsf{0}_{A\to\mathsf{0}} : A \to B$ by $\mathsf{0}_{A\to B}$, $\mathsf{0}_{AB}$ or just *zero* if no confusion can arise and call it the *zero morphism* between A and B.

2.1.2 Definition Assume that C is a category with zero object $\mathsf{0}$ and let $f : A \to B$ be a morphism in C.

(i) By a *kernel* of f one understands a morphism $i : K \to A$ such that $fi = \mathsf{0}_{KB}$ and such that for every morphism $j : L \to A$ which satisfies $fj = \mathsf{0}_{LB}$ there exists a unique morphism $l : L \to K$ making the diagram

$$\begin{tikzcd} K \arrow[r, "i"] & A \\ L \arrow[u, "l"] \arrow[ur, "j"'] & \end{tikzcd} \tag{2.1.1}$$

commute.

(ii) By a *cokernel* of f one understands a morphism $c : B \to Q$ such that $cf = \mathsf{0}_{AQ}$ and such that for every morphism $d : B \to R$ which satisfies $df = \mathsf{0}_{AR}$ there exists a unique morphism $r : Q \to R$ making the diagram

$$\begin{tikzcd} B \arrow[r, "c"] \arrow[dr, "d"'] & Q \arrow[d, "r"] \\ & R \end{tikzcd} \tag{2.1.2}$$

commute.

2.1.3 Remark By definition it is clear that the kernel of $f : A \to B$ can be identified with the equalizer $\mathrm{eq}(f, \mathsf{0}_{AB})$ and that the cokernel of $f : A \to B$ coincides with the coqualizer $\mathrm{coeq}(f, \mathsf{0}_{AB})$. Either using this observation and uniqueness (up to isomorphism) of equalizers and coequalizers or by direct argument using the universal property of the kernel respectively the cokernel one shows that kernels and cokernels are unique up to unique isomorphism. By this uniqueness property it makes sense to give the kernel and the cokernel of f each a symbol. We will from now on write $\ker(f) : \mathrm{Ker}(f) \to A$ for the kernel of f and $\mathrm{coker}(f) : B \to \mathrm{Coker}(f)$ for the cokernel of f. Note that $\mathrm{Ker}(f)$ andf $\mathrm{Coker}(f)$ are objects of the underlying category whereas $\ker(f)$ and $\mathrm{coker}(f)$ are morphisms.

2.1.4 Lemma *In a category C with zero object $\mathsf{0}$ all kernels are monomorphisms and all cokernels are epimorphisms.*

Proof. Let $f : A \to B$ be a morphism and assume that $i : K \to A$ is a kernel. Let $g, h : C \to K$ be two morphisms and assume that $ig = ih$. Denote that morphism by j. Then

$$fj = fig = fih = \mathsf{0}_{CB} \ .$$

Hence, by the universal property of the kernel there exists a unique morphism $C \to K$ making the diagram

$$\begin{tikzcd} K \arrow[r, "i"] & A \\ C \arrow[u] \arrow[ur, "j"'] & \end{tikzcd}$$

commute. But both g and h make this diagram commute, so they have to coincide. Therefore, i is a monomorphism. The argument for the cokernel is dual. $\square$

2.1.5 Still working in a category C with a zero object and all kernels and cokernels we construct two binary relations $\geqslant_{\mathrm{s}}$ and $\leqslant_{\mathrm{q}}$ between morphisms in C having the same codomain respectively the same domain.

For morphisms of the form $u : X \to Z$ and $v : Y \to Z$ we say that u *succeeds* v, in signs $u \geqslant_{\mathrm{s}} v$ if there exists a morphism $y : Y \to X$ such that the diagram

$$\begin{tikzcd} Y \arrow[d, "y"'] \arrow[dr, "v"] & \\ X \arrow[r, "u"'] & Z \end{tikzcd}$$

commutes. In other words, u to succeed v means that v factors through u. If u succeeds v and v succeeds u, we say that u and v are *s-equivalent* and write $u \equiv_{\mathrm{s}} v$.

Given two morphisms of the form $f : A \to B$ and $g : A \to C$ we say that f *precedes* g, in signs $f \leqslant_{\mathrm{q}} g$, if g factors through f, that is, if there exists a morphism $c : B \to C$ such that the diagram

$$\begin{tikzcd} & C \\ A \arrow[ur, "g"] \arrow[r, "f"'] & B \arrow[u, "c"'] \end{tikzcd}$$

commutes. If f precedes g and g precedes f, then we call f and g *q-equivalent* and write $f \equiv_{\mathrm{q}} g$.

By definition it is clear that the relations $\geqslant_{\mathrm{s}}$ and $\leqslant_{\mathrm{q}}$ are reflexive and transitive. If C is a small category, $\geqslant_{\mathrm{s}}$ and $\leqslant_{\mathrm{q}}$ are therefore preorders on the morphism sets $\mathsf{C}(-, Z)$ and $\mathsf{C}(A, -)$, respectively. Since $\geqslant_{\mathrm{s}}$ and $\leqslant_{\mathrm{q}}$ are reflexive and transitive so are s-equivalence and q-equivalence. Both are symmetric by definition, hence s-equivalence and q-equivalence are equivalence relations. In general, if f and g are q-equivalent, the *relating* morphisms b and c which fulfill the equalities $g = cf$ and $f = bg$ are neither uniquely determined nor need to be isomorphisms. An analogous observation holds for s-equivalence. The following results state conditions under which the relating morphisms for s-equivalence and for q-equivalence are isomorphisms.

2.1.6 Lemma *Let C be a category with zero object and all kernels and cokernels. Assume that A and Z are objects of C.*

(i) *Two monomorphisms $u : X \to Z$ and $v : Y \to Z$ are s-equivalent if and only there exists an isomorphism $y : Y \to X$ such that $v = uy$. In this case, the inverse $x := y^{-1}$ satisfies $u = vx$.*

(ii) *Two epimorphisms $f : A \to B$ and $g : A \to C$ are q-equivalent if and only if there exists an isomorphism $c : B \to C$ such that $g = cf$. In this case, the inverse $b := c^{-1}$ satisfies $f = bg$.*

Proof. ad (i). Assume that $u : X \to Z$ and $v : Y \to Z$ are s-equivalent. Then there exist morphisms $y : Y \to X$ and $x : X \to Y$ such that $v = uy$ and $u = vy$. Therefore,

$$u\,y\,x = u\,\mathrm{id}_X \qquad \text{and} \qquad v\,x\,y = v\,\mathrm{id}_Y \ .$$

Since both u and v are monomorphisms, the equalities $yx = \mathrm{id}_X$ and $xy = \mathrm{id}_Y$ follow, hence x and y are mutually inverse isomorphisms. The converse is clear by definition of s-equivalence.

ad (ii). The argument is dual to the one for (i). □

2.1.7 Remark Given two monomorphisms $u : X \to Z$ and $v : Y \to Z$, we will often just write $u \equiv v$ instead of $u \equiv_{\mathrm{s}} v$ to denote that u and v are s-equivalent. Similarly, for two epimorphisms $f : A \to B$ and $g : A \to C$ we usually abbreviate q-equivalence, that is $f \equiv_{\mathrm{q}} g$, by $f \equiv g$. In addition, we will often briefly say that u and v (respectively f and g) are *equivalent* instead of saying they are s-equivalent (respectively q-equivalent). By the preceding lemma, these agreements will not cause any confusion rather will they improve readability.

2.1.8 Definition Assume that C is a category with zero object 0 and that all morphisms in C have a kernel and a cokernel. Let $f : A \to B$ be a morphism.

(i) The kernel $\ker(\mathrm{coker}(f)) : \mathrm{Ker}(\mathrm{coker}(f)) \to B$ of the cokernel of f is called the *image* of f. It is denoted by $\mathrm{im}(f) : \mathrm{Im}(f) \to B$.

(ii) The cokernel $\mathrm{coker}(\ker(f)) : A \to \mathrm{Coker}(\ker(f))$ of the kernel of f is called the *coimage* of f. One denotes it by $\mathrm{coim}(f) : A \to \mathrm{Coim}(f)$.

2.1.9 Lemma *In a category C which possess a zero object 0 and all kernels and cokernels, the follow natural equivalences hold true.*

(i)

$$\ker \operatorname{coker} \ker f \equiv \ker f$$

(ii)

to be added: coker *ker* coker *case!*

question: *does one need* C *to be exact?*

2.1.10 Under the same assumptions as before that C has a zero object and possesses all kernels and cokernels consider in the following diagram a morphism $f : A \to B$ and the associated kernel, coimage, image and cokernel of f:

$$\begin{tikzcd} & A \arrow[rrr, "f"] \arrow[dr, "\mathrm{coim}(f)"'] & & & B \arrow[dr, "\mathrm{coker}(f)"] & \\ \mathrm{Ker}(f) \arrow[ur, "\ker(f)"] & & \mathrm{Coim}(f) \arrow[urr, dashed, "f'"] \arrow[r, dotted, "\overline{f}"'] & \mathrm{Im}(f) \arrow[ur, "\mathrm{im}(f)"'] & & \mathrm{Coker}(f) \end{tikzcd} \tag{2.1.3}$$

Since the composition $f \circ \ker(f)$ coincides with the zero morphism, there exists by the universal property of the coimage a unique morphism $f' : \mathrm{Coim}(f) \to B$ such that the diagram consisting of the straight and dashed arrows commutes. Since the composition

$$\mathrm{coker}(f) \circ f' \circ \mathrm{coim}(f) = \mathrm{coker}(f) \circ f$$

is the zero morphism and $\mathrm{coim}(f)$ is an epimorphism by Lemma 2.1.4, the morphism $\mathrm{coker}(f) \circ f'$ has to be the zero morphism as well. By the universal property of the image, there exists a unique morphism $\overline{f} : \mathrm{Coim}(f) \to \mathrm{Im}(f)$ making the full diagram commutate.

2.1.11 Definition A category C which has a zero object $\mathsf{0}$ and possesses all kernels and cokernels is called an *exact category* if for all morphisms $f : A \to B$ the associated unique morphism $\overline{f} : \mathrm{Coim}(f) \to \mathrm{Im}(f)$ making the diagram (2.1.3) commute is an isomorphism.

2.1.12 Definition Let

$$f_\bullet : \qquad \ldots A_{n-1} \xrightarrow{f_{n-1}} A_n \xrightarrow{f_n} A_{n+1} \ldots$$

be a sequence of objects and morphisms in the exact category C. One says that $f_\bullet$ is *exact* at A_n if the kernel of f_n is an image of f_{n-1}. If the sequence is exact at each A_n, then $f_\bullet$ is called an *exact sequence*.

From now on we always assume that the underlying category C is exact.

2.1.13 Lemma *Let*

$$A \xrightarrow{\ f\ } B \xrightarrow{\ g\ } C \tag{2.1.4}$$

be an exact sequence of objects and morphisms in C. *Then the following holds true.*

(i) *If $d : D \to B$ is a morphism such that the diagram of straight arrows in the diagram*

$$\begin{tikzcd} & & D \arrow[dl, dashed] \arrow[d] \arrow[dr, "0"] & \\ 0 \arrow[r] & A \arrow[r] & B \arrow[r] & C \end{tikzcd} \tag{2.1.5}$$

commutes and such that the horizontal sequence is exact, then there exists a unique morphism $D \to A$ making the full diagram commute.

(ii) *If $e : B \to E$ is a morphism such that the diagram of straight arrows in the diagram*

$$\begin{tikzcd} A \arrow[r] \arrow[dr, "0"'] & B \arrow[r] \arrow[d] & C \arrow[r] \arrow[dl, dashed] & 0 \\ & E & & \end{tikzcd} \tag{2.1.6}$$

commutes and such that the horizontal sequence is exact, then there exists a unique morphism $C \to E$ making the full diagram commute.

Proof. *ad* (i). Since $gd = 0$, there exists a unique morphism $\overline{d} : D \to \operatorname{Ker} g$ such that $d = \ker d \circ \overline{d}$. Since by exactness $\ker g \equiv \operatorname{im} f$, we actually have unique morphism $\overline{d} : D \to \operatorname{Im} f$ such that $d = \operatorname{im} f \circ \overline{d}$. □

2.1.14 Lemma *The sequence*

$$f_\bullet : \qquad \ldots A_{n-1} \xrightarrow{f_{n-1}} A_n \xrightarrow{f_n} A_{n+1} \ldots$$

in C is exact at A_n if and only the following conditions hold true:

(i) $f_n \circ f_{n-1} = 0_{A_{n-1} \to A_{n+1}}$.

(ii) $\operatorname{coker}(f_{n-1}) \circ \ker(f_n) = 0_{\operatorname{Ker}(f_n) \to \operatorname{Coker} f_{n-1}}$.

2.2. Additive categories

2.2.1 Definition By a *pre-additive category* one understands a category A *enriched* over the category of abelian groups. This means that for each pair of objects A, B in A the morphism set $\operatorname{Mor}(A, B)$ carries an abelian group structure

$$+_{(A,B)} : \operatorname{Mor}(A,B) \times \operatorname{Mor}(A,B) \to \operatorname{Mor}(A,B), \quad (f,g) \mapsto f + g$$

such that composition of morphisms in A is bilinear in the following sense:

(BL) If A, B, C are objects of A, $f, f' \in \operatorname{Mor}(A,B)$ and $g, g' \in \operatorname{Mor}(B,C)$, then

$$g \circ (f + f') = (g \circ f) + (g \circ f') \quad \text{and} \quad (g + g') \circ f = (g \circ f) + (g' \circ f)\,.$$

2.2.2 Usually one denotes the set of morphism between objects A and B of a pre-additive category A by $\mathrm{Hom}(A,B)$ instead of $\mathrm{Mor}(A,B)$. We will follow this conention from now on. The zero element of $\mathrm{Hom}(A,B)$ will be denoted by $0_{(A,B)}$ or briefly by 0, if no confusion can arise. In general, and as done already in the definition, we will abbreviate the group operation $+_{(A,B)}$ on $\mathrm{Hom}(A,B)$ by $+$ for clearity of exposition.

A pre-additive structure on a category imposes quite a useful relation between finite products and coproducts of its objects, namely that they have to coincide when they exist.

2.2.3 Proposition *Let A be a pre-additive category, and $A_1,\ldots,A_n$ a finite family of objects in A.*

(1) If $\prod_{l=1}^n A_l$ is a product with canonical projections $p_k : \prod_{l=1}^n A_l \to A_k$, $k=1,\ldots,n$, then it is also a coproduct where the canonical injections are given by the uniquely determined morphisms $i_k : A_k \mapsto \prod_{l=1}^n A_l$ such that

$$p_l \circ i_k = \begin{cases} \mathrm{id}_{A_k}, & \text{if } k = l, \\ 0, & \text{else.} \end{cases}$$

In addition, the equality

$$\sum_{l=1}^n i_l \circ p_l = \mathrm{id}_{\prod_{l=1}^n A_l} \tag{2.2.1}$$

holds true.

(2) If $\coprod_{l=1}^n A_l$ is a coproduct with canonical injections $i_k : \coprod_{l=1}^n A_l \to A_k$, $k=1,\ldots,n$, then it is also a product with canonical projections given by the uniquely determined morphisms $p_k : \coprod_{l=1}^n A_l \mapsto A_k$ such that

$$p_k \circ i_l = \begin{cases} \mathrm{id}_{A_k}, & \text{if } k = l, \\ 0, & \text{else.} \end{cases}$$

In addition, the equality

$$\sum_{l=1}^n i_l \circ p_l = \mathrm{id}_{\coprod_{l=1}^n A_l} \tag{2.2.2}$$

holds true.

Proof. Let us first show (1). So assume that $\prod_{l=1}^n A_l$ is a product with canonical projections p_k, and define the i_k as in (1). Then we have, for $k=1,\ldots,n$,

$$p_k \circ \Big(\sum_{l=1}^n i_l \circ p_l \Big) = \sum_{l=1}^n p_k \circ i_l \circ p_l = p_k \,.$$

By the universal property of the product, Equation (2.2.1) follows. Now let $f_k : A_k \to X$, $k=1,\ldots,n$, be a family of morphisms in A. Define $f : \prod_{l=1}^n A_l \to X$ by $f = \sum_{l=1}^n f_l \circ p_l$ and compute

$$f \circ i_k = \Big(\sum_{l=1}^n f_l \circ p_l \Big) \circ i_k = \sum_{l=1}^n f_l \circ p_l \circ i_k = f_k \,.$$

If $\tilde{f} : \prod_{l=1}^{n} A_l \to X$ is another morphism satisfying $\tilde{f} \circ i_k = f_k$ for all i, then

$$f - \tilde{f} = (f - \tilde{f}) \circ \Big(\sum_{l=1}^{n} i_l \circ p_l \Big) = \sum_{l=1}^{n} (f - \tilde{f}) \circ i_l \circ p_l =$$
$$= \sum_{l=1}^{n} (f - \tilde{f}) \circ i_l \circ p_l = \sum_{l=1}^{n} (f_l - f_l) \circ p_l = 0 \, .$$

But this entails that $\prod_{l=1}^{n} A_l$ together with the morphisms i_k fulfills the universal property of a coproduct of the family $(A_l)_{l=1}^{n}$.

One shows (2) by an analogous but dual argument. □

Since by the proposition the product and the coproduct of finitely many objects A_k, $k = 1, \ldots, n$ in a pre-additive category A coincide (up to canonical isomorphism), one denotes them by the same symbol, namely by

$$\bigoplus_{k=1}^{n} A_k ,$$

and calls the resulting object the *direct sum* of the A_k. The proposition tells also that an initial or terminal object in A has to be a zero object which we then denote by 0_{A} or 0 if no confusion can arise.

2.2.4 Definition A pre-additive category A is called *additive*, if it has the following properties:

(A0) A has a zero object.

(A1) Every finite family of objects has a product.

(A1)$^\circ$ Every finite family of objects has a coproduct.

2.2.5 Example The category Ab of abelian groups carries in a natural way the structure of an additive category. Likewise, if R is a (unital) ring, the category R-Mod of R-left modules is additive.

2.3. Abelian categories

2.3.1 Definition By an *abelian category* one understands an additive category A which fulfills the following axioms by Grothendieck:

(AB1) Every morphism has a kernel and a cokernel.

(AB2) For every morphism f the induced canonical morphism $\operatorname{coim} f \to \operatorname{im} f$ is an isomorphism.

2.3.2 Proposition *Assume that A is an abelian category, and let*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "g"'] & A \arrow[d, "r"] \\ B \arrow[r, "s"'] & Y \end{tikzcd} \tag{2.3.1}$$

be a commutative diagram in A.

(1) The diagram is cartesian if and only if the sequence

$$0 \longrightarrow X \xrightarrow{i_1 f + i_2 g} A \oplus B \xrightarrow{rp_1 - sp_2} Y \tag{2.3.2}$$

is exact.

(2) The diagram is cocartesian, if and only if

$$X \xrightarrow{i_1 f - i_2 g} A \oplus B \xrightarrow{rp_1 + sp_2} Y \longrightarrow 0 \tag{2.3.3}$$

is exact.

(3) If the diagram is cartesian, and s an epimorphism, then the diagram is even bicartesian, and f is an epimorphism, too. Moreover, one obtains in this case a commutative diagram with exact rows

$$\begin{tikzcd} 0 \arrow[r] & \ker s \arrow[r] \arrow[d, equal] & X \arrow[r, "f"] \arrow[d, "g"'] & A \arrow[r] \arrow[d, "r"] & 0 \\ 0 \arrow[r] & \ker s \arrow[r] & B \arrow[r, "s"'] & Y \arrow[r] & 0\,. \end{tikzcd} \tag{2.3.4}$$

In particular this means that the kernel of s factors through g then.

(4) If the diagram is cocartesian, and f a monomorphism, then the diagram is even bicartesian, and s is a monomorphism, too. Moreover, one obtains in this case a commutative diagram with exact rows

$$\begin{tikzcd} 0 \arrow[r] & X \arrow[r, "f"] \arrow[d, "g"'] & A \arrow[r] \arrow[d, "r"] & \operatorname{coker} f \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & B \arrow[r, "s"'] & Y \arrow[r] & \operatorname{coker} f \arrow[r] & 0. \end{tikzcd} \tag{2.3.5}$$

In particular this means that the cokernel of f factors through r then.

Proof. To prove (1), consider the sequence

$$0 \longrightarrow K \xrightarrow{k} A \oplus B \xrightarrow{rp_1 - sp_2} Y, \tag{2.3.6}$$

where k is the kernel of $rp_1 - sp_2$. Given a commutative diagram

$$\begin{tikzcd} P \arrow[r, "l"] \arrow[d, "m"'] & A \arrow[d, "r"] \\ B \arrow[r, "s"'] & Y, \end{tikzcd}$$

the morphism $P \xrightarrow{i_1 l + i_2 m} A \oplus B$ must then factor through k in a unique way. Since the diagram

$$\begin{tikzcd} K \arrow[r, "p_1 k"] \arrow[d, "p_2 k"'] & A \arrow[d, "r"] \\ B \arrow[r, "s"'] & Y \end{tikzcd}$$

commutes as well, this implies that (2.3.1) is cartesian if and only if the sequence (2.3.2) is exact.

Next let us show (3). So assume that the diagram (2.3.1) is cartesian and that s is epic. Then $rp_1 - sp_2$ must be epic as well, since $(rp_1 - sp_2)i_2 = -s$. So both sequences (2.3.2) and (2.3.3) are exact, and the diagram is bicartesian. Now assume that $hf = 0$ for some morphism h. Then $f = p_1 k$, where $k = i_1 f + i_2 g$ is monic by (1). Since $hp_1 k = 0$, the morphism hp_1 factors through the cokernel of k which is $rp_1 + sp_2$. Hence $hp_1 = h'(rp_1 + sp_2)$ for some h'. One then obtains

$$0 = hp_1 i_2 = h'(rp_1 + sp_2)i_1 = h'r \, .$$

By assumption, r is epic, hence $h' = 0$. But then $hp_1 = 0$, which entails $h = hp_1 i_i = 0$. Therefore f must be epic as well.

Now consider $l : \ker s \to B$, the kernel of s. Since $sl = 0 = r0$, and since the diagram (2.3.1) is assumed to be cartesian, there exists a unique $l' : \ker s \to X$ such that $gl' = l$ and $fl' = 0$. As a kernel, l is monic, hence so is l'. It remains to show that l' is the kernel of f. To this end assume $fj = 0$ for some morphism j. Because $sgj = rfj = 0$, gj factors through the kernel of s, hence $gj = lj' = gl'j'$, and $0 = sgj = sgl'j'$. On the other hand, $rfj = 0 = sgl'j' = rfl'j'$. By the universal property of the pullback one obtains $j = l'j'$. Since j' is monic, j' is uniquely determined by j, so l' is the kernel of f.

Statments (2) and (4) follow by dualization. □

2.4. Abeliannes of a category is a property

Introduction

One of the fundamental observations about an abelian category is that the corresponding additive structure, meaning the abelian group structures on its hom-sets, actually is uniquely determined by the underlying category and its fundamental properties. In this section, we will make this statement precise and show how to recover the additive structure, if the category satisfies certain properties.

The A-axioms

2.4.1 Given a category A we consider the following axioms:

(A0) A has a zero object.

(A1) Every finite family of objects has a product.

- (A1)° Every finite family of objects has a coproduct.
- (A2) Every morphism has a kernel.
- (A2)° Every morphism has a cokernel.
- (A3) Every monomorphism is the kernel of a morphism.
- (A3)° Every epimorphism is the cokernel of a morphism.

It is the goal of this section to prove the following fundamental result.

2.4.2 Theorem *Every abelian category* A *satisfies Axioms (A0) to (A3)°. Vice versa, if* A *is a category satisfying Axioms (A0) to (A3)°, then there exists a unique pre-additive structure on* A*, and the resulting additive category is abelian.*

IV.3. Homotopical algebra

3.1. Introduction

In this chapter, we shall introduce the formalism of *model categories*. Model categories provide an abstract setting for homotopy theory: in particular, we shall see that topological spaces form a model category. In a model category, it is possible to talk about notions such as "homotopy," and thus to pass to the homotopy category.

But many algebraic categories form model categories as well. The category of chain complexes over a ring forms one. It turns out that this observation essentially encodes classical homological algebra. We shall see, in particular, how the notion of *derived functor* can be interpreted in a model category, via this model structure on chain complexes.

Our ultimate goal in developing this theory, however, is to study the *non-abelian* case. We are interested in developing the theory of the *cotangent complex*, which is loosely speaking the derived functor of the Kähler differentials $\Omega_{S/R}$ on the category of R-algebras. This is not a functor on an additive category; however, we shall see that the non-abelian version of derived functors (in the category of *simplicial* R-algebras) allows one to construct the cotangent complex in an elegant way.

3.2. Model categories

Definition

We need to begin with the notion of a *retract* of a map.

3.2.1 Definition Let $\mathcal{C}$ be a category. Then we can form a new category Map$\mathcal{C}$ of *maps* of $\mathcal{C}$. The objects of this category are the morphisms $A \to B$ of $\mathcal{C}$, and a morphism between $A \to B$ and $C \to D$ is given by a commutative square

$$\begin{tikzcd} A \arrow[r] \arrow[d] & C \arrow[d] \\ B \arrow[r] & D \end{tikzcd} .$$

A map in $\mathcal{C}$ is a **retract** of another map in $\mathcal{C}$ if it is a retract as an object of Map$\mathcal{C}$. This means that there is a diagram:

$$\begin{tikzcd}
A \arrow[r] \arrow[rr, bend left, "Id"] \arrow[d, "f"] & B \arrow[r] \arrow[d, "g"] & A \arrow[d, "f"] \\
X \arrow[r] \arrow[rr, bend right, "Id"'] & Y \arrow[r] & X
\end{tikzcd}$$

For instance, one can prove:

3.2.2 Proposition *In any category, isomorphisms are closed under retracts.*

We leave the proof as an exercise.

3.2.3 Definition A **model category** is a category $\mathcal{C}$ equipped with three classes of maps called *cofibrations*, *fibrations*, and *weak equivalences*. They have to satisfy five axioms $M1 - M5$.

Denote cofibrations as $\hookrightarrow$, fibrations as $\twoheadrightarrow$, and weak equivalences as $\rightarrow \sim$.

(M1) $\mathcal{C}$ is closed under all limits and colimits.[1]

(M2) Each of the three classes of cofibrations, fibrations, and weak equivalences is *closed under retracts*.[2]

(M3) If *two of three* in a composition are weak equivalences, so is the third.

$$\begin{tikzcd}
{} \arrow[r, "f"] \arrow[d, "h"'] & {} \arrow[dl, "g"] \\
{} &
\end{tikzcd}$$

(M4) (*Lifts*) Suppose we have a diagram

$$\begin{tikzcd}
A \arrow[r] \arrow[d, hook, "i"] & X \arrow[d, two heads, "p"] \\
B \arrow[r] \arrow[ur, dashed] & Y
\end{tikzcd}$$

Here $i : A \to B$ is a cofibration and $p : X \to Y$ is a fibration. Then a lift exists if i or p is a weak equivalence.

[1] Many of our arguments will involve infinite colimits. The original formulation in **?** required only finite such, but most people assume infinite.

[2] Quillen initially called model categories satisfying this axiom *closed* model categories. All the model categories we consider will be closed, and we have, following **?**, omitted this axiom.

(M5) (*Factorization*) Every map can be factored in two ways:

$$\begin{tikzcd} & \cdot \arrow[dr, two heads, "\sim"] & \\ \cdot \arrow[ur, hook] \arrow[rr, "f"] \arrow[dr, hook, "\sim"'] & & \cdot \\ & \cdot \arrow[ur, two heads] & \end{tikzcd}$$

In words, it can be factored as a composite of a cofibration followed by a fibration which is a weak equivalence, or as a cofibration which is a weak equivalence followed by a fibration.

A map which is a weak equivalence and a fibration will be called an **acyclic fibration**. Denote this by $\twoheadrightarrow \sim$. A map which is both a weak equivalence and a cofibration will be called an **acyclic cofibration**, denoted $\hookrightarrow \sim$. (The word "acyclic" means for a chain complex that the homology is trivial; we shall see that this etymology is accurate when we construct a model structure on the category of chain complexes.)

3.2.4 Remark If C is a model category, then $\mathcal{C}^{op}$ is a model category, with the notions of fibrations and cofibrations reversed. So if we prove something about fibrations, we automatically know something about cofibrations.

We begin by listing a few elementary examples of model categories:

3.2.5 Example
1. Given a complete and cocomplete category $\mathcal{C}$, then we can give a model structure to $\mathcal{C}$ by taking the weak equivalences to be the isomorphisms and the cofibrations and fibrations to be all maps.
2. If R is a *Frobenius ring*, or the classes of projective and injective R-modules coincide, then the category of modules over R is a model category. The cofibrations are the injections, the fibrations are the surjections, and the weak equivalences are the *stable equivalences* (a term which we do not define). See **?**.
3. The category of topological spaces admits a model structure where the fibrations are the *Serre fibrations* and the weak equivalences are the *weak homotopy equivalences*. The cofibrations are, as we shall see, determined from this, though they can be described explicitly.

3.2.6 Remark Show that there exists a model structure on the category of sets where the injections are the cofibrations, the surjections are fibrations, and all maps are weak equivalences.

The retract argument

The axioms for a model category are somewhat complicated. We are now going to see that they are actually redundant. That is, any two of the classes of cofibrations, fibrations, and weak equivalences determine the third. We shall thus introduce a useful trick that we shall have occasion to use many times further when developing the foundations.

3.2.7 Definition Let $\mathcal{C}$ be any category. Suppose that P is a class of maps of $\mathcal{C}$. A map $f : A \to B$ has the **left lifting property** with respect to P iff: for all $p : C \to D$ in P and all diagrams

$$\begin{tikzcd} A \arrow[r] \arrow[d, "f"'] & C \arrow[d, "p"] \\ B \arrow[r] \arrow[ur, dashed, "\exists"] & D \end{tikzcd}$$

a lift represented by the dotted arrow exists, making the diagram commute. We abbreviate this property to **LLP**. There is also a notion of a **right lifting property**, abbreviated **RLP**, where f is on the right.

3.2.8 Proposition *Let P be a class of maps of $\mathcal{C}$. Then the set of maps $f : A \to B$ that have the LLP (resp. RLP) with respect to P is closed under retracts and composition.*

Proof. This will be a diagram chase. Suppose $f : A \to B$ and $g : B \to C$ have the LLP with respect to maps in P. Suppose given a diagram

$$\begin{tikzcd} A \arrow[r] \arrow[d, "g\circ f"] & X \arrow[d] \\ C \arrow[r] & Y \end{tikzcd}$$

with $X \to Y$ in P. We have to show that there exists a lift $C \to X$. We can split this into a commutative diagram:

$$\begin{tikzcd} A \arrow[r] \arrow[d, "f"] & X \arrow[dd] \\ B \arrow[ur, dashed] \arrow[d, "g"] \arrow[dr] & \\ C \arrow[r] & Y \end{tikzcd}$$

The lifting property provides a map $\phi : B \to X$ as in the dotted line in the diagram. This gives a diagram

$$\begin{tikzcd} B \arrow[r, "\phi"] \arrow[d, "g"] & X \arrow[d] \\ C \arrow[r] \arrow[ur, dashed] & Y \end{tikzcd}$$

□

and in here we can find a lift because g has the LLP with respect to p. It is easy to check that this lift is what we wanted.

The axioms of a model category imply that cofibrations have the LLP with respect to trivial fibrations, and acyclic cofibrations have the LLP with respect to fibrations. There are dual statements for fibrations. It turns out that these properties *characterize* cofibrations and fibrations (and acyclic ones).

3.2.9 Theorem *Suppose C is a model category. Then:*

(1) A map f is a cofibration iff it has the left lifting property with respect to the class of acyclic fibrations.

(2) A map is a fibration iff it has the right lifting property w.r.t. the class of acyclic cofibrations.

Proof. Suppose you have a map f, that has LLP w.r.t. all acyclic fibrations and you want it to be a cofibration. (The other direction is an axiom.) Somehow we're going to have to get it to be a retract of a cofibration. Somehow you have to use factorization. Factor f:

$$\begin{tikzcd} A \arrow[d, "f"] \arrow[dr, hook] & \\ X & X' \arrow[l, two heads, "\sim"] \end{tikzcd}$$

We had assumed that f has LLP. There is a lift:

$$\begin{tikzcd} A \arrow[r, hook, "i"] \arrow[d, "f"] & X' \arrow[d, two heads, "\sim"] \\ X \arrow[r, "Id"] \arrow[ur, dashed] & X \end{tikzcd}$$

This implies that f is a retract of i.

$$\begin{tikzcd} A \arrow[r] \arrow[d, "f"] & A \arrow[r] \arrow[d, hook, "i"] & A \arrow[d, "f"] \\ X \arrow[r, dotted, "\exists"] & X' \arrow[r] & X \end{tikzcd}$$

□

3.2.10 Theorem *(1) A map p is an acyclic fibration iff it has RLP w.r.t. cofibrations*

(2) A map is an acyclic cofibration iff it has LLP w.r.t. all fibrations.

Suppose we know the cofibrations. Then we don't know the weak equivalences, or the fibrations, but we know the maps that are both. If we know the fibrations, we know the maps that are both weak equivalences and cofibrations. This is basically the same argument. One direction is easy: if a map is an acyclic fibration, it has the lifting property by the definitions. Conversely, suppose f has RLP w.r.t. cofibrations. Factor this as a cofibration followed by an acyclic fibration.

$$\begin{tikzcd} X \arrow[r, "Id"] \arrow[d, hook] & X \arrow[d, "f"] \\ Y' \arrow[r, two heads, "p", "\sim"'] \arrow[ur, dashed] & Y \end{tikzcd}$$

f is a retract of p; it is a weak equivalence because p is a weak equivalence. It is a fibration by the previous theorem.

3.2.11 Corollary *A map is a weak equivalence iff it can be written as the product of an acyclic fibration and an acyclic cofibration.*

We can always write

$$\begin{tikzcd} & \cdot \arrow[dr, "p"] & \\ \cdot \arrow[ur, hook, "\sim"] \arrow[rr, "f"] & & \cdot \end{tikzcd}$$

By two out of three f is a weak equivalence iff p is. The class of weak equivalences is determined by the fibrations and cofibrations.

3.2.12 Example (Topological spaces) The construction here is called the Serre model structure (although it was defined by Quillen). We have to define some maps.

(1) The fibrations will be Serre fibrations.

(2) The weak equivalences will be weak homotopy equivalences

(3) The cofibrations are determined by the above classes of maps.

3.2.13 Theorem *A space equipped with these classes of maps is a model category.*

Proof. More work than you realize. M1 is not a problem. The retract axiom is also obvious. (Any class that has the lifting property also has retracts.) The third property is also obvious: *something is a weak equivalence iff when you apply some functor (homotopy), it becomes an isomorphism.* (This is important.) So we need lifting and factorization. One of the lifting axioms is also automatic, by the definition of a cofibration. Let's start with the factorizations. Introduce two classes of maps:

$$A = \{D^n \times \{0\} \to D^n \times [0,1] \mid n \geqslant 0\}$$

$$B = A \cup \{S^{n-1} \to D^n \mid n \geqslant 0, S^{-1} = \varnothing\}$$

These are compact, in a category-theory sense. By definition of Serre fibrations, a map is a fibration iff it has the right lifting property with respect to A. A map is an acyclic fibration iff it has the RLP w.r.t. B. (This was on the homework.) I need another general fact:

3.2.14 Proposition *The class of maps having the left lifting property w.r.t. a class P is closed under arbitrary coproducts, co-base change, and countable (or even transfinite) composition. By countable composition*

$$A_0 \hookrightarrow A_1 \to A_2 \to \cdots$$

we mean the map $A \to colim_n A_n$.

Suppose I have a map $f_0 : X_0 \to Y_0$. We want to produce a diagram:

$$\begin{tikzcd} X_0 \arrow[r] \arrow[dr, "f_0"'] & X_1 \arrow[d, "f_1"] \\ & Y \end{tikzcd}$$

We have $\sqcup V \to \sqcup D$ where the disjoint union is taken over commutative diagrams

$$\begin{tikzcd} V \arrow[r] \arrow[d] & X \arrow[d] \\ D \arrow[r] & Y \end{tikzcd}$$

where $V \to D$ is in A. Sometimes we call these lifting problems. For every lifting problem, we formally create a solution. This gives a diagram:

$$\begin{tikzcd}
\sqcup V \arrow[r] \arrow[d] & \sqcup D \arrow[d] \arrow[ddr] & \\
X \arrow[r] \arrow[drr] & X_1 \arrow[dr, "f_1"] & \\
& & Y
\end{tikzcd}$$

where we have subsequently made the pushout to Y. By construction, every lifting problem in X_0 can be solved in X_1.

$$\begin{tikzcd}
V \arrow[r] \arrow[d] & X_0 \arrow[r, hook, "k"] \arrow[d] & X_1 \arrow[d] \\
D \arrow[r] \arrow[ur, dashed] \arrow[urr, dotted] & Y \arrow[r] & Y
\end{tikzcd}$$

We know that every map in A is a cofibration. Also, $\sqcup V \to \sqcup D$ is a homotopy equivalence. k is an acylic cofibration because it is a weak equivalence (recall that it is a homotopy equivalence) and a cofibration.

Now we make a cone of $X_0 \to X_1 \to \cdots X_\infty$ into Y. The claim is that f is a fibration:

$$\begin{tikzcd}
X \arrow[r, hook, "\sim"] \arrow[dr] & X_\infty \arrow[d, "f"] \\
& Y
\end{tikzcd}$$

by which we mean

$$\begin{tikzcd}
V \arrow[r] \arrow[d, "\ell"'] & X_n \arrow[r] \arrow[d] & X_{n+1} \arrow[r] \arrow[d] & X_\infty \arrow[d] \\
D \arrow[r] \arrow[ur, dashed] & Y \arrow[r] & Y \arrow[r] & Y
\end{tikzcd}$$

□

where $\ell \in A$. V is compact Hausdorff. X_∞ was a colimit along closed inclusions.

So I owe you one lifting property, and the other factorization.

Part V.

Commutative Algebra, advanced topics

V.1. Flatness revisited

In the past, we have already encountered the notion of *flatness.* We shall now study it in more detail. We shall start by introducing the notion of *faithful* flatness and introduce the idea of "descent." Later, we shall consider other criteria for (normal) flatness that we have not yet explored.

We recall (definition 4.4.7) that a module M over a commutative ring R is *flat* if the functor $N \mapsto N \otimes_R M$ is an exact functor. An R-algebra is flat if it is flat as a module. For instance, we have seen that any localization of R is a flat algebra, because localization is an exact functor.

All this has not been added yet!

1.1. Faithful flatness

Faithfully flat modules

Let R be a commutative ring.

1.1.1 Definition The R-module M is **faithfully flat** if any complex $N' \to N \to N''$ of R-modules is exact if and only if the tensored sequence $N' \otimes_R M \to N \otimes_R M \to N'' \otimes_R M$ is exact.

Clearly, a faithfully flat module is flat.

1.1.2 Example The direct sum of faithfully flat modules is faithfully flat.

1.1.3 Example A (nonzero) free module is faithfully flat, because R itself is flat (tensoring with R is the identity functor).

We shall now prove several useful criteria about faithfully flat modules.

1.1.4 Proposition *An R-module M is faithfully flat if and only if it is flat and if $M \otimes_R N = 0$ implies $N = 0$ for any N.*

Proof. Suppose M faithfully flat Then M is flat, clearly. In addition, if N is any R-module, consider the sequence

$$0 \to N \to 0;$$

it is exact if and only if

$$0 \to M \otimes_R N \to 0$$

is exact. Thus $N = 0$ if and only if $M \otimes_R N = 0$.

Conversely, suppose M is flat and satisfies the additional condition. We need to show that if $N' \otimes_R M \to N \otimes_R M \to N'' \otimes_R M$ is exact, so is $N' \to N \to N''$. Since M is flat, taking homology commutes with tensoring with M. In particular, if H is the homology of $N' \to N \to N''$, then $H \otimes_R M$ is the homology of $N' \otimes_R M \to N \otimes_R M \to N'' \otimes_R M$. It follows that $H \otimes_R M = 0$, so $H = 0$, and the initial complex is exact. ☐

1.1.5 Example Another illustration of the above technique is the following observation: if M is faithfully flat and $N \to N'$ is any morphism, then $N \to N'$ is an isomorphism if and only if $M \otimes N' \to M \otimes N$ is an isomorphism. This follows because the condition that a map be an isomorphism can be phrased as the exactness of a certain (uninteresting) complex.

1.1.6 Remark (exercise) The direct sum of a flat module and a faithfully flat module is faithfully flat.

From the above result, we can get an important example of a faithfully flat algebra over a ring.

1.1.7 Example Let R be a commutative ring, and $\{f_i\}$ a finite set of elements that generate the unit ideal in R (or equivalently, the basic open sets $D(f_i) = \operatorname{Spec} R_{f_i}$ form a covering of $\operatorname{Spec} R$). Then the algebra $\prod R_{f_i}$ is faithfully flat over R (i.e., is so as a module). Indeed, as a product of localizations, it is certainly flat.

So by proposition 1.1.4, we are left with showing that if M is any R-module and $M_{f_i} = 0$ for all i, then $M = 0$. Fix $m \in M$, and consider the ideal $\operatorname{Ann}(m)$ of elements annihilating m. Since m maps to zero in each localization M_{f_i}, there is a power of f_i in $\operatorname{Ann}(m)$ for each i. This easily implies that $\operatorname{Ann}(m) = R$, so $m = 0$. (We used the fact that if the $\{f_i\}$ generate the unit ideal, so do $\left\{f_i^N\right\}$ for any $N \in \mathbb{Z}_{\geqslant 0}$.)

A functor F between two categories is said to be **faithful** if the induced map on the hom-sets $\hom(x, y) \to \hom(Fx, Fy)$ is always injective. The following result explains the use of the term "faithful."

1.1.8 Proposition *A module M is faithfully flat if and only if it is flat and the functor $N \to N \otimes_R M$ is faithful.*

Proof. Let M be flat. We need to check that M is faithfully flat if and only if the natural map

$$\hom_R(N, N') \to \hom_R(N \otimes_R M, N' \otimes_R M)$$

is injective. Suppose first M is faithfully flat and $f : N \to N'$ goes to zero $f \otimes 1_M : N \otimes_R M \to N' \otimes_R M$. We know by flatness that

$$\operatorname{im}(f) \otimes_R M = \operatorname{im}(f \otimes 1_M)$$

so that if $f \otimes 1_M = 0$, then $\operatorname{im}(f) \otimes M = 0$. Thus by faithful flatness, $\operatorname{im}(f) = 0$ by Proposition 1.1.4.

Conversely, let us suppose M flat and the functor $N \to N \otimes_R M$ faithful. Let $N \neq 0$; then $1_N \neq 0$ as maps $N \to N$. It follows that $1_N \otimes 1_M$ and $0 \otimes 1_M = 0$ are different as endomorphisms of $M \otimes_R N$. Thus $M \otimes_R N \neq 0$. By Proposition 1.1.4, we are done again. ☐

1.1.9 Example Note, however, that $\mathbb{Z} \oplus \mathbb{Z}/2$ is a $\mathbb{Z}$-module such that tensoring by it is a faithful but not exact functor.

Finally, we prove one last criterion:

1.1.10 Proposition *M is faithfully flat if and only if M is flat and $\mathfrak{m}M \neq M$ for all maximal ideals $\mathfrak{m} \subset R$.*

Proof. If M is faithfully flat, then M is flat, and $M \otimes_R R/\mathfrak{m} = M/\mathfrak{m}M \neq 0$ for all $\mathfrak{m}$ as $R/\mathfrak{m} \neq 0$, by Proposition 1.1.4. So we get one direction.

Alternatively, suppose M is flat and $M \otimes_R R/\mathfrak{m} \neq 0$ for all maximal $\mathfrak{m}$. Since every proper ideal is contained in a maximal ideal, it follows that $M \otimes_R R/I \neq 0$ for all proper ideals I. We shall use this and Proposition 1.1.4 to prove that M is faithfully flat

Let N now be any nonzero module. Then N contains a *cyclic* submodule, i.e. one isomorphic to R/I for some proper I. The injection

$$R/I \hookrightarrow N$$

becomes an injection

$$R/I \otimes_R M \hookrightarrow N \otimes_R M,$$

and since $R/I \otimes_R M \neq 0$, we find that $N \otimes_R M \neq 0$. By Proposition 1.1.4, it follows that M is faithfully flat □

1.1.11 Corollary *A nonzero finitely generated flat module over a* local *ring is faithfully flat.*

Proof. This follows from proposition 1.1.10 and Nakayama's lemma. □

A *finitely presented* flat module over a local ring is in fact free, but we do not prove this (except when the ring is noetherian, see **??**).

Proof. Indeed, let R be a local ring with maximal ideal $\mathfrak{m}$, and M a finitely generated flat R-module. Then by Nakayama's lemma, $M/\mathfrak{m}M \neq 0$, so that M must be faithfully flat. □

1.1.12 Proposition *Faithfully flat modules are closed under direct sums and tensor products.*

Proof. Exercise. □

Faithfully flat algebras

Let $\phi : R \to S$ be a morphism of rings, making S into an R-algebra.

1.1.13 Definition S is a **faithfully flat R-algebra** if it is faithfully flat as an R-module.

1.1.14 Example The map $R \to R[x]$ from a ring into its polynomial ring is always faithfully flat. This is clear.

Next, we indicate the usual "sorite" for faithfully flat morphisms:

1.1.15 Proposition *Faithfully flat morphisms are closed under composition and base change.*

That is, if $R \to S$, $S \to T$ are faithfully flat, so is $R \to T$. Similarly, if $R \to S$ is faithfully flat and R' any R-algebra, then $R' \to S \otimes_R R'$ is faithfully flat.

The reader may wish to try this proof as an exercise.

Proof. The first result follows because the composite of the two faithful and exact functors (tensoring $\otimes_R S$ and tensoring $\otimes_S T$ gives the composite $\otimes_R T$) yields a faithful and exact functor.

In the second case, let M be an R'-module. Then $M \otimes_{R'} (R' \otimes_R S)$ is canonically isomorphic to $M \otimes_R S$. From this it is clear if the functor $M \mapsto M \otimes_R S$ is faithful and exact, so is $M \mapsto M \otimes_{R'} (R' \otimes_R S)$. □

Flat maps are usually injective, but they need not be. For instance, if R is a product $R_1 \times R_2$, then the projection map $R \to R_1$ is flat. This never happens for faithfully flat maps. In particular, a quotient can never be faithfully flat.

1.1.16 Proposition *If S is a faithfully flat R-algebra, then the structure map $R \to S$ is injective.*

Proof. Indeed, let us tensor the map $R \to S$ with S, over R. We get a morphism of S-modules

$$S \to S \otimes_R S,$$

sending $s \mapsto 1 \otimes s$. This morphism has an obvious section $S \otimes_R S \to S$ sending $a \otimes b \mapsto ab$. Since it has a section, it is injective. But faithful flatness says that the original map $R \to S$ must be injective itself. □

1.1.17 Example The converse of proposition 1.1.16 definitely fails. Consider the localization $\mathbb{Z}_{(2)}$; it is a flat $\mathbb{Z}$-algebra, but not faithfully flat (for instance, tensoring with $\mathbb{Z}/3$ yields zero).

1.1.18 Remark (exercise) Suppose $\phi : R \to S$ is a flat, injective morphism of rings such that $S/\phi(R)$ is a flat R-module. Then show that ϕ is faithfully flat.

Flat morphisms need not be injective, but they are locally injective. We shall see this using:

1.1.19 Proposition *A flat local homomorphism of local rings is faithfully flat. In particular, it is injective.*

Proof. Let $\phi : R \to S$ be a local homomorphism of local rings with maximal ideals $\mathfrak{m}, \mathfrak{n}$. Then by definition $\phi(\mathfrak{m}) \subset \mathfrak{n}$. It follows that $S \neq \phi(\mathfrak{m})S$, so by Proposition 1.1.10 we win. □

The point of the above proof was, of course, the fact that the ring-homomorphism was *local.* If we just had that $\phi(\mathfrak{m})S \subsetneq S$ for every maximal ideal $\mathfrak{m} \subset R$, that would be sufficient for the argument.

1.1.20 Corollary *Let $\phi : R \to S$ be a flat morphism. Let $\mathfrak{q} \in \operatorname{Spec} S$, $\mathfrak{p} = \phi^{-1}(\mathfrak{q})$ the image in $\operatorname{Spec} R$. Then $R_{\mathfrak{p}} \to S_{\mathfrak{q}}$ is faithfully flat, hence injective.*

Proof. We only need to show that the map is flat by proposition 1.1.19. Let $M' \hookrightarrow M$ be an injection of $R_{\mathfrak{p}} \to S_{\mathfrak{q}}$-modules. Note that M', M are then R-modules as well. Then

$$M' \otimes_{R_{\mathfrak{p}}} S_{\mathfrak{q}} = (M' \otimes_R R_{\mathfrak{p}}) \otimes_{R_{\mathfrak{p}}} S_{\mathfrak{q}} = M' \otimes_R S_{\mathfrak{q}}.$$

Similarly for M. This shows that tensoring over $R_{\mathfrak{p}}$ with $S_{\mathfrak{q}}$ is the same as tensoring over R with $S_{\mathfrak{q}}$. But $S_{\mathfrak{q}}$ is flat over S, and S is flat over R, so by proposition 1.1.15, $S_{\mathfrak{q}}$ is flat over R. Thus the result is clear. □

Descent of properties under faithfully flat base change

Let S be an R-algebra. Often, things that are true about objects over R (for instance, R-modules) will remain true after base-change to S. For instance, if M is a finitely generated R-module, then $M \otimes_R S$ is a finitely generated S-module. In this section, we will show that we can conclude the *reverse* implication when S is *faithfully flat* over R.

1.1.21 Remark (exercise) Let $R \to S$ be a faithfully flat morphism of rings. If S is noetherian, so is R. The converse is false!

1.1.22 Remark (exercise) Many properties of morphisms of rings are such that if they hold after one makes a faithfully flat base change, then they hold for the original morphism. Here is a simple example. Suppose S is a faithfully flat R-algebra. Let R' be any R-algebra. Suppose $S' = S \otimes_R R'$ is finitely generated over R'. Then S is finitely generated over R.

To see that, note that R' is the colimit of its finitely generated R-subalgebras R_α. Thus S' is the colimit of the $R_\alpha \otimes_R S$, which inject into S'; finite generation implies that one of the $R_\alpha \otimes_R S \to S'$ is an isomorphism. Now use the fact that isomorphisms "descend" under faithfully flat morphisms.

In algebraic geometry, one can show that many properties of morphisms of *schemes* allow for descent under faithfully flat base-change. See **?**, volume IV-2.

Topological consequences

There are many topological consequences of faithful flatness on the Spec's. These are explored in detail in volume 4-2 of **?**. We shall only scratch the surface. The reader should bear in mind the usual intuition that flatness means that the fibers "look similar" to one other.

1.1.23 Proposition *Let $R \to S$ be a faithfully flat morphism of rings. Then the map* $\operatorname{Spec} S \to \operatorname{Spec} R$ *is surjective.*

Proof. Since $R \to S$ is injective, we may regard R as a subring of S. We shall first show that:

1.1.24 Lemma *If $I \subset R$ is any ideal, then $R \cap IS = I$.*

Proof. To see this, note that the morphism

$$R/I \to S/IS$$

is faithfully flat, since faithful flatness is preserved by base-change, and this is the base-change of $R \to S$ via $R \to R/I$. In particular, it is injective. Thus $IS \cap R = I$. ☐

Now to see surjectivity, we use a general criterion:

1.1.25 Lemma *Let $\phi : R \to S$ be a morphism of rings and suppose $\mathfrak{p} \in \operatorname{Spec} R$. Then $\mathfrak{p}$ is in the image of $\operatorname{Spec} S \to \operatorname{Spec} R$ if and only if $\phi^{-1}(\phi(\mathfrak{p})S) = \mathfrak{p}$.*

This lemma will prove the proposition.

Proof. Suppose first that $\mathfrak{p}$ is in the image of $\operatorname{Spec} S \to \operatorname{Spec} R$. In this case, there is $\mathfrak{q} \in \operatorname{Spec} S$ such that $\mathfrak{p}$ is the preimage of $\mathfrak{q}$. In particular, $\mathfrak{q} \supset \phi(\mathfrak{p})S$, so that, if we take pre-images,

$$\mathfrak{p} \supset \phi^{-1}(\phi(\mathfrak{p})S),$$

while the other inclusion is obviously true.

Conversely, suppose that $\mathfrak{p} \subset \phi^{-1}(\phi(\mathfrak{p})S)$. In this case, we know that

$$\phi(R - \mathfrak{p}) \cap \phi(\mathfrak{p})S = \varnothing.$$

Now $T = \phi(R - \mathfrak{p})$ is a multiplicatively closed subset. There is a morphism

$$R_{\mathfrak{p}} \to T^{-1}S \tag{1.1.1}$$

☐

which sends elements of $\mathfrak{p}$ into non-units, by (1.1.1) so it is a *local* homomorphism. The maximal ideal of $T^{-1}S$ pulls back to that of $R_{\mathfrak{p}}$. By the usual commutative diagrams, it follows that $\mathfrak{p}$ is the preimage of something in $\operatorname{Spec} S$. ☐

1.1.26 Remark The converse also holds. If $\phi : R \to S$ is a flat morphism of rings such that $\operatorname{Spec} S \to \operatorname{Spec} R$ is surjective, then ϕ is faithfully flat. Indeed, lemma 1.1.25 shows then that for any prime ideal $\mathfrak{p} \subset R$, $\phi(\mathfrak{p})$ fails to generate S. This is sufficient to imply that S is faithfully flat by proposition 1.1.10.

1.1.27 Remark A “slicker” argument that faithful flatness implies surjectiveness on spectra can be given as follows. Let $R \to S$ be faithfully flat. Let $\mathfrak{p} \in \operatorname{Spec} R$; we want to show that $\mathfrak{p}$ is in the image of $\operatorname{Spec} S$. Now *base change preserves faithful flatness.* So we can replace R by $R/\mathfrak{p}$, S by $S/\mathfrak{p}S$, and assume that R is a domain and $\mathfrak{p} = 0$. Indeed, the commutative diagram

$$\begin{tikzcd} \operatorname{Spec} S/\mathfrak{p}S \arrow[r] \arrow[d] & \operatorname{Spec} R/\mathfrak{p} \arrow[d] \\ \operatorname{Spec} S \arrow[r] & \operatorname{Spec} R \end{tikzcd}$$

shows that $\mathfrak{p}$ is in the image of $\operatorname{Spec} S \to \operatorname{Spec} R$ if and only if $\{0\}$ is in the image of $\operatorname{Spec} S/\mathfrak{p}S \to \operatorname{Spec} R/\mathfrak{p}$.

We can make another reduction: by localizing at $\mathfrak{p}$ (that is, $\{0\}$), we may assume that R is local and thus a field. So we have to show that if R is a field and S a faithfully flat R-algebra, then $\operatorname{Spec} S \to \operatorname{Spec} R$ is surjective. But since S is not the zero ring (by *faithful* flatness!), it is clear that S has a prime ideal and $\operatorname{Spec} S \to \operatorname{Spec} R$ is thus surjective.

In fact, one can show that the morphism $\operatorname{Spec} S \to \operatorname{Spec} R$ is actually an *identification,* that is, a quotient map. This is true more generally for faithfully flat and quasi-compact morphisms of schemes; see **?**, volume 4-2.

1.1.28 Theorem *Let $\phi : R \to S$ be a faithfully flat morphism of rings. Then* $\operatorname{Spec} S \to \operatorname{Spec} R$ *is a quotient map of topological spaces.*

In other words, a subset of $\operatorname{Spec} R$ is closed if and only if its pre-image in $\operatorname{Spec} S$ is closed.

Proof. We need to show that if $F \subset \operatorname{Spec} R$ is such that its pre-image in $\operatorname{Spec} S$ is closed, then F itself is closed. **ADD THIS PROOF** □

1.2. Faithfully flat descent

Fix a ring R, and let S be an R-algebra. Then there is a natural functor from R-modules to S-modules sending $N \mapsto S \otimes_R N$. In this section, we shall be interested in going in the opposite direction, or in characterizing the image of this functor. Namely, given an S-module, we want to "descend" to an R-module when possible; given a morphism of S-modules, we want to know when it comes from a morphism of R-modules by base change.

To be added: this entire section!

The Amitsur complex

To be added: citation needed

Suppose B is an A-algebra. Then we can construct a complex of A-modules

$$0 \to A \to B \to B \otimes_A B \to B \otimes_A B \otimes_A B \to \dots$$

as follows. For each n, we denote by $B^{\otimes n}$ the tensor product of B with itself n times (over A). There are morphisms of A-algebras

$$d_i : B^{\otimes n} \to B^{\otimes n+1}, \quad 0 \leqslant i \leqslant n+1$$

where the map sends

$$b_1 \otimes \cdots \otimes b_n \mapsto b_1 \otimes \cdots \otimes b_{i-1} \otimes 1 \otimes b_i \otimes \cdots \otimes b_n,$$

so that the 1 is placed in the ith spot. Then the coboundary $\partial : B^{\otimes n} \to B^{\otimes n+1}$ is defined as $\sum (-1)^i d_i$. It is easy to check that this forms a complex of A-modules.

1.2.1 Definition The above complex of B-modules is called the **Amitsur complex** of B over A, and we denote it $\mathcal{A}_{B/A}$. It is clearly functorial in B; a map of A-algebras $B \to C$ induces a morphism of complexes $\mathcal{A}_{B/A} \to \mathcal{A}_{C/A}$.

Note that the Amitsur complex behaves very nicely with respect to base-change. If A' is an A-algebra and $B' = B \otimes_A A'$ is the base extension, then $\mathcal{A}_{B'/A'} = \mathcal{A}_{B/A} \otimes_A A'$, which follows easily from the fact that base-change commutes with tensor products.

In general, the Amitsur complex is not even exact. For instance, if it is exact in degree one, then the map $A \to B$ is necessarily injective. If, however, the morphism is *faithfully flat*, then we do get exactness:

1.2.2 Theorem *If B is a faithfully flat A-algebra, then the Amitsur complex of B/A is exact. In fact, if M is any A-module, then $\mathcal{A}_{B/A} \otimes_A M$ is exact.*

Proof. We prove this first under the assumption that $A \to B$ has a section. In this case, we will even have:

1.2.3 Lemma *Suppose $A \to B$ is a morphism of rings with a section $B \to A$. Then the Amitsur complex $\mathcal{A}_{B/A}$ is homotopically trivial. (In particular, $\mathcal{A}_{B/A} \otimes_A M$ is acyclic for all M.)*

Proof. Let $s : B \to A$ be the section; by assumption, this is a morphism of A-algebras. We shall define a chain contraction of $\mathcal{A}_{B/A}$. To do this, we must define a collection of morphisms of A-modules $h_{n+1} : B^{\otimes n+1} \to B^{\otimes n}$, and this we do by sending

$$b_1 \otimes \cdots \otimes b_{n+1} \mapsto s(b_{n+1})\,(b_1 \otimes \cdots \otimes b_n)\,.$$

It is still necessary to check that the $\{h_{n+1}\}$ form a chain contraction; in other words, that $\partial h_n + h_{n+1}\partial = 1_{B^{\otimes n}}$. By linearity, we need only check this on elements of the form $b_1 \otimes \cdots \otimes b_n$. Then we find

$$\partial h_n(b_1 \otimes b_n) = s(b_1) \sum (-1)^i b_2 \otimes \cdots \otimes 1 \otimes \cdots \otimes b_n$$

where the 1 is in the ith place, while

$$h_{n+1}\partial(b_1 \otimes \cdots \otimes b_n) = b_1 \otimes \cdots \otimes b_n + \sum_{i>0} s(b_1)(-1)^{i-1} b_2 \otimes \cdots \otimes 1 \otimes \cdots \otimes b_n$$

where again the 1 is in the ith place. The assertion is from this clear. Note that if $\mathcal{A}_{B/A}$ is contractible, we can tensor the chain homotopy with M to see that $\mathcal{A}_{B/A} \otimes_A M$ is chain contractible for any M. ∎

With this lemma proved, we see that the Amitsur complex $\mathcal{A}_{B/A}$ (or even $\mathcal{A}_{B/A} \otimes_A M$) is acyclic whenever B/A admits a section. Now if we make the base-change by the morphism $A \to B$, we get the morphism $B \to B \otimes_A B$. That is,

$$B \otimes_A \left(\mathcal{A}_{B/A} \otimes_A M\right) = \mathcal{A}_{B \otimes_A B/B} \otimes_B (M \otimes_A B).$$

The latter is acyclic because $B \to B \otimes_A B$ admits a section (namely, $b_1 \otimes b_2 \mapsto b_1 b_2$). So the complex $\mathcal{A}_{B/A} \otimes_A M$ becomes acyclic after base-changing to B; this, however, is a faithfully flat base-extension, so the original complex was itself exact. ∎

1.2.4 Remark A powerful use of the Amitsur complex in algebraic geometry is to show that the cohomology of a quasi-coherent sheaf on an affine scheme is trivial. In this case, the Cech complex (of a suitable covering) turns out to be precisely the Amitsur complex (with the faithfully flat morphism $A \to \prod A_{f_i}$ for the $\{f_i\}$ a family generating the unit ideal). This argument generalizes to showing that the *étale* cohomology of a quasi-coherent sheaf on an affine is trivial; cf. **?**.

Descent for modules

Let $A \to B$ be a faithfully flat morphism of rings. Given an A-module M, we have a natural way of getting a B-module $M_B = M \otimes_A B$. We want to describe the image of this functor; alternatively, given a B-module, we want to describe the image of this functor.

Given an A-module M and the associated B-module $M_B = M \otimes_A B$, there are two ways of getting $B \otimes_A B$-modules from M_B, namely the two tensor products $M_B \otimes_B (B \otimes_A B)$ according as we pick the first map $b \mapsto b \otimes 1$ from $B \to B \otimes_A B$ or the second $b \mapsto 1 \otimes b$. We shall denote these by $M_B \otimes_A B$ and $B \otimes_A M_B$ with the action clear. But these are naturally isomorphic because both are obtained from M by base-extension $A \rightrightarrows B \otimes_A B$, and the two maps are the same. Alternatively, these two tensor products are $M \otimes_A B \otimes_A B$ and $B \otimes_A M \otimes_A B$ and these are clearly isomorphic by the braiding isomorphism[1] of the first two factors as $B \otimes_A B$-modules (with the $B \otimes_A B$ part acting on the B's in the above tensor product!).

1.2.5 Definition The **category of descent data** for the faithfully flat extension $A \to B$ is defined as follows. An object in this category consists of the following data:

1. A B-module N.
2. An isomorphism of $B \otimes_A B$-modules $\phi : N \otimes_A B \simeq B \otimes_A N$. This isomorphism is required to make the following diagram[2] of $B \otimes_A B \otimes_A B$-modules commutative:

$$\begin{tikzcd} B \otimes_A B \otimes_A N \arrow[rr, "\phi_{23}"] \arrow[dr, "\phi_{13}"'] & & B \otimes_A N \otimes_A B \arrow[dl, "\phi_{12}"] \\ & N \otimes_A B \otimes_A B & \end{tikzcd} \tag{1.2.1}$$

Here ϕ_{ij} means that the permutation of the ith and jth factors of the tensor product is done using the isomorphism ϕ.

A morphism between objects $(N, \phi), (N', \psi)$ is a morphism of B-modules $f : N \to N'$ that makes the diagram

$$\begin{tikzcd} N \otimes_A B \arrow[r, "\phi"] \arrow[d, "f \otimes 1"] & B \otimes_A N \arrow[d, "1 \otimes f"] \\ N' \otimes_A B \arrow[r, "\psi"] & B \otimes_A N' \end{tikzcd} \tag{1.2.2}$$

[1] It is *not* the braiding isomorphism $M_B \otimes_A B \simeq B \otimes_A M_B$, which is not an isomorphism of $B \otimes_A B$-modules. This is the isomorphism that sends $m \otimes b \otimes b'$ to $b \otimes m \otimes b'$.

[2] This is the cocycle condition.

As we have seen, there is a functor F from A-modules to descent data. Strictly speaking, we should check the commutativity of (1.2.1), but this is clear: for $N = M \otimes_A B$, (1.2.1) looks like

$$\begin{tikzcd} B \otimes_A B \otimes_A M \otimes_A B \arrow[rr, "\phi_{23}"] \arrow[dr, "\phi_{13}"'] & & B \otimes_A M \otimes_A B \otimes_A B \arrow[dl, "\phi_{12}"] \\ & M \otimes_A B \otimes_A B \otimes_A B & \end{tikzcd}$$

Here all the maps are just permutations of the factors (that is, the braiding isomorphisms in the structure of symmetric tensor category on the category of A-modules), so it clearly commutes.

The main theorem is:

1.2.6 Theorem (Descent for modules) *The above functor from A-modules to descent data for $A \to B$ is an equivalence of categories.*

We follow **?** in the proof.

Proof. We start by describing the inverse functor from descent data to A-modules. Recall that if M is an A-module, then M can be characterized as the submodule of M_B consisting of $m \in M_B$ such that $1 \otimes m$ and $m \otimes 1$ corresponded to the same thing in $M_B \otimes_A B \simeq B \otimes_A M_B$. (The case $M = A$ was particularly transparent: elements of A were elements $x \in B$ such that $x \otimes 1 = 1 \otimes x$ in $B \otimes_A B$.) In other words, we had the exact sequence

$$0 \to M \to M_B \to M_B \otimes_A B.$$

We want to imitate this for descent data. Namely, we want to construct a functor G from descent data to A-modules. Given descent data (N, ϕ) where $\phi : N \otimes_A B \simeq B \otimes_A N$ is an isomorphism of $B \otimes_A B$-modules, we define GN to be

$$GN = \ker(N \overset{n \mapsto 1 \otimes n - \psi(n \otimes 1)}{\to} B \otimes_A N).$$

It is clear that this is an A-module, and that it is functorial in the descent data. We have also shown that $GF(M)$ is naturally isomorphic to M for any A-module M.

We need to show the analog for $FG(N, \phi)$; in other words, we need to show that any descent data arises via the F-construction. Even before that, we need to describe a natural transformation from $FG(N, \phi)$ to the identity. Fix a descent data (N, ϕ). Then $G(N, \phi)$ gives an A-submodule $M \subset N$. We get a morphism

$$f : M_B = M \otimes_A B \to N$$

by the universal property. This sends $m \otimes b \mapsto bm$. The claim is that this is a map of descent data. In other words, we have to show that (1.2.2) commutes. The diagram looks like

$$\begin{tikzcd} M_B \otimes_A B \arrow[r] \arrow[d, "f \otimes 1"] & B \otimes_A M_B \arrow[d, "1 \otimes f"] \\ N \otimes_A B \arrow[r, "\phi"] & B \otimes_A N \end{tikzcd}.$$

In other words, if $m \otimes b \in M_B$ and $b' \in B$, we have to show that $\phi(bm \otimes b') = (1 \otimes f)(b \otimes m \otimes b') = b \otimes b'm$.

However,

$$\phi(bm \otimes b') = (b \otimes b')\phi(m \otimes 1) = (b \otimes b')(1 \otimes m) = b \otimes b'm$$

in view of the definition of $M = GN$ as the set of elements such that $\phi(m \otimes 1) = 1 \otimes m$, and the fact that ϕ is an isomorphism of $B \otimes_A B$-modules. The equality we wanted to prove is thus clear.

So we have the two natural transformations between FG, GF and the respective identity functors. We have already shown that one of them is an isomorphism. Now we need to show that if (N, ϕ) is descent data as above, and $M = G(N, \phi)$, the map $F(M) \to (N, \phi)$ is an *isomorphism*. In other words, we have to show that the map

$$M \otimes_A B \to N$$

is an isomorphism.

Here we shall draw a commutative diagram. Namely, we shall essentially use the Amitsur complex for the faithfully flat map $B \to B \otimes_A B$. We shall obtain a commutative an exact diagram:

$$\begin{tikzcd}
0 \arrow[r] & M \otimes_A B \arrow[r] \arrow[d] & N \otimes_A B \arrow[r] \arrow[d, "\phi"] & N \otimes_A B \otimes_A B \arrow[d, "\phi_{13}^{-1}"] \\
0 \arrow[r] & N \arrow[r] & B \otimes_A N \arrow[r] & B \otimes_A B \otimes_A N
\end{tikzcd}.$$

Here the map

$$N \otimes_A B \to N \otimes_A B \otimes_A B$$

sends $n \otimes b \mapsto n \otimes 1 \otimes b - \phi(1 \otimes n) \otimes b$. Consequently the first row is exact, B being flat over A. The bottom map

$$B \otimes_A N \to B \otimes_A N \otimes_A N$$

sends $b \otimes n \mapsto b \otimes 1 \otimes n - 1 \otimes b \otimes n$. It follows by the Amitsur complex that the bottom row is exact too. We need to check that the diagram commutes. Since the two vertical maps on the right are isomorphisms, it will follow that $M \otimes_A B \to N$ is an isomorphism, and we shall be done.

Fix $n \otimes b \in N \otimes_A B$. We need to figure out where it goes in $B \otimes_A B \otimes_A N$ under the two maps. Going right gives $n \otimes 1 \otimes b - \phi_{12}(1 \otimes n \otimes b)$. Going down then gives $\phi_{13}^{-1}(n \otimes 1 \otimes b) - \phi_{13}^{-1}\phi_{12}(1 \otimes n \otimes b) = \phi_{13}^{-1}(n \otimes 1 \otimes b) - \phi_{23}^{-1}(1 \otimes n \otimes b)$, where we have used the cocycle condition. So this is one of the maps $N \otimes_A B \to B \otimes_A B \otimes_A N$.

Now we consider the other way $n \otimes b$ can map to $B \otimes_A B \otimes_A N$.

Going down gives $\phi(n \otimes b)$, and then going right gives the difference of two maps $N \otimes_A B \to B \otimes_A B \otimes_A N$, which are the same as above. □

Example: Galois descent

To be added: this section

1.3. The Tor functor

Introduction

Fix M. The functor $N \mapsto N \otimes_R M$ is a right-exact functor on the category of R-modules. We can thus consider its *left-derived functors* as in **??**. Recall:

1.3.1 Definition The derived functors of the tensor product functor $N \mapsto N \otimes_R M$ are denoted by $\mathrm{Tor}^i_R(N, M), i \geqslant 0$. We shall sometimes denote omit the subscript R.

So in particular, $\mathrm{Tor}^0_R(M, N) = M \otimes N$. A priori, Tor is only a functor of the first variable, but in fact, it is not hard to see that Tor is a covariant functor of two variables M, N. In fact, $\mathrm{Tor}^i_R(M, N) \simeq \mathrm{Tor}^i_R(N, M)$ for any two R-modules M, N. For proofs, we refer to **??**. **ADD: THEY ARE NOT IN THAT CHAPTER YET.**

Let us recall the basic properties of Tor that follow from general facts about derived functors. Given an exact sequence

$$0 \to N' \to N \to N'' \to 0$$

we have a long exact sequence

$$\mathrm{Tor}^i(N', M) \to \mathrm{Tor}^i(N, M) \to \mathrm{Tor}^i(N'', M) \to \mathrm{Tor}^{i-1}(N', M) \to \dots$$

Since Tor is symmetric, we can similarly get a long exact sequence if we are given a short exact sequence of M's.

Recall, moreover, that Tor can be computed explicitly (in theory). If we have modules M, N, and a projective resolution $P_* \to N$, then $\mathrm{Tor}^i_R(M, N)$ is the ith homology of the complex $M \otimes P_*$. We can use this to compute Tor in the case of abelian groups.

1.3.2 Example We compute $\mathrm{Tor}^*_{\mathbb{Z}}(A, B)$ whenever A, B are abelian groups and B is finitely generated. This immediately reduces to the case of B either $\mathbb{Z}$ or $\mathbb{Z}/d\mathbb{Z}$ for some d by the structure theorem. When $B = \mathbb{Z}$, there is nothing to compute (derived functors are not very interesting on projective objects!). Let us compute $\mathrm{Tor}^*_{\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z})$ for an abelian group A.

Actually, let us be more general and consider the case where the ring is replaced by $\mathbb{Z}/m$ for some m such that $d \mid m$. Then we will compute $\mathrm{Tor}^*_{\mathbb{Z}/m}(A, \mathbb{Z}/d)$ for any $\mathbb{Z}/m$-module A. The case $m = 0$ will handle the ring $\mathbb{Z}$. Consider the projective resolution

$$\cdots \xrightarrow{m/d} \mathbb{Z}/m\mathbb{Z} \xrightarrow{d} \mathbb{Z}/m\mathbb{Z} \xrightarrow{m/d} \mathbb{Z}/m\mathbb{Z} \xrightarrow{d} \mathbb{Z}/m\mathbb{Z} \longrightarrow \mathbb{Z}/d\mathbb{Z} \longrightarrow 0.$$

We apply $A \otimes_{\mathbb{Z}/m\mathbb{Z}} \cdot$. Since tensoring (over $\mathbb{Z}/m$!) with $\mathbb{Z}/m\mathbb{Z}$ does nothing, we obtain the complex

$$\cdots \xrightarrow{m/d} A \xrightarrow{d} A \xrightarrow{m/d} A \xrightarrow{d} A \longrightarrow 0.$$

The groups $\mathrm{Tor}_n^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z})$ are simply the homology groups (ker/im) of the complex, which are simply

$$\begin{aligned}
\mathrm{Tor}_0^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong A/dA \\
\mathrm{Tor}_n^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong {}_dA/(m/d)A \quad n \text{ odd, } n \geqslant 1 \\
\mathrm{Tor}_n^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong {}_{m/d}A/dA \quad n \text{ even, } n \geqslant 2,
\end{aligned}$$

where ${}_kA = \{a \in A \mid ka = 0\}$ denotes the set of elements of A killed by k.

The symmetry of the tensor product also provides with a simple proof that Tor commutes with filtered colimits.

1.3.3 Proposition *Let M be an R-module, $\{N_i\}$ a filtered system of R-modules. Then the natural morphism*

$$\varinjlim_i \mathrm{Tor}_R^i(M, N_i) \to \mathrm{Tor}_R^i(M, \varinjlim_i N_i)$$

is an isomorphism.

Proof. We can see this explicitly. Let us compute the Tor functors by choosing a projective resolution $P_* \to M$ of M (note that which factor we use is irrelevant, by symmetry!). Then the left side is the colimit $\varinjlim H(P_* \otimes N_i)$, while the right side is $H(P_* \otimes \varinjlim N_i)$. But tensor products commute with filtered (or arbitrary) colimits, since the tensor product admits a right adjoint. Moreover, we know that homology commutes with filtered colimits. Thus the natural map is an isomorphism. ☐

Tor **and flatness**

Tor provides a simple way of detecting flatness. Indeed, one of the basic applications of this is that for a flat module M, the tor-functors vanish for $i \geqslant 1$ (whatever be N). Indeed, recall that $\mathrm{Tor}(M, N)$ is computed by taking a projective resolution of N,

$$\cdots \to P_2 \to P_1 \to P_0 \to M \to 0$$

tensoring with M, and taking the homology. But tensoring with M is exact if we have flatness, so the higher Tor modules vanish.

The converse is also true. In fact, something even stronger holds:

1.3.4 Proposition *M is flat iff* $\mathrm{Tor}^1(M, R/I) = 0$ *for all finitely generated ideals $I \subset R$.*

Proof. We have just seen one direction. Conversely, suppose $\mathrm{Tor}^i(M, R/I) = 0$ for all finitely generated ideals I and $i > 0$. Then the result holds, first of all, for all ideals I, because of proposition 1.3.3 and the fact that R/I is always the colimit of R/J as J ranges over finitely generated ideals $J \subset I$.

We now show that $\mathrm{Tor}^i(M, N) = 0$ whenever N is finitely generated. To do this, we induct on the number of generators of N. When N has one generator, it is cyclic and we are done. Suppose we have proved the result whenever for modules that have $n-1$ generators or less, and suppose N has n generators. Then we can consider an exact sequence of the form

$$0 \to N' \hookrightarrow N \twoheadrightarrow N'' \to 0$$

where N' has $n-1$ generators and N'' is cyclic. Then the long exact sequence shows that $\mathrm{Tor}^i(M, N) = 0$ for all $i \geqslant 1$.

Thus we see that $\mathrm{Tor}^i(M, N) = 0$ whenever N is finitely generated. Since any module is a filtered colimit of finitely generated ones, we are done by proposition 1.3.3. □

Note that there is an exact sequence $0 \to I \to R \to R/I \to 0$ and so

$$\mathrm{Tor}_1(M, R) = 0 \to \mathrm{Tor}_1(M, R/I) \to I \otimes M \to M$$

is exact, and by this:

1.3.5 Corollary *If the map*

$$I \otimes M \to M$$

is injective for all ideals I, then M is flat.

1.4. Flatness over noetherian rings

We shall be able to obtain simpler criterion for flatness when the ring in question is noetherian local. For instance, we have already seen:

1.4.1 Theorem *If M is a finitely generated module over a noetherian local ring R (with residue field k), then M is free if and only if* $\mathrm{Tor}_1(k, M) = 0$.

In particular, flatness is the same thing as the vanishing of *one* Tor module, and it equates to freeness. Now, we want to generalize this result to the case where M is not necessarily finitely generated over R, but finitely generated over an R-algebra that is also noetherian local. In particular, we shall get useful criteria for when an extension of noetherian local *rings* (which in general is not finite, or even finitely generated) is flat.

We shall prove two main criteria. The *local criterion* is a direct generalization of the above result (the vanishing of one Tor group). The *infinitesimal criterion* reduces checking flatness of M to checking flatness of $M \otimes_R R/\mathfrak{m}^t$ over $R/\mathfrak{m}^t$; in particular, it reduces to the case where the base ring is *artinian.* Armed with these, we will be able to prove a rather difficult theorem that states that we can always find lots of flat extensions of noetherian local rings.

Flatness over a noetherian local ring

We shall place ourselves in the following situation. R, S are noetherian local rings with maximal ideals $\mathfrak{m} \subset R, \mathfrak{n} \subset S$, and S is an R-algebra (and the morphism $R \to S$ is *local*, so $\mathfrak{m}S \subset \mathfrak{n}$). We will want to know when a S-module is flat over R. In particular, we want a criterion for when S is flat over R.

1.4.2 Theorem *The finitely generated S-module M is flat over R iff*

$$\mathrm{Tor}^1_R(k, M) = 0.$$

In this case, M is even free.

It is actually striking how little the condition that M is a finitely generated S-module enters, or how irrelevant it seems in the statement. The argument will, however, use the fact that M is *separated* with respect to the $\mathfrak{m}$-adic topology, which relies on Krull's intersection theorem (note that since $\mathfrak{m}S \subset \mathfrak{n}$, the $\mathfrak{m}$-adic topology on M is separated).

Proof. Necessity is immediate. What we have to prove is sufficiency.

First, we claim that if N is an R-module of finite length, then

$$\mathrm{Tor}^1_R(N, M) = 0. \tag{1.4.1}$$

This is because N has by dévissage (proposition 2.2.12) a finite filtration N_i whose quotients are of the form $R/\mathfrak{p}$ for $\mathfrak{p}$ prime and (by finite length hypothesis) $\mathfrak{p} = \mathfrak{m}$. So we have a filtration on M whose successive quotients are isomorphic to k. We can then climb up the filtration to argue that $\mathrm{Tor}^1(N_i, M) = 0$ for each i.

Indeed, the claim (1.4.1) is true $N_0 = 0 \subset N$ trivially. We climb up the filtration piece by piece inductively; if $\mathrm{Tor}^1_R(N_i, M) = 0$, then the exact sequence

$$0 \to N_i \to N_{i+1} \to k \to 0$$

yields an exact sequence

$$\mathrm{Tor}^1_R(N_i, M) \to \mathrm{Tor}^1_R(N_{i+1}, M) \to 0$$

from the long exact sequence of Tor and the hypothesis on M. The claim is proved.

Now we want to prove that M is flat. The idea is to show that $I \otimes_R M \to M$ is injective for any ideal $I \subset R$. We will use some diagram chasing and the Krull intersection theorem on the kernel K of this map, to interpolate between it and various quotients by powers of $\mathfrak{m}$. First we write some exact sequences.

We have an exact sequence

$$0 \to \mathfrak{m}^t \cap I \to I \to I/I \cap \mathfrak{m}^t \to 0$$

which we tensor with M:

$$\mathfrak{m}^t \cap I \otimes M \to I \otimes M \to I/I \cap \mathfrak{m}^t \otimes M \to 0.$$

The sequence

$$0 \to I/I \cap \mathfrak{m}^t \to R/\mathfrak{m}^t \to R/(I + \mathfrak{m}^t) \to 0$$

is also exact, and tensoring with M yields an exact sequence:

$$0 \to I/I \cap \mathfrak{m}^t \otimes M \to M/\mathfrak{m}^t M \to M/(\mathfrak{m}^t + I)M \to 0$$

because $\mathrm{Tor}^1_R(M, R/(I + \mathfrak{m}^t)) = 0$ by (1.4.1), as $R/(I + \mathfrak{m}^t)$ is of finite length.

Let us draw the following commutative diagram:

$$\begin{tikzcd} & & 0 \arrow[d] \\ \mathfrak{m}^t \cap I \otimes M \arrow[r] & I \otimes M \arrow[r] & I/I \cap \mathfrak{m}^t \otimes M \arrow[d] \\ & & M/\mathfrak{m}^t M \end{tikzcd} \qquad (1.4.2)$$

□

Here the column and the row are exact. As a result, if an element in $I \otimes M$ goes to zero in M (a fortiori in $M/\mathfrak{m}^t M$) it must come from $\mathfrak{m}^t \cap I \otimes M$ for all t. Thus, by the Artin-Rees lemma, it belongs to $\mathfrak{m}^t(I \otimes M)$ for all t, and the Krull intersection theorem (applied to S, since $\mathfrak{m}S \subset \mathfrak{n}$) implies it is zero.

The infinitesimal criterion for flatness

1.4.3 Theorem *Let R be a noetherian local ring, S a noetherian local R-algebra. Let M be a finitely generated module over S. Then M is flat over R iff $M/\mathfrak{m}^t M$ is flat over $R/\mathfrak{m}^t$ for all $t > 0$.*

Proof. One direction is easy, because flatness is preserved under base-change $R \to R/\mathfrak{m}^t$. For the other direction, suppose $M/\mathfrak{m}^t M$ is flat over $R/\mathfrak{m}^t$ for all t. Then, we need to show that if $I \subset R$ is any ideal, then the map $I \otimes_R M \to M$ is injective. We shall argue that the kernel is zero using the Krull intersection theorem.

Fix $t \in \mathbb{N}$. As before, the short exact sequence of $R/\mathfrak{m}^t$-modules $0 \to I/(\mathfrak{m}^t \cap I) \cap R/\mathfrak{m}^t \to R/(\mathfrak{m}^t \cap I) \to 0$ gives an exact sequence (because $M/\mathfrak{m}^t M$ is $R/\mathfrak{m}^t$-flat)

$$0 \to I/I \cap \mathfrak{m}^t \otimes M \to M/\mathfrak{m}^t M \to M/(\mathfrak{m}^t + I)M \to 0$$

which we can fit into a diagram, as in (1.4.2)

$$\begin{tikzcd} & & 0 \arrow[d] \\ \mathfrak{m}^t \cap I \otimes M \arrow[r] & I \otimes M \arrow[r] & I/I \cap \mathfrak{m}^t \otimes M \arrow[d] \\ & & M/\mathfrak{m}^t M \end{tikzcd} .$$

The horizontal sequence was always exact, as before. The vertical sequence can be argued to be exact by tensoring the exact sequence

$$0 \to I/I \cap \mathfrak{m}^t \to R/\mathfrak{m}^t \to R/(I + \mathfrak{m}^t) \to 0$$

of $R/\mathfrak{m}^t$-modules with $M/\mathfrak{m}^t M$, and using flatness of $M/\mathfrak{m}^t M$ over $R/\mathfrak{m}^t$ (and **??**). Thus we get flatness of M as before. ☐

Incidentally, if we combine the local and infinitesimal criteria for flatness, we get a little more.

1.4.4 Remark (comment) The gr criterion for flatness

Suppose $(R, \mathfrak{m})$ is a noetherian local ring and $(S, \mathfrak{n})$ a local R-algebra. As usual, we are interested in criteria for when a finitely generated S-module M is flat over R.

We can, of course, endow M with the $\mathfrak{m}$-adic topology. Then M is a filtered module over the filtered ring R (with the $\mathfrak{m}$-adic topology). We have morphisms for each i,

$$\mathfrak{m}^i/\mathfrak{m}^{i+1} \otimes_{R/\mathfrak{m}} M/\mathfrak{m}M \to \mathfrak{m}^i M/\mathfrak{m}^{i+1} M$$

that induce map

$$\mathrm{gr}(R) \otimes_{R/\mathfrak{m}} M/\mathfrak{m}M \to \mathrm{gr}(M).$$

If M is flat over

Generalizations of the local and infinitesimal criteria

In the previous subsecs, we obtained results that gave criteria for when, given a local homomorphism of noetherian local rings $(R, \mathfrak{m}) \to (S, \mathfrak{n})$, a finitely generated S-module was R-flat. These criteria generally were related to the Tor groups of the module with respect to $R/\mathfrak{m}$. We are now interested in generalizing the above results to the setting where $\mathfrak{m}$ is replaced by an ideal that *maps into the Jacobson radical of S*. In other words,

$$\phi : R \to S$$

will be a homomorphism of noetherian rings, and $J \subset R$ will be an ideal such that $\phi(J)$ is contained in every maximal ideal of S.

Ideally, we are aiming for results of the following type:

1.4.5 Theorem (Generalized local criterion for flatness) *Let $\phi : R \to S$ be a morphism of noetherian rings, $J \subset R$ an ideal with $\phi(J)$ contained in the Jacobson radical of S. Let M be a finitely generated S-module. Then M is R-flat if and only if M/JM is R/J-flat and* $\mathrm{Tor}_1^R(R/J, M) = 0$.

Note that this is a generalization of theorem 1.4.2. In that case, R/J was a field and the R/J-flatness of M/JM was automatic. One key step in the proof of theorem 1.4.2 was to go from the hypothesis that $\mathrm{Tor}_1(M,k)=0$ to $\mathrm{Tor}_1(M,N)=0$ whenever N was an R-module of *finite length.* We now want to do the same in this generalized case; the analogy would be that, under the hypotheses of theorem 1.4.5, we would like to conclude that $\mathrm{Tor}_1^R(M,N)=0$ whenever N is a finitely generated R-module *annihilated by* I. This is not quite as obvious because we cannot generally find a filtration on N whose successive quotients are R/J (unlike in the case where J was maximal). Therefore we shall need two lemmas.

1.4.6 Remark One situation where the strong form of the local criterion, theorem 1.4.5, is used is in Grothendieck's proof (cf. EGA IV-11, **?**) that the locus of points where a coherent sheaf is flat is open (in commutative algebra language, if A is noetherian and M finitely generated over a finitely generated A-algebra B, then the set of primes $\mathfrak{q}\in\operatorname{Spec}B$ such that $M_{\mathfrak{q}}$ is A-flat is open in $\operatorname{Spec}B$).

1.4.7 Lemma (Serre) *Suppose R is a ring, S an R-algebra, and M an S-module. Then the following are equivalent:*

1. *$M\otimes_R S$ is S-flat and* $\mathrm{Tor}_1^R(M,S)=0$.
2. $\mathrm{Tor}_1^R(M,N)=0$ *whenever N is any S-module.*

We follow **?**.

Proof. Let P be an S-module (considered as fixed), and Q any (variable) R-module. Recall that there is a homology spectral sequence

$$\mathrm{Tor}_p^S(\mathrm{Tor}_q^R(Q,S),P)\implies\mathrm{Tor}_{p+q}^R(Q,P).$$

Recall that this is the Grothendieck spectral sequence of the composite functors

$$Q\mapsto Q\otimes_R S,\quad Q'\mapsto Q'\otimes_S P$$

because

$$(Q\otimes_R S)\otimes_S P\simeq Q\otimes_R P.$$

To be added: This, and generalities on spectral sequences, need to be added! From this spectral sequence, it will be relatively easy to deduce the result.

1. Suppose $M\otimes_R S$ is S-flat and $\mathrm{Tor}_1^R(M,S)=0$. We want to show that 2 holds, so let N be any S-module. Consider the E_2 page of the above spectral sequence $\mathrm{Tor}_p^S(\mathrm{Tor}_q^R(M,S),N)\implies\mathrm{Tor}_{p+q}^R(M,N)$. In the terms such that $p+q=1$, we have the two terms $\mathrm{Tor}_0^S(\mathrm{Tor}_1^R(M,S),N)$, $\mathrm{Tor}_1^S(\mathrm{Tor}_0^R(M,$ But by hypotheses these are both zero. It follows that $\mathrm{Tor}_1^R(M,N)=0$.
2. Suppose $\mathrm{Tor}_1^R(M,N)=0$ for each S-module N. Since this is a homology spectral sequence, this implies that the E_2^{10} term vanishes (since nothing will be able to hit this term). In particular $\mathrm{Tor}_1^S(M\otimes_R S,N)=0$ for each S-module N. It follows that $M\otimes_R S$ is S-flat. Hence the higher terms $\mathrm{Tor}_p^S(M\otimes_R S,N)=0$ as well, so the botton row of the E_2 page (except $(0,0)$) is thus entirely zero. It follows that the E_{01}^2 term vanishes if E_∞^{01} is trivial. This gives that $\mathrm{Tor}_1^R(M,S)\otimes_S N=0$ for every S-module N, which clearly implies $\mathrm{Tor}_1^R(M,S)=0$. ∎

As a result, we shall be able to deduce the result alluded to in the motivation following the statement of theorem 1.4.5.

1.4.8 Lemma *Let R be a noetherian ring, $J \subset R$ an ideal, M an R-module. Then TFAE:*

1. $\mathrm{Tor}_1^R(M, R/J) = 0$ *and M/JM is R/J-flat.*
2. $\mathrm{Tor}_1^R(M, N) = 0$ *for any finitely generated R-module N annihilated by a power of J.*

Proof. This is immediate from lemma 1.4.7, once one notes that any N as in the statement admits a finite filtration whose successive quotients are annihilated by J. □

Proof of theorem 1.4.5. Only one direction is nontrivial, so suppose M is a finitely generated S-module, with M/JM flat over R/J and $\mathrm{Tor}_1^R(M, R/J) = 0$. We know by the lemma that $\mathrm{Tor}_1^R(M, N) = 0$ whenever N is finitely generated and annihilated by a power of J.

So as to avoid repeating the same argument over and over, we encapsulate it in the following lemma.

1.4.9 Lemma *Let the hypotheses be as in theorem 1.4.5 Suppose for every ideal $I \subset R$, and every $t \in \mathbb{N}$, the map*

$$I/I \cap J^t \otimes M \to M/J^t M$$

is an injection. Then M is R-flat.

Proof. Indeed, then as before, the kernel of $I \otimes_R M \to M$ lives inside the image of $(I \cap J^t) \otimes M \to I \otimes_R M$ for *every* t; by the Artin-Rees lemma, and the Krull intersection theorem (since $\bigcap J^t(I \otimes_R M) = \{0\}$), it follows that this kernel is zero. □

It is now easy to finish the proof. Indeed, we can verify the hypotheses of the lemma by noting that

$$I/I \cap J^t \otimes M \to I \otimes M$$

is obtained by tensoring with M the sequence

$$0 \to I/I \cap J^t \to R/(I \cap J^t) \to R/(I + J^t) \to 0.$$

Since $\mathrm{Tor}_1^R(M, R/(I + J^t)) = 0$, we find that the map as in the lemma is an injection, and so we are done. □

The reader can similarly formulate a version of the infinitesimal criterion in this more general case using lemma 1.4.9 and the argument in theorem 1.4.3. (In fact, the spectral sequence argument of this section is not necessary.) We shall not state it here, as it will appear as a component of theorem 1.4.10. We leave the details of the proof to the reader.

The final statement of the flatness criterion

We shall now bundle the various criteria for flatness into one big result, following **?**:

1.4.10 Theorem *Let A, B be noetherian rings, $\phi : A \to B$ a morphism making B into an A-algebra. Let I be an ideal of A such that $\phi(I)$ is contained in the Jacobson radical of B. Let M be a finitely generated B-module. Then the following are equivalent:*

1. *M is A-flat.*
2. *(Local criterion) M/IM is A/I-flat and $\mathrm{Tor}_1^A(M, A/I) = 0$.*
3. *(Infinitesimal criterion) M/I^nM is A/I^n-flat for each n.*
4. *(Associated graded criterion) M/IM is A/I-flat and $M/IM \otimes_{A/I} I^n/I^{n+1} \to I^nM/I^{n+1}M$ is an isomorphism for each n.*

The last criterion can be phrased as saying that the I-adic *associated graded* of M is determined by M/IM.

Proof. We have already proved that the first three are equivalent. It is easy to see that flatness of M implies that

$$M/IM \otimes_{A/I} I^n/I^{n+1} \to I^nM/I^{n+1}M \tag{1.4.3}$$

is an isomorphism for each n. Indeed, this easily comes out to be the quotient of $M \otimes_A I^n$ by the image of $M \otimes_A I^{n+1}$, which is $I^nM/I^{n+1}M$ since the map $M \otimes_A I^n \to I^nM$ is an isomorphism. Now we need to show that this last condition implies flatness. To do this, we may (in view of the infinitesimal criterion) assume that I is *nilpotent*, by base-changing to A/I^n. We are then reduced to showing that $\mathrm{Tor}_1^A(M, A/I) = 0$ (by the local criterion). Then we are, finally, reduced to showing:

1.4.11 Lemma *Let A be a ring, $I \subset A$ be a nilpotent ideal, and M any A-module. If* (1.4.3) *is an isomorphism for each n, then* $\mathrm{Tor}_1^A(M, A/I) = 0$.

Proof. This is equivalent to the assertion, by a diagram chase, that

$$I \otimes_A M \to M$$

is an injection. We shall show more generally that $I^n \otimes_A M \to M$ is an injection for each n. When $n \gg 0$, this is immediate, I being nilpotent. So we can use descending induction on n.

Suppose $I^{n+1} \otimes_A M \to I^{n+1}M$ is an isomorphism. Consider the diagram

$$\begin{tikzcd}
 & I^{n+1} \otimes_A M \arrow[r] \arrow[d] & I^n \otimes_A M \arrow[r] \arrow[d] & I^n/I^{n+1} \otimes_A M \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & I^{n+1}M \arrow[r] & I^nM \arrow[r] & I^nM/I^{n+1}M \arrow[r] & 0.
\end{tikzcd}$$

□

By hypothesis, the outer two vertical arrows are isomorphisms. Thus the middle vertical arrow is an isomorphism as well. This completes the induction hypothesis. □

Here is an example of the above techniques:

1.4.12 Proposition *Let $(A, \mathfrak{m}), (B, \mathfrak{n}), (C, \mathfrak{n}')$ be noetherian local rings. Suppose given a commutative diagram of local homomorphisms*

$$\begin{tikzcd} B \arrow[rr] & & C \\ & A \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Suppose B, C are flat A-algebras, and $B/\mathfrak{m}B \to C/\mathfrak{m}C$ is a flat morphism. Then $B \to C$ is flat.

Geometrically, this means that flatness can be checked fiberwise if both objects are flat over the base. This will be a useful technical fact.

Proof. We will use the associated graded criterion for flatness with the ideal $I = \mathfrak{m}B \subset B$. (Note that we are *not* using the criterion with the maximal ideal here!) Namely, we shall show that

$$I^n/I^{n+1} \otimes_{B/I} C/IC \to I^nC/I^{n+1}C \tag{1.4.4}$$

is an isomorphism. By theorem 1.4.10, this will do it. Now we have:

$$\begin{aligned} I^n/I^{n+1} \otimes_{B/I} C/IC &\simeq \mathfrak{m}^nB/\mathfrak{m}^{n+1}B \otimes_{B/\mathfrak{m}B} C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A B/\mathfrak{m}B \otimes_B C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A B \otimes_B C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A C/\mathfrak{m}C \\ &\simeq \mathfrak{m}^nC/\mathfrak{m}^{n+1}C \simeq I^nC/I^{n+1}C. \end{aligned}$$

In this chain of equalities, we have used the fact that B, C were flat over A, so their associated gradeds with respect to $\mathfrak{m} \subset A$ behave nicely. It follows that (1.4.4) is an isomorphism, completing the proof. ∎

Flatness over regular local rings

Here we shall prove a result that implies geometrically, for instance, that a finite morphism between smooth varieties is always flat.

1.4.13 Theorem ("Miracle" flatness theorem) *Let $(A, \mathfrak{m})$ be a regular local (noetherian) ring. Let $(B, \mathfrak{n})$ be a Cohen-Macaulay, local A-algebra such that*

$$\dim B = \dim A + \dim B/\mathfrak{m}B.$$

Then B is flat over A.

Recall that *inequality* $\leqslant$ always holds in the above for any morphism of noetherian local rings (**??**), and equality always holds with flatness supposed. We get a partial converse.

Proof. We shall work by induction on $\dim A$. Let $x \in \mathfrak{m}$ be a non-zero-divisor, so the first element in a regular sequence of parameters. We are going to show that $(A/(x), B/(x))$ satisfies the same hypotheses. Indeed, note that

$$\dim B/(x) \leqslant \dim A/(x) + \dim B/\mathfrak{m}B$$

by the usual inequality. Since $\dim A/(x) = \dim A - 1$, we find that quotienting by x drops the dimension of B by at least one: that is, $\dim B/(x) \leqslant \dim B - 1$. By the principal ideal theorem, we have equality,

$$\dim B/(x) = \dim B - 1.$$

The claim is that x is a non-zero-divisor in B, and consequently we can argue by induction. Indeed, but B is *Cohen-Macaulay.* Thus, any zero divisor in B lies in a *minimal* prime (since all associated primes of B are minimal); thus quotienting by a zero divisor would not bring down the degree. So x is a non-zero-divisor in B.

In other words, we have found $x \in A$ which is both A-regular and B-regular (i.e. non-zero-divisors on both), and such that the hypotheses of the theorem apply to the pair $(A/(x), B/(x))$. It follows that $B/(x)$ is flat over $A/(x)$ by the inductive hypothesis. The next lemma will complete the proof. ∎

1.4.14 Lemma *Suppose $(A, \mathfrak{m})$ is a noetherian local ring, $(B, \mathfrak{n})$ a noetherian local A-algebra, and M a finite B-module. Suppose $x \in A$ is a regular element of A which is also regular on M. Suppose moreover M/xM is $A/(x)$-flat. Then M is flat over A.*

Proof. This follows from the associated graded criterion for flatness (see the omnibus result theorem 1.4.10). Indeed, if we use the notation of that result, we take $I = (x)$. We are given that M/xM is $A/(x)$-flat. So we need to show that

$$M/xM \otimes_{A/(x)} (x^n)/(x^{n+1}) \to x^n M/x^{n+1}M$$

is an isomorphism for each n. This, however, is implied because $(x^n)/(x^{n+1})$ is isomorphic to $A/(x)$ by regularity, and multiplication

$$M \xrightarrow{x^n} x^n M, \quad xM \xrightarrow{x^n} x^{n+1}M$$

are isomorphisms by M-regularity. ∎

Example: construction of flat extensions

As an illustration of several of the techniques in this chapter and previous ones, we shall show, following **?** (volume III, chapter 0) that, given a local ring and an extension of its residue field, one may find a flat extension of this local ring with the bigger field as *its* residue field. One application of this is in showing (in the context of Zariski's Main Theorem) that the fibers of a birational projective morphism of noetherian schemes (where the target is normal) are *geometrically* connected. We shall later give another application in the theory of étale morphisms.

1.4.15 Theorem *Let $(R, \mathfrak{m})$ be a noetherian local ring with residue field k. Suppose K is an extension of k. Then there is a noetherian local R-algebra $(S, \mathfrak{n})$ with residue field K such that S is flat over R and $\mathfrak{n} = \mathfrak{m}S$.*

Proof. Let us start by motivating the theorem when K is generated over k by *one* element. This case can be handled directly, but the general case will require a somewhat tricky passage to the limit. There are two cases.

1. First, suppose $K = k(t)$ for $t \in K$ *transcendental* over k. In this case, we will take S to be a suitable localization of $R[t]$. Namely, we consider the prime[3] ideal $\mathfrak{m}R[t] \subset R[t]$, and let $S = (R[t])_{\mathfrak{m}R[t]}$. Then S is clearly noetherian and local, and moreover $\mathfrak{m}S$ is the maximal ideal of S. The residue field of S is $S/\mathfrak{m}S$, which is easily seen to be the quotient field of $R[t]/\mathfrak{m}R[t] = k[t]$, and is thus isomorphic to K. Moreover, as a localization of a polynomial ring, S is flat over R. Thus we have handled the case of a purely transcendental extension generated by one element.

2. Let us now suppose $K = k(a)$ for $a \in K$ *algebraic* over k. Then a satisfies a monic irreducible polynomial $\overline{p}(T)$ with coefficients in k. We lift $\overline{p}$ to a monic polynomial $p(T) \in R[T]$. The claim is that then, $S = R[T]/(p(T))$ will suffice.

   Indeed, S is clearly flat over R (in fact, it is free of rank $\deg p$). As it is finite over R, S is noetherian. Moreover, $S/\mathfrak{m}S = k[T]/(p(T)) \simeq K$. It follows that $\mathfrak{m}S \subset S$ is a maximal ideal and that the residue field is K. Since any maximal ideal of S contains $\mathfrak{m}S$ by Nakayama,[4] we see that S is local as well. Thus we have showed that S satisfies all the conditions we want.

So we have proved the theorem when K is generated by one element over k. In general, we can iterate this procedure finitely many times, so that the assertion is clear when K is a finitely generated extension of k. Extending to infinitely generated extensions is trickier.

Let us first argue that we can write K/k as a "transfinite limit" of monogenic extensions. Consider the set of well-ordered collections $\mathcal{C}'$ of subfields between k and K (containing k) such that if $L \in \mathcal{C}'$ has an immediate predecessor L', then L/L' is generated by one element. First, such collections $\mathcal{C}'$ clearly exist; we can take the one consisting only of k. The set of such collections is clearly a partially ordered set such that every chain has an upper bound. By Zorn's lemma, there is a *maximal* such collection of subfields, which we now call $\mathcal{C}$.

The claim is that $\mathcal{C}$ has a maximal field, which is K. Indeed, if it had no maximal element, we could adjoin the union $\bigcup_{F \in \mathcal{C}} F$ to $\mathcal{C}$ and make $\mathcal{C}$ bigger, contradicting maximality. If this maximal field of $\mathcal{C}$ were not K, then we could add another element to this maximal subfield and get a bigger collection than $\mathcal{C}$, contradiction.

So thus we have a set of fields K_α (with α, the index, ranging over a well-ordered set) between k and K, such that if α has a successor α', then K'_α is generated by one element over K_α. Moreover K is the largest of the K_α, and k is the smallest.

[3] It is prime because the quotient is the domain $k[t]$.
[4] **To be added: citation needed**

We are now going to define a collection of rings R_α by transfinite induction on α. We start the induction with $R_0 = R$ (where 0 is the smallest allowed α). The inductive hypothesis that we will want to maintain is that R_α is a noetherian local ring with maximal ideal $\mathfrak{m}_\alpha$, flat over R and satisfying $\mathfrak{m}R_\alpha = \mathfrak{m}_\alpha$; we require, moreover, that the residue field of R_α be K_α. Thus if we can do this at each step, we will be able to work up to K and get the ring S that we want. We are, moreover, going to construct the R_α such that whenever $\beta < \alpha$, R_α is a R_β-algebra.

Let us assume that R_β has been defined for all $\beta < \alpha$ and satisfies the conditions. Then we want to define R_α in an appropriate way. If we can do this, then we will have proved the result. There are two cases:

1. α has an immediate predecessor α_{pre}. In this case, we can define R_α from $R_{\alpha_{pre}}$ as above (because $K_\alpha/K_{\alpha_{pre}}$ is monogenic).
2. α has no immediate predecessor. Then we define $R_\alpha = \varinjlim_{\beta<\alpha} R_\beta$. The following lemma will show that R_α satisfies the appropriate hypotheses.

This completes the proof, modulo lemma 1.4.16. □

We shall need the following lemma to see that we preserve noetherianness when we pass to the limit.

1.4.16 Lemma *Suppose given an inductive system $\{(A_\alpha, \mathfrak{m}_\alpha)\}$ of noetherian rings and flat local homomorphisms, starting with A_0. Suppose moreover that $\mathfrak{m}_\alpha A_\beta = \mathfrak{m}_\beta$ whenever $\alpha < \beta$.*

Then $A = \varinjlim A_\alpha$ is a noetherian local ring, flat over each A_α. Moreover, if $\mathfrak{m} \subset A$ is the maximal ideal, then $\mathfrak{m}_\alpha A = \mathfrak{m}$. The residue field of A is $\varinjlim A_\alpha/\mathfrak{m}_\alpha$.

Proof. First, it is clear that A is a local ring (**?? To be added: reference!**) with maximal ideal equal to $\mathfrak{m}_\alpha A$ for any α in the indexing set, and that A has the appropriate residue field. Since filtered colimits preserve flatness, flatness of A is also clear. We need to show that A is noetherian; this is the crux of the lemma.

To prove that A is noetherian, we are going to show that its $\mathfrak{m}$-adic completion $\hat{A}$ is noetherian. Fortunately, we have a convenient criterion for this. If $\hat{\mathfrak{m}} = \mathfrak{m}\hat{A}$, then $\hat{A}$ is complete with respect to the $\hat{\mathfrak{m}}$-adic topology. So if we show that $\hat{A}/\hat{\mathfrak{m}}$ is noetherian and $\hat{\mathfrak{m}}/\hat{\mathfrak{m}^2}$ is a finitely generated $\hat{A}$-module, we will have shown that $\hat{A}$ is noetherian by corollary 8.1.13.

But $\hat{A}/\hat{\mathfrak{m}}$ is a field, so obviously noetherian. Also, $\hat{\mathfrak{m}}/\hat{\mathfrak{m}}^2 = \mathfrak{m}/\mathfrak{m}^2$, and by flatness of A, this is

$$A \otimes_{A_\alpha} \mathfrak{m}_\alpha/\mathfrak{m}_\alpha^2$$

for any α. Since A_α is noetherian, we see that this is finitely generated. The criterion corollary 8.1.13 now shows that the completion $\hat{A}$ is noetherian.

Finally, we need to deduce that A is itself noetherian. To do this, we shall show that $\hat{A}$ is faithfully flat over A. Since noetherianness "descends" under faithfully flat extensions (**To be added: citation needed**), this will be enough. It suffices to show that $\hat{A}$ is *flat* over each A_α. For this, we use the infinitesimal criterion; we have that

$$\hat{A} \otimes_{A_\alpha} A_\alpha/\mathfrak{m}_\alpha^t = \hat{A}/\hat{\mathfrak{m}^t} = A/\mathfrak{m}^t = A/A\mathfrak{m}_\alpha^t,$$

which is flat over $A_\alpha/\mathfrak{m}_\alpha^t$ since A is flat over A_α.

It follows that $\hat{A}$ is flat over each A_α. If we want to see that $A \to \hat{A}$ is flat, we let $I \subset A$ be a finitely generated ideal; we shall prove that $I \otimes_A \hat{A} \to \hat{A}$ is injective (which will establish flatness). We know that there is an ideal $I_\alpha \subset A_\alpha$ for some A_α such that

$$I = I_\alpha A = I_\alpha \otimes_{A_\alpha} A.$$

Then

$$I \otimes_A \hat{A} = I_\alpha \otimes_{A_\alpha} \hat{A}$$

which injects into $\hat{A}$ as $A_\alpha \to \hat{A}$ is flat.

1.4.17 Remark (comment) Let us first show that A is *separated* with respect to the $\mathfrak{m}$-adic topology. Fix $x \in A$. Then x lies in the subring A_α for some fixed α depending on α (note that $A_\alpha \to A$ is injective since a flat morphism of local rings is *faithfully flat*). If $x \in \mathfrak{m}^n = A\mathfrak{m}_\alpha^n$, then $x \in \mathfrak{m}_\alpha^n$ by faithful flatness and lemma 1.1.24. So if $x \in \mathfrak{m}^n$ for all n, then $x \in \mathfrak{m}_\alpha^n$ for all n; the separatedness of A_α with respect to the $\mathfrak{m}_\alpha$-adic topology now shows $x = 0$.

Generic flatness

Suppose given a module M over a noetherian *domain* R. Then $M \otimes_R K(R)$ is a finitely generated free module over the field $K(R)$. Since $K(R)$ is the inductive limit $\varinjlim R_f$ as f ranges over $(R - \{0\})/R^*$ and $K(R) \otimes_R M \simeq \varinjlim_{f \in (R-\{0\})/R^*} M_f$, it follows by the general theory of **??** that there exists $f \in R - \{0\}$ such that M_f is free over R_f.

Here $\operatorname{Spec} R_f = D(f) \subset \operatorname{Spec} R$ should be thought of as a "big" subset of $\operatorname{Spec} R$ (in fact, as one can check, it is *dense* and open). So the moral of this argument is that M is "generically free." If we had the language of schemes, we could make this more precise. But the idea is that localizing at M corresponds to restricting the *sheaf* associated to M to $D(f) \subset \operatorname{Spec} R$; on this dense open subset, we get a free sheaf. (The reader not comfortable with such "finitely presented" arguments will find another one below, that also works more generally.)

Now we want to generalize this to the case where M is finitely generated not over R, but over a finitely generated R-algebra. In particular, M could itself be a finitely generated R-algebra!

1.4.18 Theorem (Generic freeness) *Let S be a finitely generated algebra over the noetherian domain R, and let M be a finitely generated S-module. Then there is $f \in R - \{0\}$ such that M_f is a free (in particular, flat) R-module.*

Proof. We shall first reduce the result to one about rings instead of modules. By Hilbert's basis theorem, we know that S is noetherian. By dévissage (proposition 2.2.12), there is a finite filtration of M by S-submodules,

$$0 = M_0 \subset M_1 \subset \cdots \subset M_k = M$$

such that the quotients M_{i+1}/M_i are isomorphic to quotients $S/\mathfrak{p}_i$ for the $\mathfrak{p}_i \in \operatorname{Spec} S$.

Since localization is an exact functor, it will suffice to show that there exists an f such that $(S/\mathfrak{p}_i)_f$ is a free R-module for each f. Indeed, it is clear that if a module admits a finite filtration all of whose successive quotients are free, then the module itself is free. We may thus even reduce to the case where $M = S/\mathfrak{p}$.

So we are reduced to showing that if we have a finitely generated *domain* T over R, then there exists $f \in R - \{0\}$ such that T_f is a free R-module. If $R \to T$ is not injective, then the result is obvious (localize at something nonzero in the kernel), so we need only handle the case where $R \to T$ is a monomorphism.

By the Noether normalization theorem, there are d elements of $T \otimes_R K(R)$, which we denote by $t_1, \dots, t_d$, which are algebraically independent over $K(R)$ and such that $T \otimes_R K(R)$ is integral over $K(R)[t_1, \dots, t_d]$. (Here d is the transcendence degree of $K(T)/K(R)$.) If we localize at some highly divisible element, we can assume that $t_1, \dots, t_d$ all lie in T itself. *Let us assume that the result for domains is true whenever the transcendence degree is $< d$, so that we can induct.*

Then we know that $R[t_1, \dots, t_d] \subset T$ is a polynomial ring. Moreover, each of the finitely many generators of T/R satisfies a monic polynomial equation over $K(R)[t_1, \dots, t_d]$ (by the integrality part of Noether normalization). If we localize R at a highly divisible element, we may assume that the coefficients of these polynomials belong to $R[t_1, \dots, t_d]$. We have thus reduced to the following case. T is a finitely generated domain over R, *integral* over the polynomial ring $R[t_1, \dots, t_d]$. In particular, it is a finitely generated module over the polynomial ring $R[t_1, \dots, t_d]$. Thus we have some r and an exact sequence

$$0 \to R[t_1, \dots, t_d]^r \to T \to Q \to 0,$$

where Q is a torsion $R[t_1, \dots, t_d]^r$-module. Since the polynomial ring is free, we are reduced to showing that by localizing at a suitable element of R, we can make Q free.

But now we can do an inductive argument. Q has a finite filtration by T-modules whose quotients are isomorphic to $T/\mathfrak{p}$ for nonzero primes $\mathfrak{p}$ with $\mathfrak{p} \neq 0$ as T is torsion; these are still domains finitely generated over R, but such that the associated transcendence degree is *less* than d. We have already assumed the statement proven for domains where the transcendence degree is $< d$. Thus we can find a suitable localization that makes all these free, and thus Q free; it follows that with this localization, T becomes free too. $\square$

V.2. Homological theory of local rings

We will now apply general homological algebra to commutative algebra proper. The use of homological machinery provides a new and elegant characterization of regular local rings (among noetherian local rings, they are the ones with finite global dimension) and leads to proofs of several difficult results about them. For instance, we will be able to prove the rather important result (which one repeatedly uses in algebraic geometry) that a regular local ring is a UFD. As another example, the aforementioned criterion leads to a direct proof of the otherwise non-obvious that a localization of a regular local ring at a prime ideal is still a regular local ring.

Note: right now, the material on regular local rings is still missing! It should be added.

2.1. Depth

In this section, we first introduce the notion of *depth* for local rings via the Ext functor, and then show that depth can be measured as the length of a maximal *regular sequence*. After this, we study the theory of regular sequences in general (on not-necessarily-local rings), and show that the depth of a module can be bounded in terms of both its dimension and its associated primes.

Depth over local rings

Throughout, let $(R, \mathfrak{m})$ be a noetherian local ring. Let $k = R/\mathfrak{m}$ be the residue field.

Let $M \neq 0$ be a finitely generated R-module. We are going to define an arithmetic invariant of M, called the *depth*, that will measure in some sense the torsion of M.

2.1.1 Definition The **depth** of M is equal to the smallest integer i such that $\mathrm{Ext}^i(k, M) \neq 0$. If there is no such integer, we set $\operatorname{depth} M = \infty$.

We shall give another characterization of this shortly that makes no reference to Ext functors, and is purely elementary. We will eventually see that there is always such an i (at least if $M \neq 0$), so $\operatorname{depth} M < \infty$.

2.1.2 Example (Depth zero) Let us characterize when a module M has depth zero. Depth zero is equivalent to saying that $\mathrm{Ext}^0(k, M) = \hom_R(k, M) \neq 0$, i.e. that there is a nontrivial morphism

$$k \to M.$$

As $k = R/\mathfrak{m}$, the existence of such a map is equivalent to the existence of a nonzero x such that $\mathrm{Ann}(x) = \mathfrak{m}$, i.e. $\mathfrak{m} \in \mathrm{Ass}(M)$. So depth zero is equivalent to having $\mathfrak{m} \in \mathrm{Ass}(M)$.

Suppose now that $\mathrm{depth}(M) \neq 0$. In particular, $\mathfrak{m} \notin \mathrm{Ass}(M)$. Since $\mathrm{Ass}(M)$ is finite, prime avoidance implies that $\mathfrak{m} \not\subset \bigcup_{\mathfrak{p} \in \mathrm{Ass}(M)} \mathfrak{p}$. Thus $\mathfrak{m}$ contains an element which is a non-zero-divisor on M (see proposition 2.2.13). So we find:

2.1.3 Proposition *M has depth zero iff every element in $\mathfrak{m}$ is a zero divisor on M.*

Now suppose $\mathrm{depth}\, M \neq 0$. There is $a \in \mathfrak{m}$ which is a non-zero-divisor on M, i.e. such that there is an exact sequence

$$0 \to M \xrightarrow{a} M \to M/aM \to 0.$$

For each i, there is an exact sequence in Ext groups:

$$\mathrm{Ext}^{i-1}(k, M/aM) \to \mathrm{Ext}^i(k, M) \xrightarrow{a} \mathrm{Ext}^i(k, M) \to \mathrm{Ext}^i(k, M/aM) \to \mathrm{Ext}^{i+1}(k, M). \quad (2.1.1)$$

However, the map $a : \mathrm{Ext}^i(k, M) \to \mathrm{Ext}^i(k, M)$ is zero as multiplication by a kills k. (If a kills a module N, then it kills $\mathrm{Ext}^*(N, M)$ for all M.) We see from this that

$$\mathrm{Ext}^i(k, M) \hookrightarrow \mathrm{Ext}^i(k, M/aM)$$

is injective, and

$$\mathrm{Ext}^{i-1}(k, M/aM) \twoheadrightarrow \mathrm{Ext}^i(k, M)$$

is surjective.

2.1.4 Corollary *If $a \in \mathfrak{m}$ is a non-zero-divisor on M, then*

$$\mathrm{depth}(M/aM) = \mathrm{depth}\, M - 1.$$

Proof. When $\mathrm{depth}\, M = \infty$, this is easy (left to the reader) from the exact sequence. Suppose $\mathrm{depth}(M) = n$. We would like to see that $\mathrm{depth}\, M/aM = n - 1$. That is, we want to see that $\mathrm{Ext}^{n-1}(k, M/aM) \neq 0$, but $\mathrm{Ext}^i(k, M/aM) = 0$ for $i < n - 1$. This is direct from the sequence (2.1.1) above. In fact, surjectivity of $\mathrm{Ext}^{n-1}(k, M/aM) \to \mathrm{Ext}^n(k, M)$ shows that $\mathrm{Ext}^{n-1}(k, M/aM) \neq 0$. Now let $i < n - 1$. Then in (2.1.1), $\mathrm{Ext}^i(k, M/aM)$ is sandwiched between two zeros, so it is zero. ∎

The moral of the above discussion is that one quotients out by a non-zero-divisor, the depth drops by one. In fact, we have described a recursive algorithm for computing $\mathrm{depth}(M)$.

1. If $\mathfrak{m} \in \mathrm{Ass}(M)$, output zero.
2. If $\mathfrak{m} \notin \mathrm{Ass}(M)$, choose an element $a \in \mathfrak{m}$ which is a non-zero-divisor on M. Output $\mathrm{depth}(M/aM) + 1$.

If one wished to apply this in practice, one would probably start by looking for a non-zero-divisor $a_1 \in \mathfrak{m}$ on M, and then looking for one on M/a_1M, etc. From this we make:

2.1.5 Definition Let $(R, \mathfrak{m})$ be a local noetherian ring, M a finite R-module. A sequence $a_1, \ldots, a_n \in \mathfrak{m}$ is said to be **M-regular** iff:

1. a_1 is a non-zero-divisor on M
2. a_2 is a non-zero-divisor on M/a_1M
3. ...
4. a_i is a non-zero-divisor on $M/(a_1, \dots, a_{i-1})M$ for all i.

A regular sequence $a_1, \dots, a_n$ is **maximal** if it can be extended no further, i.e. there is no a_{n+1} such that $a_1, \dots, a_{n+1}$ is M-regular.

We now get the promised "elementary" characterization of depth.

2.1.6 Corollary depth(M) *is the length of every maximal M-regular sequence. In particular, all M-regular sequences have the same length.*

Proof. If $a_1, \dots, a_n$ is M-regular, then

$$\operatorname{depth} M/(a_1, \dots, a_i)M = \operatorname{depth} M - i$$

for each i, by an easy induction on i and the corollary 2.1.4. From this the result is clear, because depth zero occurs precisely when $\mathfrak{m}$ is an associated prime (proposition 2.1.3). But it is also clear that a regular sequence $a_1, \dots, a_n$ is maximal precisely when every element of $\mathfrak{m}$ acts as a zero divisor on $M/(a_1, \dots, a_n)M$, that is, $\mathfrak{m} \in \operatorname{Ass}(M/(a_1, \dots, a_n)M)$. □

2.1.7 Remark We could *define* the depth via the length of a maximal M-regular sequence.

Finally, we can bound the depth in terms of the dimension.

2.1.8 Corollary *Let $M \neq 0$. Then the depth of M is finite. In fact,*

$$\operatorname{depth} M \leqslant \dim M. \tag{2.1.2}$$

Proof. When $\operatorname{depth} M = 0$, the assertion is obvious. Otherwise, there is $a \in \mathfrak{m}$ which is a non-zero-divisor on M. We know that

$$\operatorname{depth} M/aM = \operatorname{depth} M - 1$$

and (by proposition 7.2.2)

$$\dim M/aM = \dim M - 1.$$

By induction on $\dim M$, we have that $\operatorname{depth} M/aM \leqslant \dim M/aM$. From this the induction step is clear, because depth and dim both drop by one after quotienting. □

Generally, the depth is not the dimension.

2.1.9 Example Given any M, adding k makes the depth zero: $M \oplus k$ has $\mathfrak{m}$ as an associated prime. But the dimension does not jump to zero just by adding a copy of k. If M is a direct sum of pieces of differing dimensions, then the bound (2.1.2) does not exhibit equality. In fact, if M, M' are finitely generated modules, then we have

$$\operatorname{depth} M \oplus M' = \min\left(\operatorname{depth} M, \operatorname{depth} M'\right),$$

which follows at once from the definition of depth in terms of vanishing Ext groups.

2.1.10 Remark (exercise) Suppose R is a noetherian local ring whose depth (as a module over itself) is zero. If R is reduced, then R is a field.

Finally, we include a result that states that the depth does not depend on the ring so much as the module.

2.1.11 Proposition (Depth and change of rings) *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a morphism of noetherian local rings. Suppose M is a finitely generated S-module, which is also finitely generated as an R-module. Then $\operatorname{depth}_R M = \operatorname{depth}_S M$.*

Proof. It is clear that we have the inequality $\operatorname{depth}_R M \leqslant \operatorname{depth}_S M$, by the interpretation of depth via regular sequences. Let $x_1, \dots, x_n \in R$ be a maximal M-sequence. We need to show that it is a maximal M-sequence in S as well. By quotienting, we may replace M with $M/(x_1, \dots, x_n)M$; we then have to show that if M has depth zero as an R-module, it has depth zero as an S-module.

But then $\hom_R(R/\mathfrak{m}, M) \neq 0$. This is a R-submodule of M, consisting of elements killed by $\mathfrak{m}$, and in fact it is a S-submodule. We are going to show that $\mathfrak{n}$ annihilates some element of it, which will imply that $\operatorname{depth}_S M = 0$.

To see this, note that $\hom_R(R/\mathfrak{m}, M)$ is artinian as an R-module (as it is killed by $\mathfrak{m}$). As a result, it is an artinian S-module, which means it contains $\mathfrak{n}$ as an associated prime, proving the claim and the result. ∎

Regular sequences

In the previous subsec, we defined the notion of *depth* of a finitely generated module over a noetherian local ring using the Ext functors. We then showed that the depth was the length of a maximal regular sequence.

Now, although it will not be necessary for the main results in this chapter, we want to generalize this to the case of a non-local ring. Most of the same arguments go through, though there are some subtle differences. For instance, regular sequences remain regular under permutation in the local case, but not in general. Since there will be some repetition, we shall try to be brief.

We start by generalizing the idea of a regular sequence which is not required to be contained in the maximal ideal of a local ring. Let R be a noetherian ring, and M a finitely generated R-module.

2.1.12 Definition A sequence $x_1, \dots, x_n \in R$ is M**-regular** (or is an M**-sequence**) if for each $k \leqslant n$, x_k is a non-zero-divisor on the R-module $M/(x_1, \dots, x_{k-1})M$ and also $(x_1, \dots, x_n)M \neq M$.

So x_1 is a non-zero-divisor on M, by the first part. That is, the homothety $M \xrightarrow{x_1} M$ is injective. The last condition is also going to turn out to be necessary for us. In the previous subsec, it was automatic as $\mathfrak{m}M \neq M$ (unless $M = 0$) by Nakayama's lemma as M was assumed finitely generated.

The property of being a regular sequence is inherently an inductive one. Note that $x_1, \dots, x_n$ is a regular sequence on M if and only if x_1 is a zero divisor on M and $x_2, \dots, x_n$ is an M/x_1M-sequence.

2.1.13 Definition If M is an R-module and $I \subset R$ an ideal, then we write $\operatorname{depth}_I M$ for the length of the length-maximizing M-sequence contained in I. When R is local and $I \subset R$ the maximal ideal, then we just write $\operatorname{depth} M$ as before.

While we will in fact have a similar characterization of depth in terms of Ext, in this section we *define* it via regular sequences.

2.1.14 Example The basic example one is supposed to keep in mind is the polynomial ring $R = R_0[x_1, \dots, x_n]$ and $M = R$. Then the sequence $x_1, \dots, x_n$ is regular in R.

2.1.15 Example Let $(R, \mathfrak{m})$ be a regular local ring, and let $x_1, \dots, x_n$ be a regular system of parameters in R (i.e. a system of generators for $\mathfrak{m}$ of minimal size). Then we have seen that the $\{x_i\}$ form a regular sequence on R, in any order. This is because each quotient $R/(x_1, \dots, x_i)$ is itself regular, hence a domain.

As before, we have a simple characterization of depth zero:

2.1.16 Proposition *Let R be noetherian, M finitely generated. If M is an R-module with $IM \neq M$, then M has depth zero if and only if I is contained in an element of* $\operatorname{Ass}(M)$*.*

Proof. This is analogous to proposition 2.1.3. Note than an ideal consists of zero divisors on M if and only if it is contained in an associated prime (proposition 2.2.13). ∎

The above proof used proposition 2.2.13, a key fact which will be used repeatedly in the sequel. This is one reason the theory of depth works best for finitely generated modules over noetherian rings.

The first observation to make is that regular sequences are *not* preserved by permutation. This is one nice characteristic that we would like but is not satisfied.

2.1.17 Example Let k be a field. Consider $R = k[x, y]/((x-1)y, yz)$. Then x, z is a regular sequence on R. Indeed, x is a non-zero-divisor and $R/(x) = k[z]$. However, z, x is not a regular sequence because z is a zero divisor in R.

Nonetheless, regular sequences *are* preserved by permutation for local rings under suitable noetherian hypotheses:

2.1.18 Proposition *Let R be a noetherian local ring and M a finite R-module. Then if $x_1, \dots, x_n$ is a M-sequence contained in the maximal ideal, so is any permutation $x_{\sigma(1)}, \dots, x_{\sigma(n)}$.*

Proof. It is clearly enough to check this for a transposition. Namely, if we have an M-sequence

$$x_1, \dots, x_i, x_{i+1}, \dots x_n$$

we would like to check that so is

$$x_1, \dots, x_{i+1}, x_i, \dots, x_n.$$

It is here that we use the inductive nature. Namely, all we need to do is check that

$$x_{i+1}, x_i, \dots, x_n$$

is regular on $M/(x_1, \dots, x_{i-1})M$, since the first part of the sequence will automatically be regular. Now $x_{i+2}, \dots, x_n$ will automatically be regular on $M/(x_1, \dots, x_{i+1})M$. So all we need to show is that x_{i+1}, x_i is regular on $M/(x_1, \dots, x_{i-1})M$.

The moral of the story is that we have reduced to the following lemma.

2.1.19 Lemma *Let R be a noetherian local ring. Let N be a finite R-module and $a, b \in R$ an N-sequence contained in the maximal ideal. Then so is b, a.*

Proof. We can prove this as follows. First, a will be a non-zero-divisor on N/bN. Indeed, if not then we can write

$$an = bn'$$

for some $n, n' \in N$ with $n \notin bN$. But b is a non-zero-divisor on N/aN, which means that $bn' \in aN$ implies $n' \in aN$. Say $n' = an''$. So $an = ban''$. As a is a non-zero-divisor on N, we see that $n = bn''$. Thus $n \in bN$, contradiction. This part has not used the fact that R is local.

Now we claim that b is a non-zero-divisor on N. Suppose $n \in N$ and $bn = 0$. Since b is a non-zero-divisor on N/aN, we have that $n \in aN$, say $n = an'$. Thus

$$b(an') = a(bn') = 0.$$

The fact that $N \xrightarrow{a} N$ is injective implies that $bn' = 0$. So we can do the same and get $n' = an''$, $n'' = an^{(3)}, n^{(3)} = an^{(4)}$, and so on. It follows that n is a multiple of $a, a^2, a^3, \dots$, and hence in $\mathfrak{m}^j N$ for each j where $\mathfrak{m} \subset R$ is the maximal ideal. The Krull intersection theorem now implies that $n = 0$.

Together, these arguments imply that b, a is an N-sequence, proving the lemma. □

The proof of the result is now complete. □

One might wonder what goes wrong, and why permutations do not preserve regular sequences in general; after all, oftentimes we can reduce results to their analogs for local rings. Yet the fact that regularity is preserved by permutations for local rings does not extend to arbitrary rings. The problem is that regular sequences do *not* localize. Well, they almost do, but the final condition that $(x_1, \dots, x_n)M \neq M$ doesn't get preserved. We can state:

2.1.20 Proposition *Suppose $x_1, \dots, x_n$ is an M-sequence. Let N be a flat R-module. Then if $(x_1, \dots, x_n)M \otimes_R N \neq M \otimes N$, then $x_1, \dots, x_n$ is an $M \otimes_R N$-sequence.*

Proof. This is actually very easy now. The fact that $x_i : M/(x_1, \dots, x_{i-1})M \to M/(x_1, \dots, x_{i-1})M$ is injective is preserved when M is replaced by $M \otimes_R N$ because the functor $- \otimes_R N$ is exact. □

In particular, it follows that if we have a good reason for supposing that $(x_1, \dots, x_n)M \otimes N \neq M \otimes N$, then we'll already be done. For instance, if N is the localization of R at a prime ideal containing the x_i. Then we see that automatically $x_1, \dots, x_n$ is an $M_{\mathfrak{p}} = M \otimes_R R_{\mathfrak{p}}$-sequence.

Finally, we have an analog of the previous correspondence between depth and the vanishing of Ext. Since the argument is analogous to corollary 2.1.6, we omit it.

2.1.21 Theorem *Let R be a ring. Suppose M is an R-module and $IM \neq M$. All maximal M-sequences in I have the same length. This length is the smallest value of r such that* $\mathrm{Ext}^r(R/I, M) \neq 0$.

2.1.22 Remark (exercise) Suppose I is an ideal in R. Let M be an R-module such that $IM \neq M$. Show that $\mathrm{depth}_I M \geqslant 2$ if and only if the natural map

$$M \simeq \hom(R, M) \to \hom(I, M)$$

is an isomorphism.

Powers of regular sequences

Regular sequences don't necessarily behave well with respect to permutation or localization without additional hypotheses. However, in all cases they behave well with respect to taking powers. The upshot of this is that the invariant called *depth* that we will soon introduce is invariant under passing to the radical.

We shall deduce this from the following easy fact.

2.1.23 Lemma *Suppose we have an exact sequence of R-modules*

$$0 \to M' \to M \to M'' \to 0.$$

Suppose the sequence $x_1, \dots, x_n \in R$ is M'-regular and M''-regular. Then it is M-regular.

The converse is not true, of course.

Proof. Morally, this is the snake lemma. For instance, the fact that multiplication by x_1 is injective on M', M'' implies by the snake diagram that $M \xrightarrow{x_1} M$ is injective. However, we don't a priori know that a simple inductive argument on n will work to prove this. The reason is that it needs to be seen that quotienting each term by $(x_1, \dots, x_{n-1})$ will preserve exactness. However, a general fact will tell us that this is indeed the case. See below.

Anyway, this general fact now lets us induct on n. If we assume that $x_1, \dots, x_{n-1}$ is M-regular, we need only prove that $x_n : M/(x_1, \dots, x_{n-1})M \to M/(x_1, \dots, x_{n-1})$ is injective. (It is not surjective or the sequence would not be M''-regular.) But we have the exact sequence by the next lemma,

$$0 \to M'/(x_1 \dots x_{n-1})M' \to M/(x_1 \dots x_{n-1})M \to M''/(x_1 \dots x_{n-1})M'' \to 0$$

and the injectivity of x_n on the two ends implies it at the middle by the snake lemma. $\square$

So we need to prove:

2.1.24 Lemma *Suppose $0 \to M' \to M \to M'' \to 0$ is a short exact sequence. Let $x_1, \dots, x_m$ be an M''-sequence. Then the sequence*

$$0 \to M'/(x_1 \dots x_m)M' \to M/(x_1 \dots x_m)M \to M''/(x_1 \dots x_m)M'' \to 0$$

is exact as well.

One argument here uses the fact that the Tor functors vanish when one has a regular sequence like this. We can give a direct argument.

Proof. By induction, this needs only be proved when $m = 1$, since we have the recursive description of regular sequences: in general, $x_2 \dots x_m$ will be regular on M''/x_1M''. In any case, we have exactness except possibly at the left as the tensor product is right-exact. So let $m' \in M'$; suppose m' maps to a multiple of x_1 in M. We need to show that m' is a multiple of x_1 in M'.

Suppose m' maps to x_1m. Then x_1m maps to zero in M'', so by regularity m maps to zero in M''. Thus m comes from something, $\overline{m}'$, in M'. In particular $m' - x_1\overline{m}'$ maps to zero in M, so it is zero in M'. Thus indeed m' is a multiple of x_1 in M'. ∎

With this lemma proved, we can state:

2.1.25 Proposition *Let M be an R-module and $x_1, \dots, x_n$ an M-sequence. Then $x_1^{a_1}, \dots, x_n^{a_n}$ is an M-sequence for any $a_1, \dots, a_n \in \mathbb{Z}_{>0}$.*

Proof. We will use:

2.1.26 Lemma *Suppose $x_1, \dots, x_i, \dots, x_n$ and $x_1, \dots, x_i', \dots, x_n$ are M-sequences for some M. Then so is $x_1, \dots, x_ix_i', \dots, x_n$.*

Proof. As usual, we can mod out by $(x_1 \dots x_{i-1})$ and thus assume that $i = 1$. We have to show that if $x_1, \dots, x_n$ and $x_1', \dots, x_n$ are M-sequences, then so is $x_1x_1', \dots, x_n$.

We have an exact sequence

$$0 \to x_1M/x_1x_1'M \to M/x_1x_1'M \to M/x_1M \to 0.$$

Now $x_2, \dots, x_n$ is regular on the last term by assumption, and also on the first term, which is isomorphic to $M/x_1'M$ as x_1 acts as a non-zero-divisor on M. So $x_2, \dots, x_n$ is regular on both ends, and thus in the middle. This means that

$$x_1x_1', \dots, x_n$$

is M-regular. That proves the lemma. ∎

So we now can prove the proposition. It is trivial if $\sum a_i = n$ (i.e. if all are 1) it is clear. In general, we can use complete induction on $\sum a_i$. Suppose we know the result for smaller values of $\sum a_i$. We can assume that some $a_j > 1$. Then the sequence

$$x_1^{a_1}, \dots x_j^{a_j}, \dots x_n^{a_n}$$

is obtained from the sequences

$$x_1^{a_1}, \dots, x_j^{a_j - 1}, \dots, x_n^{a_n}$$

and

$$x_1^{a_1}, \dots, x_j^{1}, \dots, x_n^{a_n}$$

by multiplying the middle terms. But the complete induction hypothesis implies that both those two sequences are M-regular, so we can apply the lemma. $\square$

In general, the product of two regular sequences is not a regular sequence. For instance, consider a regular sequence x, y in some finitely generated module M over a noetherian local ring. Then y, x is regular, but the product sequence xy, xy is *never* regular.

Depth

We make the following definition slightly differently than in the local case:

2.1.27 Definition Suppose I is an ideal such that $IM \neq M$. Then we define the **I-depth of** M to be the maximum length of a maximal M-sequence contained in I. When R is a local ring and I the maximal ideal, then that number is simply called the **depth** of M.

The **depth** of a proper ideal $I \subset R$ is its depth on R.

The definition is slightly awkward, but it turns out that all maximal M-sequences in I have the same length, as we saw in theorem 2.1.21. So we can use any of them to compute the depth.

The first thing we can prove using the above machinery is that depth is really a "geometric" invariant, in that it depends only on the radical of I.

2.1.28 Proposition *Let R be a ring, $I \subset R$ an ideal, and M an R-module with $IM \neq M$. Then* $\mathrm{depth}_I M = \mathrm{depth}_{\mathrm{Rad}(I)} M$.

Proof. The inequality $\mathrm{depth}_I M \leqslant \mathrm{depth}_{\mathrm{Rad} I} M$ is trivial, so we need only show that if $x_1, \dots, x_n$ is an M-sequence in $\mathrm{Rad}(I)$, then there is an M-sequence of length n in I. For this we just take a high power

$$x_1^N, \dots, x_n^N$$

where N is large enough such that everything is in I. We can do this as powers of M-sequences are M-sequences (proposition 2.1.25). $\square$

This was a fairly easy consequence of the above result on powers of regular sequences. On the other hand, we want to give another proof, because it will let us do more. Namely, we will show that depth is really a function of prime ideals.

For convenience, we set the following condition: if $IM = M$, we define

$$\mathrm{depth}_I(M) = \infty.$$

2.1.29 Proposition *Let R be a noetherian ring, $I \subset R$ an ideal, and M a finitely generated R-module. Then*

$$\mathrm{depth}_I M = \min_{\mathfrak{p} \in V(I)} \mathrm{depth}_{\mathfrak{p}} M.$$

So the depth of I on M can be calculated from the depths at each prime containing I. In this sense, it is clear that $\mathrm{depth}_I(M)$ depends only on $V(I)$ (and the depths on those primes), so clearly it depends only on I *up to radical.*

Proof. In this proof, we shall use the fact that the length of every maximal M-sequence is the same (theorem 2.1.21).

It is obvious that we have an inequality

$$\mathrm{depth}_I \leqslant \min_{\mathfrak{p} \in V(I)} \mathrm{depth}_{\mathfrak{p}} M$$

as each of those primes contains I. We are to prove that there is a prime $\mathfrak{p}$ containing I with

$$\mathrm{depth}_I M = \mathrm{depth}_{\mathfrak{p}} M.$$

But we shall actually prove the stronger statement that there is $\mathfrak{p} \supset I$ with $\mathrm{depth}_{\mathfrak{p}} M_{\mathfrak{p}} = \mathrm{depth}_I M$. Note that localization at a prime can only increase depth because an M-sequence in $\mathfrak{p}$ leads to an M-sequence in $M_{\mathfrak{p}}$ thanks to Nakayama's lemma and the flatness of localization.

So let $x_1, \dots, x_n \in I$ be a M-sequence of maximum length. Then I acts by zero divisors on $M/(x_1, \dots, x_n)M$ or we could extend the sequence further. In particular, I is contained in an associated prime of $M/(x_1, \dots, x_n)M$ by elementary commutative algebra (basically, prime avoidance).

Call this associated prime $\mathfrak{p} \in V(I)$. Then $\mathfrak{p}$ is an associated prime of $M_{\mathfrak{p}}/(x_1, \dots, x_n)M_{\mathfrak{p}}$, and in particular acts only by zero divisors on this module. Thus the $M_{\mathfrak{p}}$-sequence $x_1, \dots, x_n$ can be extended no further in $\mathfrak{p}$. In particular, since the depth can be computed as the length of *any* maximal $M_{\mathfrak{p}}$-sequence,

$$\mathrm{depth}_{\mathfrak{p}} M_{\mathfrak{p}} = \mathrm{depth}_I M.$$

□

Perhaps we should note a corollary of the argument above:

2.1.30 Corollary *Hypotheses as above, we have* $\mathrm{depth}_I M \leqslant \mathrm{depth}_{\mathfrak{p}} M_{\mathfrak{p}}$ *for any prime* $\mathfrak{p} \supset I$. *However, there is at least one* $\mathfrak{p} \supset I$ *where equality holds.*

We are thus reduced to analyzing depth in the local case.

2.1.31 Remark (exercise) If $(R, \mathfrak{m})$ is a local noetherian ring and M a finitely generated R-module, then show that $\mathrm{depth}\, M = \mathrm{depth}_{\hat{R}} \hat{M}$, where $\hat{M}$ is the $\mathfrak{m}$-adic completion. (Hint: use $\hat{M} = M \otimes_R \hat{R}$, and the fact that $\hat{R}$ is flat over R.)

Depth and dimension

Consider an R-module M, which is always assumed to be finitely generated. Let $I \subset R$ be an ideal with $IM \neq M$. We deduce from the previous subsecs:

2.1.32 Proposition *Let M be a finitely generated module over the noetherian ring R. Then*

$$\operatorname{depth}_I M \leqslant \dim M$$

for any ideal $I \subset R$ with $IM \neq M$.

Proof. We have proved this when R is a *local* ring (corollary 2.1.8). Now we just use corollary 2.1.30 to reduce to the local case. ☐

This does not tell us much about how $\operatorname{depth}_I M$ depends on I, though; it just says something about how it depends on M. In particular, it is not very helpful when trying to estimate $\operatorname{depth} I = \operatorname{depth}_I R$. Nonetheless, there is a somewhat stronger result, which we will need in the future. We start by stating the version in the local case.

2.1.33 Proposition *Let $(R, \mathfrak{m})$ be a noetherian local ring. Let M be a finite R-module. Then the depth of $\mathfrak{m}$ on M is at most the dimension of $R/\mathfrak{p}$ for $\mathfrak{p}$ an associated prime of M:*

$$\operatorname{depth} M \leqslant \min_{\mathfrak{p} \in \operatorname{Ass}(M)} \dim R/\mathfrak{p}.$$

This is sharper than the bound $\operatorname{depth} M \leqslant \dim M$, because each $\dim R/\mathfrak{p}$ is at most $\dim M$ (by definition).

Proof. To prove this, first assume that the depth is zero. In that case, the result is immediate. We shall now argue inductively. Assume that that this is true for modules of smaller depth. We will quotient out appropriately to shrink the support and change the associated primes. Namely, choose a M-regular (non-zero-divisor on M) $x \in R$. Then $\operatorname{depth} M/xM = \operatorname{depth} M - 1$.

Let $\mathfrak{p}_0$ be an associated prime of M. We claim that $\mathfrak{p}_0$ is *properly* contained in an associated prime of M/xM. We will prove this below. Thus $\mathfrak{p}_0$ is properly contained in some $\mathfrak{q}_0 \in \operatorname{Ass}(M/xM)$.

Now we know that $\operatorname{depth} M/xM = \operatorname{depth} M - 1$. Also, by the inductive hypothesis, we know that $\dim R/\mathfrak{q}_0 \geqslant \operatorname{depth} M/xM = \operatorname{depth} M - 1$. But the dimension of $R/\mathfrak{q}_0$ is strictly smaller than that of $R/\mathfrak{p}_0$, so at least $\dim R/\mathfrak{p}_0 + 1 \geqslant \operatorname{depth} M$. This proves the lemma, modulo the result:

2.1.34 Lemma *Let $(R, \mathfrak{m})$ be a noetherian local ring. Let M be a finitely generated R-module, $x \in \mathfrak{m}$ an M-regular element. Then each element of $\operatorname{Ass}(M)$ is properly contained in an element of $\operatorname{Ass}(M/xM)$.*

So if we quotient by a regular element, we can make the associated primes jump up.

Proof. Let $\mathfrak{p}_0 \in \mathrm{Ass}(M)$; we want to show $\mathfrak{p}_0$ is properly contained in something in $\mathrm{Ass}(M/xM)$.

Indeed, $x \notin \mathfrak{p}_0$, so $\mathfrak{p}_0$ cannot itself be an associated prime. However, $\mathfrak{p}_0$ annihilates a nonzero element of M/xM. To see this, consider a maximal principal submodule of M annihilated by $\mathfrak{p}_0$. Let this submodule be Rz for some $z \in M$. Then if z is a multiple of x, say $z = xz'$, then Rz' would be a larger submodule of M annihilated by $\mathfrak{p}_0$—here we are using the fact that x is a non-zero-divisor on M. So the image of this z in M/xM is nonzero and is clearly annihilated by $\mathfrak{p}_0$. It follows $\mathfrak{p}_0$ is contained in an element of $\mathrm{Ass}(M/xM)$, necessarily properly. ∎

2.1.35 Remark (exercise) Another argument for lemma 2.1.34 is given in §16 of **?**, vol. IV, by reducing to the coprimary case. Here is a sketch.

The strategy is to use the existence of an exact sequence

$$0 \to M' \to M \to M'' \to 0$$

with $\mathrm{Ass}(M'') = \mathrm{Ass}(M) - \{\mathfrak{p}_0\}$ and $\mathrm{Ass}(M') = \{\mathfrak{p}_0\}$. Quotienting by x preserves exactness, and we get

$$0 \to M'/xM' \to M/xM \to M''/xM'' \to 0.$$

Now $\mathfrak{p}_0$ is properly contained in every associated prime of M'/xM' (as it acts nilpotently on M'). It follows that any element of $\mathrm{Ass}(M'/xM') \subset \mathrm{Ass}(M/xM)$ will do the job.

In essence, the point is that the result is *trivial* when $\mathrm{Ass}(M) = \{\mathfrak{p}_0\}$.

2.1.36 Remark (exercise) Here is a simpler argument for lemma 2.1.34, following **?**. Let $\mathfrak{p}_0 \in \mathrm{Ass}(M)$, as before. Again as before, we want to show that $\hom_R(R/\mathfrak{p}_0, M/xM) \neq 0$. But we have an exact sequence

$$0 \to \hom_R(R/\mathfrak{p}_0, M) \xrightarrow{x} \hom_R(R/\mathfrak{p}_0, M) \to \hom_R(R/\mathfrak{p}_0, M/xM),$$

and since the first map is not surjective (by Nakayama), the last object is nonzero.

Finally, we can globalize the results:

2.1.37 Proposition *Let R be a noetherian ring, $I \subset R$ an ideal, and M a finitely generated module. Then $\mathrm{depth}_I M$ is at most the length of every chain of primes in* $\mathrm{Spec} R$ *that starts at an associated prime of M and ends at a prime containing I.*

Proof. Currently omitted.

2.1.38 Remark (comment) Consider a chain of primes $\mathfrak{p}_0 \subset \cdots \subset \mathfrak{p}_k$ where $\mathfrak{p}_0$ is an associated prime and $\mathfrak{p}_k$ contains I. The goal is to show that

$$\mathrm{depth}_I M \leqslant k.$$

By localization, we can assume that $\mathfrak{p}_k$ is the maximal ideal of R; recall that localization can only increase the depth. We can also assume I is this maximal ideal, by increasing it.

In this case, the result follows from the local version (proposition 2.1.33).

2.2. Cohen-Macaulayness

Cohen-Macaualay modules over a local ring

For a local noetherian ring, we have discussed two invariants of a module: dimension and depth. They generally do not coincide, and Cohen-Macaulay modules will be those where they do.

Let $(R, \mathfrak{m})$ be a noetherian local ring.

2.2.1 Definition A finitely generated R-module M is **Cohen-Macaulay** if $\operatorname{depth} M = \dim M$. The ring R is called **Cohen-Macaulay** if it is Cohen-Macaulay as a module over itself.

We already know that the inequality $\leqslant$ always holds. If there is a system of parameters for M (i.e., a sequence $x_1, \dots, x_r \in \mathfrak{m}$ such that $M/(x_1, \dots, x_r)M$ is artinian) which is a regular sequence on M, then M is Cohen-Macaulay: we see in fact that $\dim M = \operatorname{depth} M = r$. This is the distinguishing trait of Cohen-Macaulay rings.

Let us now give a few examples:

2.2.2 Example (Regular local rings are Cohen-Macaulay) If R is regular, then $\operatorname{depth} R = \dim R$, so R is Cohen-Macaulay.

Indeed, we have seen that if $x_1, \dots, x_n$ is a regular system of parameters for R (i.e. a minimal set of generators for $\mathfrak{m}$), then $n = \dim R$ and the $\{x_i\}$ form a regular sequence. See the remark after corollary 9.1.12; the point is that $R/(x_1, \dots, x_{i-1})$ is regular for each i (by the aforementioned corollary), and hence a domain, so x_i acts on it by a non-zero-divisor.

The next example easily shows that a Cohen-Macaulay ring need not be regular, or even a domain:

2.2.3 Example (Local artinian rings are Cohen-Macaulay) Any local artinian ring, because the dimension is zero for an artinian ring.

2.2.4 Example (Cohen-Macaulayness and completion) A finitely generated module M is Cohen-Macaulay if and only if its completion $\hat{M}$ is; this follows from remark 2.1.31.

Here is a slightly harder example.

2.2.5 Example A normal local domain $(R, \mathfrak{m})$ of dimension 2 is Cohen-Macaulay. This is a special case of Serre's criterion for normality.

Here is an argument. If $x \in \mathfrak{m}$ is nonzero, we want to show that $\operatorname{depth} R/(x) = 1$. To do this, we need to show that $\mathfrak{m} \notin \operatorname{Ass}(R/(x))$ for each such x, because then $\operatorname{depth} R/(x) \geqslant 1$ (which is all we need). However, suppose the contrary; then there is y not divisible by x such that $\mathfrak{m}y \subset (x)$. So $y/x \notin R$, but $\mathfrak{m}(y/x) \subset R$.

This, however, implies $\mathfrak{m}$ is principal. Indeed, we either have $\mathfrak{m}(y/x) = R$, in which case $\mathfrak{m}$ is generated by x/y, or $\mathfrak{m}(y/x) \subset \mathfrak{m}$. The latter would imply that y/x is integral over R (as multiplication by it stabilizes a finitely generated R-module), and by normality $y/x \in R$. We have seen much of this argument before.

2.2.6 Example Consider $\mathbb{C}[x, y]/(xy)$, the coordinate ring of the union of two axes intersecting at the origin. This is not regular, as its localization at the origin is not a domain. We will later show that this is a Cohen-Macaulay ring, though.

2.2.7 Remark (comment) Indeed, we can project the associated variety $X = V(xy)$ onto the affine line by adding the coordinates. This corresponds to the map

$$\mathbb{C}[z] \to \mathbb{C}[x, y]/(xy)$$

sending $z \to x + y$. This makes $\mathbb{C}[x, y]/(xy)$ into a free $\mathbb{C}[z]$-module of rank two (with generators $1, x$), as one can check. So by the previous result (strictly speaking, its extension to non-domains), the ring in question is Cohen-Macaulay.

2.2.8 Example $R = \mathbb{C}[x, y, z]/(xy, xz)$ is not Cohen-Macaulay (at the origin). The associated variety looks geometrically like the union of the plane $x = 0$ and the line $y = z = 0$ in affine 3-space. Here there are two components of different dimensions intersecting. Let's choose a regular sequence (that is, regular after localization at the origin). The dimension at the origin is clearly two because of the plane. First, we need a non-zero-divisor in this ring, which vanishes at the origin, say $x + y + z$. (Check this.) When we quotient by this, we get

$$S = \mathbb{C}[x, y, z]/(xy, xz, x + y + z) = \mathbb{C}[y, z]/((y + z)y, (y + z)z).$$

The claim is that S localized at the ideal corresponding to $(0, 0)$ has depth zero. We have $y + z \neq 0$, which is killed by both y, z, and hence by the maximal ideal at zero. In particular the maximal ideal at zero is an associated prime, which implies the claim about the depth.

As it happens, a Cohen-Macaulay variety is always equidimensional. The rough reason is that each irreducible piece puts an upper bound on the depth given by the dimension of the piece. If any piece is too small, the total depth will be too small.

Here is the deeper statement:

2.2.9 Proposition *Let $(R, \mathfrak{m})$ be a noetherian local ring, M a finitely generated, Cohen-Macaulay R-module. Then:*

1. *For each $\mathfrak{p} \in \mathrm{Ass}(M)$, we have $\dim M = \dim R/\mathfrak{p}$.*
2. *Every associated prime of M is minimal (i.e. minimal in $\mathrm{supp}\, M$).*
3. *$\mathrm{supp}\, M$ is equidimensional.*

In general, there may be nontrivial inclusion relations among the associated primes of a general module. However, this cannot happen for a Cohen-Macaulay module.

Proof. The first statement implies all the others. (Recall that *equidimensional* means that all the irreducible components of $\mathrm{supp}\, M$, i.e. the $\mathrm{Spec}\, R/\mathfrak{p}$, have the same dimension.) But this in turn follows from the bound of proposition 2.1.33. ☐

Next, we would like to obtain a criterion for when a quotient of a Cohen-Macaulay module is still Cohen-Macaulay. The answer will be similar to theorem 9.1.14 for regular local rings.

2.2.10 Proposition *Let M be a Cohen-Macaulay module over the local noetherian ring $(R, \mathfrak{m})$. If $x_1, \dots, x_n \in \mathfrak{m}$ is a M-regular sequence, then $M/(x_1, \dots, x_n)M$ is Cohen-Macaulay of dimension (and depth)* $\dim M - n$.

Proof. Indeed, we reduce to the case $n = 1$ by induction. But then, because x_1 is a non-zero-divisor on M, we have $\dim M/x_1M = \dim M - 1$ and $\operatorname{depth} M/x_1M = \operatorname{depth} M - 1$. Thus

$$\dim M/x_1M = \operatorname{depth} M/x_1M. \qquad \square$$

So, if we are given a Cohen-Macaulay module M and want one of a smaller dimension, we just have to find $x \in \mathfrak{m}$ not contained in any of the minimal primes of $\operatorname{supp} M$ (these are the only associated primes). Then, M/xM will do the job.

The non-local case

More generally, we would like to make the definition:

2.2.11 Definition A general noetherian ring R is **Cohen-Macaulay** if $R_\mathfrak{p}$ is Cohen-Macaulay for all $\mathfrak{p} \in \operatorname{Spec} R$.

We should check that these definitions coincide for a local noetherian ring. This, however, is not entirely obvious; we have to show that localization preserves Cohen-Macaulayness. In this subsec, we shall do that, and we shall furthermore show that Cohen-Macaulay rings are *catenary*, or more generally that Cohen-Macaulay modules are catenary. (So far we have seen that they are equidimensional, in the local case.)

We shall deduce this from the following result, which states that for a Cohen-Macaulay module, we can choose partial systems of parameters in any given prime ideal in the support.

2.2.12 Proposition *Let M be a Cohen-Macaulay module over the local noetherian ring $(R, \mathfrak{m})$, and let $\mathfrak{p} \in \operatorname{supp} M$. Let $x_1, \dots, x_r \in \mathfrak{p}$ be a maximal M-sequence contained in $\mathfrak{p}$. Then:*

1. *$\mathfrak{p}$ is an associated and minimal prime of $M/(x_1, \dots, x_r)M$.*
2. $\dim R/\mathfrak{p} = \dim M - r$

Proof. We know (proposition 2.2.10) that $M/(x_1, \dots, x_r)M$ is a Cohen-Macaulay module too. Clearly $\mathfrak{p}$ is in its support, since all the $x_i \in \mathfrak{p}$. The claim is that $\mathfrak{p}$ is an associated prime—or minimal prime, it is the same thing—of $M/(x_1, \dots, x_r)M$. If not, there is $x \in \mathfrak{p}$ that is a non-zero-divisor on this quotient, which means that $\{x_1, \dots, x_r\}$ was not maximal as claimed.

Now we need to verify the assertion on the dimension. Clearly $\dim M/(x_1, \dots, x_r)M = \dim M - r$, and moreover $\dim R/\mathfrak{p} = \dim M/(x_1, \dots, x_r)$ by proposition 2.2.9. Combining these gives the second assertion. $\square$

2.2.13 Corollary *Hypotheses as above,* $\dim M_\mathfrak{p} = r = \dim M - \dim R/\mathfrak{p}$. *Moreover, $M_\mathfrak{p}$ is a Cohen-Macaulay module over $R_\mathfrak{p}$.*

This result shows that definition 2.2.11 is a reasonable definition.

Proof. Indeed, if we consider the conclusions of **??**, we find that $x_1, \ldots, x_r$ becomes a system of parameters for $M_{\mathfrak{p}}$: we have that $M_{\mathfrak{p}}/(x_1, \ldots, x_r)M_{\mathfrak{p}}$ is an artinian $R_{\mathfrak{p}}$-module, while the sequence is also regular. The first claim follows, as does the second: any module with a system of parameters that is a regular sequence is Cohen-Macaulay. ☐

As a result, we can get the promised result that a Cohen-Macaulay ring is catenary.

2.2.14 Proposition *If M is Cohen-Macaulay over the local noetherian ring R, then* $\operatorname{supp} M$ *is a catenary space.*

In other words, if $\mathfrak{p} \subset \mathfrak{q}$ are elements of $\operatorname{supp} M$, then every maximal chain of prime ideals from $\mathfrak{p}$ to $\mathfrak{q}$ has the same length.

Proof. We will show that $\dim R/\mathfrak{p} = \dim R/\mathfrak{q} + \dim R_{\mathfrak{q}}/\mathfrak{p}R_{\mathfrak{q}}$, a claim that suffices to establish catenariness. We will do this by using the dimension formulas computed earlier.

Namely, we know that M is catenary over R, so by corollary 2.2.13

$$\dim_{R_{\mathfrak{q}}} M_{\mathfrak{q}} = \dim M - \dim R/\mathfrak{q}, \quad \dim_{R_{\mathfrak{p}}} M_{\mathfrak{p}} = \dim M - \dim R/\mathfrak{p}.$$

Moreover, $M_{\mathfrak{q}}$ is Cohen-Macaulay over $R_{\mathfrak{q}}$. As a result, we have (in view of the previous equation)

$$\dim_{R_{\mathfrak{p}}} M_{\mathfrak{p}} = \dim_{R_{\mathfrak{q}}} M_{\mathfrak{q}} - \dim R_{\mathfrak{q}}/\mathfrak{p}R_{\mathfrak{q}} = \dim M - \dim R/\mathfrak{q} - \dim R_{\mathfrak{q}}/\mathfrak{p}R_{\mathfrak{q}}.$$

Combining, we find

$$\dim M - \dim R/\mathfrak{p} = \dim M - \dim R/\mathfrak{q} - \dim R_{\mathfrak{q}}/\mathfrak{p}R_{\mathfrak{q}},$$

which is what we wanted. ☐

It thus follows that any Cohen-Macaulay ring, and thus any *quotient* of a Cohen-Macaualay ring, is catenary. In particular, it follows any non-catenary local noetherian ring cannot be expressed as a quotient of a Cohen-Macaulay (e.g. regular) local ring.

It also follows immediately that if R is any regular (not necessarily local) ring, then R is catenary, and the same goes for any quotient of R. In particular, since a polynomial ring over a field is regular, we find:

2.2.15 Proposition *Any affine ring is catenary.*

Reformulation of Serre's criterion

Much earlier, we proved criteria for a noetherian ring to be reduced and (more interestingly) normal. We can state them more cleanly using the theory of depth developed.

2.2.16 Definition Let R be a noetherian ring, and let $k \in \mathbb{Z}_{\geqslant 0}$.

1. We say that R satisfies **condition** R_k if, for every prime ideal $\mathfrak{p} \in \operatorname{Spec} R$ with $\dim R_{\mathfrak{p}} \leqslant k$, the local ring $R_{\mathfrak{p}}$ is regular.
2. R satisfies **condition** S_k if $\operatorname{depth} R_{\mathfrak{p}} \geqslant \inf(k, \dim R_{\mathfrak{p}})$ for all $\mathfrak{p} \in \operatorname{Spec} R$.

A Cohen-Macaulay ring satisfies all the conditions S_k, and conversely. The condition R_k means geometrically that the associated variety is regular (i.e., smooth, at least if one works over an algebraically closed field) outside a subvariety of codimension $\geqslant k$.

Recall that, according to **??**, a noetherian ring is *reduced* iff:

1. For any minimal prime $\mathfrak{p} \subset R$, $R_{\mathfrak{p}}$ is a field.
2. Every associated prime of R is minimal.

Condition 1 can be restated as follows. The ideal $\mathfrak{p} \subset R$ is minimal if and only if it is zero-dimensional, and $R_{\mathfrak{p}}$ is regular if and only if it is a field. So the first condition is that *for every height zero prime, $R_{\mathfrak{p}}$ is regular.* In other words, it is the condition R_0.

For the second condition, $\mathfrak{p} \in \operatorname{Ass}(R)$ iff $\mathfrak{p} \in \operatorname{Ass}(R_{\mathfrak{p}})$, which is equivalent to $\operatorname{depth} R_{\mathfrak{p}} = 0$. So the second condition states that for primes $\mathfrak{p} \in \operatorname{Spec} R$ of height at least 1, $\mathfrak{p} \notin \operatorname{Ass}(R_{\mathfrak{p}})$, or $\operatorname{depth}(R_{\mathfrak{p}}) \geqslant 1$. This is the condition S_1.

We find:

2.2.17 Proposition *A noetherian ring is reduced if and only if it satisfies R_0 and S_1.*

In particular, for a Cohen-Macaulay ring, checking if it is reduced is easy; one just has to check R_0 (if the localizations at minimal primes are reduced).

Serre's criterion for normality is in the same spirit, but harder. Recall that a noetherian ring is *normal* if it is a finite direct product of integrally closed domains.

The earlier form of Serre's criterion (see theorem 4.5.14) was:

2.2.18 Proposition *Let R be a local ring. Then R is normal iff*

1. *R is reduced.*
2. *For every height one prime $\mathfrak{p} \in \operatorname{Spec} R$, $R_{\mathfrak{p}}$ is a DVR (i.e. regular).*
3. *For every non-zero-divisor $x \in R$, every associated prime of $R/(x)$ is minimal.*

In view of the criterion for reducedness, these conditions are equivalent to:

1. For every prime $\mathfrak{p}$ of height $\leqslant 1$, $R_{\mathfrak{p}}$ is regular.
2. For every prime $\mathfrak{p}$ of height $\geqslant 1$, $\operatorname{depth} R_{\mathfrak{p}} \geqslant 1$ (necessary for reducedness)
3. $\operatorname{depth} R_{\mathfrak{p}} \geqslant 2$ for $\mathfrak{p}$ containing but not minimal over any principal ideal (x) for x a non-zero-divisor. This is the last condition of the proposition; to say $\operatorname{depth} R_{\mathfrak{p}} \geqslant 2$ is to say that $\operatorname{depth} R_{\mathfrak{p}}/(x)R_{\mathfrak{p}} \geqslant 1$, or $\mathfrak{p} \notin \operatorname{Ass}(R_{\mathfrak{p}}/(x)R_{\mathfrak{p}})$.

Combining all this, we find:

2.2.19 Theorem (Serre's criterion) *A noetherian ring is normal if and only if it satisfies the conditions R_1 and S_2.*

Again, for a Cohen-Macaulay ring, the last condition is automatic, as the depth is the codimension.

2.3. Projective dimension and free resolutions

We shall introduce the notion of *projective dimension* of a module; this will be the smallest projective resolution it admits (if there is none such, the dimension is ∞). We can think of it as measuring how far a module is from being projective. Over a noetherian *local* ring, we will show that the projective dimension can be calculated very simply using the Tor functor (which is an elaboration of the story that a projective module over a local ring is free).

Ultimately we want to show that a noetherian local ring is regular if and only if every finitely generated module admits a finite free resolution. Although we shall not get to that result until the next section, we will at least relate projective dimension to a more familiar invariant of a module: *depth.*

Introduction

Let R be a commutative ring, M an R-module.

2.3.1 Definition The **projective dimension** of M is the largest integer n such that there exists a module N with

$$\mathrm{Ext}^n(M, N) \neq 0.$$

We allow ∞, if arbitrarily large such n exist. We write $\mathrm{pd}(M)$ for the projective dimension. For convenience, we set $\mathrm{pd}(0) = -\infty$.

So, if $m > n = \mathrm{pd}(M)$, then we have $\mathrm{Ext}^m(M, N) = 0$ for *all* modules N, and n is the smallest integer with this property. As an example, note that $\mathrm{pd}(M) = 0$ if and only if M is projective and nonzero. Indeed, we have seen that the Ext groups $\mathrm{Ext}^i(M, N), i > 0$ vanish always for M projective, and conversely.

To compute $\mathrm{pd}(M)$ in general, one can proceed as follows. Take any M. Choose a surjection $P \twoheadrightarrow M$ with P projective; call the kernel K and draw a short exact sequence

$$0 \to K \to P \to M \to 0.$$

For any R-module N, we have a long exact sequence

$$\mathrm{Ext}^{i-1}(P, N) \to \mathrm{Ext}^{i-1}(K, N) \to \mathrm{Ext}^i(M, N) \to \mathrm{Ext}^i(P, N).$$

If $i > 0$, the right end vanishes; if $i > 1$, the left end vanishes. So if $i > 1$, this map $\mathrm{Ext}^{i-1}(K, N) \to \mathrm{Ext}^i(M, N)$ is an *isomorphism.*

Suppose that $\mathrm{pd}(K) = d \geqslant 0$. We find that $\mathrm{Ext}^{i-1}(K, N) = 0$ for $i - 1 > d$. This implies that $\mathrm{Ext}^i(M, N) = 0$ for such $i > d + 1$. In particular, $\mathrm{pd}(M) \leqslant d + 1$. This argument is completely reversible if $d > 0$. Then we see from these isomorphisms that

$$\boxed{\mathrm{pd}(M) = \mathrm{pd}(K) + 1}, \quad \text{unless } \mathrm{pd}(M) = 0 \tag{2.3.1}$$

If M is projective, the sequence $0 \to K \to P \to M \to 0$ splits, and $\mathrm{pd}(K) = 0$ too.

The upshot is that we can compute projective dimension by choosing a projective resolution.

2.3.2 Proposition *Let M be an R-module. Then* $\mathrm{pd}(M) \leqslant n$ *iff there exists a finite projective resolution of M having $n + 1$ terms,*

$$0 \to P_n \to \cdots \to P_1 \to P_0 \to M \to 0.$$

Proof. Induction on n. When $n = 0$, M is projective, and we can use the resolution $0 \to M \to M \to 0$.

Suppose $\mathrm{pd}(M) \leqslant n$, where $n > 0$. We can get a short exact sequence

$$0 \to K \to P_0 \to M \to 0$$

with P_0 projective, so $\mathrm{pd}(K) \leqslant n-1$ by (2.3.1). The inductive hypothesis implies that there is a projective resolution of K of length $\leqslant n - 1$. We can splice this in with the short exact sequence to get a projective resolution of M of length n.

The argument is reversible. Choose any projective resolution

$$0 \to P_n \to \cdots \to P_1 \to P_0 \to M \to 0$$

and split into short exact sequences, and then one argue inductively to show that $\mathrm{pd}(M) \leqslant n$.□

Let $\mathrm{pd}(M) = n$. Choose any projective resolution $\cdots \to P_2 \to P_1 \to P_0 \to M$. Choose $K_i = \ker(P_i \to P_{i-1})$ for each i. Then there is a short exact sequence $0 \to K_0 \to P_0 \to M \to 0$. Moreover, there are exact sequences

$$0 \to K_i \to P_i \to K_{i-1} \to 0$$

for each i. From these, and from (2.3.1), we see that the projective dimensions of the K_i drop by one as i increments. So K_{n-1} is projective if $\mathrm{pd}(M) = n$ as $\mathrm{pd}(K_{n-1}) = 0$. In particular, we can get a projective resolution

$$0 \to K_{n-1} \to P_{n-1} \to \cdots \to P_0 \to M \to 0$$

which is of length n. In particular, if one has a (possibly infinite) projective resolution M, one can stop after going out n terms, because the kernels will become projective. In other words, the projective resolution can be made to *break off* at the nth term. This applies to *any* projective resolution. Conversely, since any module has a (possibly infinite) projective resolution, we find:

2.3.3 Proposition *We have* $\mathrm{pd}(M) \leqslant n$ *if any projective resolution*

$$\cdots \to P_1 \to P_0 \to M \to 0$$

breaks off at the n*th stage: that is, the kernel of* $P_{n-1} \to P_{n-2}$ *is projective.*

If $\mathrm{pd}(M) \leqslant n$, then by definition we have $\mathrm{Ext}^{n+1}(M, N) = 0$ for *any* module N. By itself, this does not say anything about the Tor functors. However, the criterion for projective dimension enables us to show:

2.3.4 Proposition *If* $\mathrm{pd}(M) \leqslant n$*, then* $\mathrm{Tor}_m(M, N) = 0$ *for* $m > n$*.*

One can define an analog of projective dimension with the Tor functors, called *flat dimension*, and it follows that the flat dimension is at most the projective dimension.

In fact, we have more generally:

2.3.5 Proposition *Let* F *be a right-exact functor on the category of* R*-modules, and let* $\{L_i F\}$ *be its left derived functors. If* $\mathrm{pd}(M) \leqslant n$*, then* $L_i F(M) = 0$ *for* $i > n$*.*

Clearly this implies the claim about Tor functors.

Proof. Recall how $L_i F(M)$ can be computed. Namely, one chooses a projective resolution $P_\bullet \to M$ (any will do), and compute the homology of the complex $F(P_\bullet)$. However, we can choose $P_\bullet \to M$ such that $P_i = 0$ for $i > n$ by proposition 2.3.2. Thus $F(P_\bullet)$ is concentrated in degrees between 0 and n, and the result becomes clear when one takes the homology. □

In general, flat modules are not projective (e.g. $\mathbb{Q}$ is flat, but not projective, over $\mathbb{Z}$), and while one can use projective dimension to bound "flat dimension" (the analog for Tor-vanishing), one cannot use the flat dimension to bound the projective dimension. For a local ring, we will see that it is possible in the next subsec.

Tor and projective dimension

Over a noetherian *local* ring, there is a much simpler way to test whether a finitely generated module is projective. This is a special case of the very general flatness criterion theorem 1.4.10, but we can give a simple direct proof. So we prefer to keep things self-contained.

2.3.6 Theorem *Let* M *be a finitely generated module over the noetherian local ring* $(R, \mathfrak{m})$*, with residue field* $k = R/\mathfrak{m}$*. Then, if* $\mathrm{Tor}_1(M, k) = 0$*,* M *is free.*

In particular, projective—or even flat—modules which are of finite type over R are automatically free. This is a strengthening of the earlier theorem (**??**) that a finitely generated projective module over a local ring is free.

Proof. Indeed, we can find a free module F and a surjection $F \to M$ such that $F \otimes_R k \to M \otimes_R k$ is an isomorphism. To do this, choose elements of M that form a basis of $M \otimes_R k$, and then define a map $F \to M$ via these elements; it is a surjection by Nakayama's lemma.

Let K be the kernel of $F \twoheadrightarrow M$, so there is an exact sequence

$$0 \to K \to F \to M \to 0.$$

We want to show that $K = 0$, which will imply that $M = 0$. By Nakayama's lemma, it suffices to show that $K \otimes_R k = 0$. But we have an exact sequence

$$\mathrm{Tor}_1(M, k) \to K \otimes_R k \to F \otimes_R k \to M \otimes_R k \to 0.$$

The last map is an isomorphism, and $\mathrm{Tor}_1(M, k) = 0$, which implies that $K \otimes_R k = 0$. The result is now proved. □

As a result, we can compute the projective dimension of a module in terms of Tor.

2.3.7 Corollary *Let M be a finitely generated module over the noetherian local ring R with residue field k. Then $\mathrm{pd}(M)$ is the largest integer n such that $\mathrm{Tor}_n(M, k) \neq 0$. It is also the smallest integer n such that $\mathrm{Tor}_{n+1}(M, k) = 0$.*

There is a certain symmetry: if Ext replaces Tor, then one has the definition of depth. We will show later that there is indeed a useful connection between projective dimension and depth.

Proof. We will show that if $\mathrm{Tor}_{n+1}(M, k) = 0$, then $\mathrm{pd}(M) \leqslant n$. This implies the claim, in view of proposition 2.3.4. Choose a (possibly infinite) projective resolution

$$\cdots \to P_1 \to P_0 \to M \to 0.$$

Since R is noetherian, we can assume that each P_i is *finitely generated.*

Write $K_i = \ker(P_i \to P_{i-1})$, as before; these are finitely generated R-modules. We want to show that K_{n-1} is projective, which will establish the claim, as then the projective resolution will "break off." But we have an exact sequence

$$0 \to K_0 \to P_0 \to M \to 0,$$

which shows that $\mathrm{Tor}_n(K_0, k) = \mathrm{Tor}_{n+1}(M, k) = 0$. Using the exact sequencese $0 \to K_i \to P_i \to K_{i-1} \to 0$, we inductively work downwards to get that $\mathrm{Tor}_1(K_{n-1}, k) = 0$. So K_{n-1} is projective by theorem 2.3.6. □

In particular, we find that if $\mathrm{pd}(k) \leqslant n$, then $\mathrm{pd}(M) \leqslant n$ for all M. This is because if $\mathrm{pd}(k) \leqslant n$, then $\mathrm{Tor}_{n+1}(M, k) = 0$ by using the relevant resolution of k (see proposition 2.3.4, but for k).

2.3.8 Corollary *Suppose there exists n such that $\mathrm{Tor}_{n+1}(k, k) = 0$. Then every finitely generated R-module has a finite free resolution of length at most n.*

We have thus seen that k is in some sense the "worst" R-module, in that it is as far from being projective, or that it has the largest projective dimension. We can describe this worst-case behavior with the next concept:

2.3.9 Definition Given a ring R, the **global dimension** is the sup of the projective dimensions of all finitely generated R-modules.

So, to recapitulate: the global dimension of a noetherian local ring R is the projective dimension of its residue field k, or even the *flat* dimension of the residue field.

Minimal projective resolutions

Usually projective resolutions are non-unique; they are only unique up to chain homotopy. We will introduce a certain restriction that enforces uniqueness. These "minimal" projective resolutions will make it extremely easy to compute the groups $\mathrm{Tor}_\bullet(\cdot, k)$.

Let $(R, \mathfrak{m})$ be a local noetherian ring with residue field k, M a finitely generated R-module. All tensor products will be over R.

2.3.10 Definition A projective resolution $P_\bullet \to M$ of finitely generated modules is **minimal** if for each i, the induced map $P_i \otimes k \to P_{i-1} \otimes k$ is zero, and the map $P_0 \otimes k \to M/\mathfrak{m}M$ is an isomorphism.

In other words, the complex $P_\bullet \otimes k$ is isomorphic to $M \otimes k$. This is equivalent to saying that for each i, the map $P_i \to \ker(P_{i-1} \to P_{i-2})$ is an isomorphism modulo $\mathfrak{m}$.

2.3.11 Proposition *Every M (over a local noetherian ring) has a minimal projective resolution.*

Proof. Start with a module M. Then $M/\mathfrak{m}M$ is a finite-dimensional vector space over k, of dimension say d_0. We can choose a basis for that vector space, which we can lift to M. That determines a map of free modules

$$R^{d_0} \to M,$$

which is a surjection by Nakayama's lemma. It is by construction an isomorphism modulo $\mathfrak{m}$. Then define $K = \ker(R^{d_0} \to M)$; this is finitely generated by noetherianness, and we can do the same thing for K, and repeat to get a map $R^{d_1} \twoheadrightarrow K$ which is an isomorphism modulo $\mathfrak{m}$. Then

$$R^{d_1} \to R^{d_0} \to M \to 0$$

is exact, and minimal; we can continue this by the same procedure. ∎

2.3.12 Proposition *Minimal projective resolutions are unique up to isomorphism.*

Proof. Suppose we have one minimal projective resolution:

$$\cdots \to P_2 \to P_1 \to P_0 \to M \to 0$$

and another:

$$\cdots \to Q_2 \to Q_1 \to Q_0 \to M \to 0.$$

There is always a map of projective resolutions $P_* \to Q_*$ by general homological algebra. There is, equivalently, a commutative diagram

$$\begin{tikzcd} \dots \arrow[r] \arrow[d] & P_2 \arrow[r] \arrow[d] & P_1 \arrow[r] \arrow[d] & P_0 \arrow[r] \arrow[d] & M \arrow[r] \arrow[d, "\mathrm{id}"] & 0 \\ \dots \arrow[r] & Q_2 \arrow[r] & Q_1 \arrow[r] & Q_0 \arrow[r] & M \arrow[r] & 0 \end{tikzcd}$$

If both resolutions are minimal, the claim is that this map is an isomorphism. That is, $\phi_i : P_i \to Q_i$ is an isomorphism, for each i.

To see this, note that P_i, Q_i are finite free R-modules.[1] So ϕ_i is an isomorphism iff ϕ_i is an isomorphism modulo the maximal ideal, i.e. if

$$P_i/\mathfrak{m}P_i \to Q_i/\mathfrak{m}Q_i$$

is an isomorphism. Indeed, if ϕ_i is an isomorphism, then its tensor product with $R/\mathfrak{m}$ obviously is an isomorphism. Conversely suppose that the reductions mod $\mathfrak{m}$ make an isomorphism. Then the ranks of P_i, Q_i are the same, and ϕ_i is an n-by-n matrix whose determinant is not in the maximal ideal, so is invertible. This means that ϕ_i is invertible by the usual formula for the inverse matrix.

So we are to check that $P_i/\mathfrak{m}P_i \to Q_i/\mathfrak{m}Q_i$ is an isomorphism for each i. This is equivalent to the assertion that

$$(Q_i/\mathfrak{m}Q_i)^\vee \to (P_i/\mathfrak{m}P_i)^\vee$$

is an isomorphism. But this is the map

$$\hom_R(Q_i, R/\mathfrak{m}) \to \hom_R(P_i, R/\mathfrak{m}).$$

If we look at the chain complexes $\hom(P_*, R/\mathfrak{m}), \hom(Q_*, R/\mathfrak{m})$, the cohomologies compute the Ext groups of $(M, R/\mathfrak{m})$. But all the maps in this chain complex are zero because the resolution is minimal, and we have that the image of P_i is contained in $\mathfrak{m}P_{i-1}$ (ditto for Q_i). So the cohomologies are just the individual terms, and the maps $\hom_R(Q_i, R/\mathfrak{m}) \to \hom_R(P_i, R/\mathfrak{m})$ correspond to the identities on $\mathrm{Ext}^i(M, R/\mathfrak{m})$. So these are isomorphisms.[2] □

2.3.13 Corollary *If $\dots \to P_2 \to P_1 \to P_0 \to M$ is a minimal projective resolution of M, then the ranks* $\mathrm{rank}(P_i)$ *are well-defined (i.e. don't depend on the choice of the minimal resolution).*

Proof. Immediate from the proposition. In fact, the ranks are the dimensions (as $R/\mathfrak{m}$-vector spaces) of $\mathrm{Ext}^i(M, R/\mathfrak{m})$. □

[1] We are using the fact that a finite projective module over a local ring is *free*.

[2] We are sweeping under the rug the statement that Ext can be computed via *any* projective resolution. More precisely, if you take any two projective resolutions, and take the induced maps between the projective resolutions, hom them into $R/\mathfrak{m}$, then the maps on cohomology are isomorphisms.

The Auslander-Buchsbaum formula

2.3.14 Theorem (Auslander-Buschsbaum formula) *Let R be a local noetherian ring, M a finitely generated R-module of finite projective dimension. If* $\mathrm{pd}(R) < \infty$*, then* $\mathrm{pd}(M) = \mathrm{depth}(R) - \mathrm{depth}(M)$*.*

Proof. Induction on $\mathrm{pd}(M)$. When $\mathrm{pd}(M) = 0$, then M is projective, so isomorphic to R^n for some n. Thus $\mathrm{depth}(M) = \mathrm{depth}(R)$.

Assume $\mathrm{pd}(M) > 0$. Choose a surjection $P \twoheadrightarrow M$ and write an exact sequence

$$0 \to K \to P \to M \to 0,$$

where $\mathrm{pd}(K) = \mathrm{pd}(M) - 1$. We also know by induction that

$$\mathrm{pd}(K) = \mathrm{depth}\, R - \mathrm{depth}(K).$$

What we want to prove is that

$$\mathrm{depth}\, R - \mathrm{depth}\, M = \mathrm{pd}(M) = \mathrm{pd}(K) + 1.$$

This is equivalent to wanting know that $\mathrm{depth}(K) = \mathrm{depth}(M) + 1$. In general, this may not be true, though, but we will prove it under minimality hypotheses.

Without loss of generality, we can choose that P is *minimal*, i.e. becomes an isomorphism modulo the maximal ideal $\mathfrak{m}$. This means that the rank of P is $\dim M/\mathfrak{m}M$. So $K = 0$ iff $P \to M$ is an isomorphism; we've assumed that M is not free, so $K \neq 0$.

Recall that the depth of M is the smallest value i such that$\mathrm{Ext}^i(R/\mathfrak{m}, M) \neq 0$. So we should look at the long exact sequence from the above short exact sequence:

$$\mathrm{Ext}^i(R/\mathfrak{m}, P) \to \mathrm{Ext}^i(R/\mathfrak{m}, M) \to \mathrm{Ext}^{i+1}(R/\mathfrak{m}, K) \to \mathrm{Ext}^{i+1}(R/\mathfrak{m}, P).$$

Now P is just a direct sum of copies of R, so $\mathrm{Ext}^i(R/\mathfrak{m}, P)$ and $\mathrm{Ext}^{i+1}(R/\mathfrak{m}, P)$ are zero if $i+1 < \mathrm{depth}\, R$. In particular, if $i+1 < \mathrm{depth}\, R$, then the map $\mathrm{Ext}^i(R/\mathfrak{m}, M) \to \mathrm{Ext}^{i+1}(R/\mathfrak{m}, K)$ is an isomorphism. So we find that $\mathrm{depth}\, M + 1 = \mathrm{depth}\, K$ in this case.

We have seen that *if* $\mathrm{depth}\, K < \mathrm{depth}\, R$, *then* by taking i over all integers $< \mathrm{depth}\, K$, we find that

$$\mathrm{Ext}^i(R/\mathfrak{m}, M) = \begin{cases} 0 & \text{if } i+1 < \mathrm{depth}\, K \\ \mathrm{Ext}^{i+1}(R/\mathfrak{m}, K) & \text{if } i+1 = \mathrm{depth}\, K \end{cases}.$$

In particular, we are **done** unless $\mathrm{depth}\, K \geqslant \mathrm{depth}\, R$. By the inductive hypothesis, this is equivalent to saying that K is projective.

So let us consider the case where K is projective, i.e. $\mathrm{pd}(M) = 1$. We want to show that $\mathrm{depth}\, M = d - 1$ if $d = \mathrm{depth}\, R$. We need a slightly different argument in this case. Let $d = \mathrm{depth}(R) = \mathrm{depth}(P) = \mathrm{depth}(K)$ since P, K are free. We have a short exact sequence

$$0 \to K \to P \to M \to 0$$

and a long exact sequence of Ext groups:

$$0 \to \mathrm{Ext}^{d-1}(R/\mathfrak{m}, M) \to \mathrm{Ext}^d(R/\mathfrak{m}, K) \to \mathrm{Ext}^d(R/\mathfrak{m}, P).$$

We know that $\mathrm{Ext}^d(R/\mathfrak{m}, K)$ is nonzero as K is free and R has depth d. However, $\mathrm{Ext}^i(R/\mathfrak{m}, K) = \mathrm{Ext}^i(R/\mathfrak{m}, P) = 0$ for $i < d$. This implies that $\mathrm{Ext}^{i-1}(R/\mathfrak{m}, M) = 0$ for $i < d$.

We will show:

The map $\mathrm{Ext}^d(R/\mathfrak{m}, K) \to \mathrm{Ext}^d(R/\mathfrak{m}, P)$ is zero.

This will imply that the depth of M is *precisely* $d-1$. This is because the matrix $K \to P$ is given by multiplication by a matrix with coefficients in $\mathfrak{m}$ as $K/\mathfrak{m}K \to P/\mathfrak{m}P$ is zero. In particular, the map on the Ext groups is zero, because it is annihilated by $\mathfrak{m}$. ☐

2.3.15 Example Consider the case of a *regular* local ring R of dimension n. Then $\mathrm{depth}(R) = n$, so we have

$$\mathrm{pd}(M) + \mathrm{depth}(M) = n,$$

for every finitely generated R-module M. In particular, $\mathrm{depth}(M) = n$ if and only if M is free.

2.3.16 Example (The Cohen-Macaulay locus is open) Let R be a regular noetherian ring (i.e. one all of whose localizations are regular). Let M be a finitely generated R-module. We consider the locus $Z \subset \operatorname{Spec} R$ consisting of prime ideals $\mathfrak{p} \in \operatorname{Spec} R$ such that $M_{\mathfrak{p}}$ is a Cohen-Macaulay R-module. We want to show that this is an *open* subset.

Namely, over a local ring $(A, \mathfrak{m})$, define the *codepth* of a finitely generated A-module N as $\mathrm{codepth} N = \dim N - \mathrm{depth}\, N \geqslant 0$; we have that $\mathrm{codepth} N = 0$ if and only if N is Cohen-Macaulay. We are going to show that the function $\mathfrak{p} \mapsto \mathrm{codepth}_{R_{\mathfrak{p}}} M_{\mathfrak{p}}$ is upper semicontinuous on $\operatorname{Spec} R$. To do this, we use the Auslander-Buchsbaum formula $\mathrm{depth}_{R_{\mathfrak{p}}} M_{\mathfrak{p}} = \dim R_{\mathfrak{p}} - \mathrm{pd}_{R_{\mathfrak{p}}} M_{\mathfrak{p}}$ (see example 2.3.15). We will show below that $\mathfrak{p} \mapsto \mathrm{pd}_{R_{\mathfrak{p}}} M_{\mathfrak{p}}$ is upper semi-continuous on $\operatorname{Spec} R$. Thus, we have

$$\mathrm{codepth}_{R_{\mathfrak{p}}} M_{\mathfrak{p}} = -\left(\dim R_{\mathfrak{p}} - \dim_{R_{\mathfrak{p}}} M_{\mathfrak{p}}\right) + \mathrm{pd}_{R_{\mathfrak{p}}} M_{\mathfrak{p}},$$

where the second term is upper semi-continuous. The claim is that the first term is upper semi-continuous. If we consider $\operatorname{supp} M \subset \operatorname{Spec} R$, then the bracketed difference measures the *local codimension* of $\operatorname{supp} M \subset \operatorname{Spec} R$. Namely, $\dim R_{\mathfrak{p}} - \dim \operatorname{supp} M_{\mathfrak{p}}$ is the local codimension because $R_{\mathfrak{p}}$ is regular, and consequently $\operatorname{Spec} R_{\mathfrak{p}}$ is biequidimensional (**To be added: argument**). The local codimension of any set is always lower semi-continuous (**To be added: reference in the section on topological dim**). As a result, the codepth is upper semi-continuous.

We just need to prove the assertion that $\mathfrak{p} \mapsto \mathrm{pd}_{R_{\mathfrak{p}}} M_{\mathfrak{p}}$ is upper semi-continuous. That is, we need to show that if $M_{\mathfrak{p}}$ admits a projective resolution of length n by finitely generated modules, then there is a projective resolution of length n of $M_{\mathfrak{q}}$ for $\mathfrak{q}$ in some Zariski neighborhood. But a projective resolution of $M_{\mathfrak{p}}$ "descends" to a projective (even free) resolution of M_g for some $g \notin \mathfrak{p}$, which gives the result by localization.

If R is the *quotient* of a regular ring, the same result holds (because the Cohen-Macaulay locus behaves properly with respect to quotients). In particular, this result holds for R an affine ring.

2.3.17 Example Let $R = \mathbb{C}[x_1, \dots, x_n]/\mathfrak{p}$ for $\mathfrak{p}$ prime. Choose an injection $R' \to R$ where $R' = \mathbb{C}[y_1, \dots, y_m]$ and R is a finitely generated R'-module. This exists by the Noether normalization lemma.

We wanted to show:

2.3.18 Theorem *R is Cohen-Macaulay*[3] *iff R is a projective R'-module.*

We shall use the fact that projectiveness can be tested locally at every maximal ideal.

Proof. Choose a maximal ideal $\mathfrak{m} \subset R'$. We will show that $R_{\mathfrak{m}}$ is a free $R'_{\mathfrak{m}}$-module via the injection of rings $R'_{\mathfrak{m}} \hookrightarrow R_{\mathfrak{m}}$ (where $R_{\mathfrak{m}}$ is defined as R localized at the multiplicative subset of elements of $R' - \mathfrak{m}$) at each $\mathfrak{m}$ iff Cohen-Macaulayness holds.

Now $R'_{\mathfrak{m}}$ is a regular local ring, so its depth is m. By the Auslander-Buchsbaum formula, $R_{\mathfrak{m}}$ is projective as an $R'_{\mathfrak{m}}$-module iff

$$\operatorname{depth}_{R'_{\mathfrak{m}}} R_{\mathfrak{m}} = m.$$

Now R is a projective module iff the above condition holds for all maximal ideals $\mathfrak{m} \subset R'$. The claim is that this is equivalent to saying that $\operatorname{depth} R_{\mathfrak{n}} = m = \dim R_{\mathfrak{n}}$ for every maximal ideal $\mathfrak{n} \subset R$ (depth over R!).

These two statements are almost the same, but one is about the depth of R as an R-module, and another as an R'-module. □

Issue: There may be several maximal ideals of R lying over the maximal ideal $\mathfrak{m} \subset R'$.

The problem is that $R_{\mathfrak{m}}$ is not generally local, and not generally equal to $R_{\mathfrak{n}}$ if $\mathfrak{n}$ lies over $\mathfrak{m}$. Fortunately, depth makes sense even over semi-local rings (rings with finitely many maximal ideals).

Let us just assume that this does not occur, though. Let us assume that $R_{\mathfrak{m}}$ is a local ring for every maximal ideal $\mathfrak{m} \subset R$. Then we are reduced to showing that if $S = R_{\mathfrak{m}}$, then the depth of S as an $R'_{\mathfrak{m}}$-module is the same as the depth as an $R_{\mathfrak{m}}$-module. That is, the depth doesn't depend too much on the ring, since $R'_{\mathfrak{m}}, R_{\mathfrak{m}}$ are "pretty close." If you believe this, then you believe the theorem, by the first paragraph.

Let's prove this claim in a more general form:

2.3.19 Proposition *Let $\phi : S' \to S$ be a local*[4] *map of local noetherian rings such that S is a finitely generated S'-module. Then, for any finitely generated S-module M,*

$$\operatorname{depth}_S M = \operatorname{depth}_{S'} M.$$

With this, the theorem will be proved.

2.3.20 Remark This result generalizes to the semi-local case, which is how one side-steps the issue above.

[3] That is, its localizations at any prime—or, though we haven't proved yet, at any maximal ideal—are.

[4] I.e. ϕ sends non-units into non-units.

Proof. By induction on $\mathrm{depth}_{S'} M$. There are two cases.

Let $\mathfrak{m}', \mathfrak{m}$ be the maximal ideals of S', S. If $\mathrm{depth}_{S'}(M) > 0$, then there is an element a in $\mathfrak{m}'$ such that

$$M \stackrel{\phi(a)}{\to} M$$

is injective. Now $\phi(a) \in \mathfrak{m}$. So $\phi(a)$ is a non-zero-divisor, and we have an exact sequence

$$0 \to M \stackrel{\phi(a)}{\to} M \to M/\phi(a)M \to 0.$$

Thus we find

$$\mathrm{depth}_S M > 0.$$

Moreover, we find that $\mathrm{depth}_S M = \mathrm{depth}_S(M/\phi(a)M)+1$ and $\mathrm{depth}_{S'} M = \mathrm{depth}_{S'}(M/\phi(a)M))+1$. The inductive hypothesis now tells us that

$$\mathrm{depth}_S M = \mathrm{depth}_{S'} M.$$

The hard case is where $\mathrm{depth}_{S'} M = 0$. We need to show that this is equivalent to $\mathrm{depth}_S M = 0$. So we know at first that $\mathfrak{m}' \in \mathrm{Ass}(M)$. That is, there is an element $x \in M$ such that $\mathrm{Ann}_{S'}(x) = \mathfrak{m}'$. Now $\mathrm{Ann}_S(x) \subsetneq S$ and contains $\mathfrak{m}'S$.

$Sx \subset M$ is a submodule, surjected onto by S by the map $a \to ax$. This map actually, as we have seen, factors through $S/\mathfrak{m}'S$. Here S is a finite S'-module, so $S/\mathfrak{m}'S$ is a finite $S'/\mathfrak{m}'$-module. In particular, it is a finite-dimensional vector space over a field. It is thus a local artinian ring. But Sx is a module over this local artinian ring. It must have an associated prime, which is a maximal ideal in $S/\mathfrak{m}'S$. The only maximal ideal can be $\mathfrak{m}/\mathfrak{m}'S$. It follows that $\mathfrak{m} \in \mathrm{Ass}(Sx) \subset \mathrm{Ass}(M)$.

In particular, $\mathrm{depth}_S M = 0$ too, and we are done. □

2.3.21 Remark (comment) We shall eventually prove:

2.3.22 Proposition *Let $R = \mathbb{C}[X_1, \dots, X_n]/\mathfrak{p}$ for $\mathfrak{p}$ prime. Choose an injective map $\mathbb{C}[y_1, \dots, y_n] \hookrightarrow R$ making R a finite module. Then R is Cohen-Macaulay iff R is projective as a module over $\mathbb{C}[y_1, \dots, y_n]$.*[5]

The picture is that the inclusion $\mathbb{C}[y_1, \dots, y_m] \hookrightarrow \mathbb{C}[x_1, \dots, x_n]/\mathfrak{p}$ corresponds to a map

$$X \to \mathbb{C}^m$$

for $X = V(\mathfrak{p}) \subset \mathbb{C}^n$. This statement of freeness is a statement about how the fibers of this finite map stay similar in some sense.

[5] In fact, this is equivalent to freeness, although we will not prove it. Any projective finite module over a polynomial ring over a field is free, though this is a hard theorem.

2.4. Serre's criterion and its consequences

We would like to prove Serre's criterion for regularity.

2.4.1 Theorem *Let $(R, \mathfrak{m})$ be a local noetherian ring. Then R is regular iff $R/\mathfrak{m}$ has finite projective dimension. In this case,* $\mathrm{pd}(R/\mathfrak{m}) = \dim R$.

To be added: proof

First consequences

2.4.2 Proposition *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a flat, local homomorphism of noetherian local rings. If S is regular, so is R.*

Proof. Let $n = \dim S$. Let M be a finitely generated R-module, and consider a resolution

$$P_n \to P_{n-1} \to \cdots \to P_0 \to M \to 0,$$

where all the $\{P_i\}$ are finite free R-modules. If we can show that the kernel of $P_n \to P_{n-1}$ is projective, then it will follow that M has finite projective dimension. Since M was arbitrary, it will follow that R is regular too, by Serre's criterion.

Let K be the kernel, so there is an exact sequence

$$0 \to K \to P_n \to P_{n-1} \to \cdots \to P_0 \to M \to 0,$$

which we can tensor with S, by flatness:

$$0 \to K \otimes_R S \to P_n \otimes_R S \to P_{n-1} \otimes_R S \to \cdots \to P_0 \otimes_R S \to M \otimes_R S \to 0.$$

Because any finitely generated S-module has projective dimension $\leqslant n$, it follows that $K \otimes_R S$ is projective, and in particular flat.

But now S is *faithfully flat* over R (see **??**), and it follows that K is R-flat. Thus K is projective over R, proving the claim. □

2.4.3 Theorem *The localization of a regular local ring at a prime ideal is regular.*

Geometrically, this means that to test whether a nice scheme (e.g. a variety) is regular (i.e., all the local rings are regular), one only has to test the *closed* points.

Proof. Let $(R, \mathfrak{m})$ be a regular local ring. Let $\mathfrak{p} \in \operatorname{Spec} R$ be a prime ideal; we wish to show that $R_{\mathfrak{p}}$ is regular. To do this, let M be a finitely generated $R_{\mathfrak{p}}$-module. Then we can find a finitely generated R-submodule $N \subset M$ such that the natural map $N_{\mathfrak{p}} \to M$ is an isomorphism. If we take a finite free resolution of N by R-modules and localize at $\mathfrak{p}$, we get a finite free resolution of M by $R_{\mathfrak{p}}$-modules.

It now follows that M has finite projective dimension as an $R_{\mathfrak{p}}$-module. By Serre's criterion, this implies that $R_{\mathfrak{p}}$ is regular. □

Regular local rings are factorial

We now aim to prove that a regular local ring is factorial.

First, we need:

2.4.4 Definition Let R be a noetherian ring and M a f.gen. R-module. Then M is **stably free** if $M \oplus R^k$ is free for some k.

Stably free obviously implies "projective." Free implies stably free, clearly—take $k = 0$. Over a local ring, a finitely generated projective module is free, so all three notions are equivalent. Over a general ring, these notions are generally different.

We will need the following lemma:

2.4.5 Lemma *Let M be an R-module with a finite free resolution. If M is projective, it is stably free.*

Proof. There is an exact sequence

$$0 \to F_k \to F_{k-1} \to \cdots \to F_1 \to F_0 \to M \to 0$$

with the F_i free and finitely generated, by assumption.

We induct on the length k of the resolution. We know that if N is the kernel of $F_0 \to M$, then N is projective (as the sequence $0 \to N \to F_0 \to M \to 0$ splits) so there is a resolution

$$0 \to F_k \to \cdots \to F_1 \to N \to 0.$$

By the inductive hypothesis, N is stably free. So there is a free module R^d such that $N \oplus R^d$ is free.

We know that $M \oplus N = F_0$ is free. Thus $M \oplus N \oplus R^d = F_0 \oplus R^d$ is free and $N \oplus R^d$ is free. Thus M is stably free. ☐

2.4.6 Remark Stably freeness does **not** generally imply freeness, though it does over a local noetherian ring.

Nonetheless,

2.4.7 Proposition *Stably free does imply free for invertible modules.*

Proof. Let I be stably free and invertible. We must show that $I \simeq R$. Without loss of generality, we can assume that $\operatorname{Spec} R$ is connected, i.e. R has no nontrivial idempotents. We will assume this in order to talk about the **rank** of a projective module.

We know that $I \oplus R^n \simeq R^m$ for some m. We know that $m = n + 1$ by localization. So $I \oplus R^n \simeq R^{n+1}$ for some n. We will now need to construct the **exterior powers**, for which we digress:

2.4.8 Definition Let R be a commutative ring and M an R-module. Then $\wedge M$, the **exterior algebra on** M, is the free (noncommutative) graded R-algebra generated by M (with product $\wedge$) with just enough relations such that $\wedge$ is anticommutative (and, *more strongly*, $x \wedge x = 0$ for x degree one).

Clearly $\wedge M$ is a quotient of the **tensor algebra** $T(M)$, which is by definition $R \oplus M \oplus M \otimes M \oplus \cdots \oplus M^{\otimes n} \oplus \ldots$. The tensor algebra is a graded R-algebra in an obvious way: $(x_1 \otimes \cdots \otimes x_a).(y_1 \otimes \cdots \otimes y_b) = x_1 \otimes \cdots \otimes x_a \otimes y_1 \otimes \cdots \otimes y_b$. This is an associative R-algebra. Then

$$\wedge M = T(M)/(x \otimes x,\ x, y \in M).$$

The grading on $\wedge M$ comes from the grading of $T(M)$.

We are interested in basically one example:

2.4.9 Example Say $M = R^m$. Then $\wedge^m M = R$. If $e_1, \ldots, e_m \in M$ are generators, then $e_1 \wedge \cdots \wedge e_m$ is a generator. More generally, $\wedge^k M$ is free on $e_{i_1} \wedge \cdots \wedge e_{i_k}$ for $i_1 < \cdots < i_k$.

We now make:

2.4.10 Definition If M is a projective R-module of rank n, then

$$\det(M) = \wedge^n M.$$

If M is free, then $\det(M)$ is free of rank one. So, as we see by localization, $\det(M)$ is always an invertible module for M locally free (i.e. projective) and $\wedge^{n+1} M = 0$.

2.4.11 Lemma $\det(M \oplus N) = \det M \otimes \det N$.

Proof. This isomorphism is given by wedging $\wedge^{\text{top}} M \otimes \wedge^{\text{top}} N \to \wedge^{\text{top}}(M \oplus N)$. This is easily checked for oneself. ☐

Anyway, let us finally go back to the proof. If $I \oplus R^n = R^{n+1}$, then taking determinants shows that

$$\det I \otimes R = R,$$ ☐

so $\det I = R$. But this is I as I is of rank one. So I is free.

2.4.12 Theorem *A regular local ring is factorial.*

Let R be a regular local ring of dimension n. We want to show that R is factorial. Choose a prime ideal $\mathfrak{p}$ of height one. We'd like to show that $\mathfrak{p}$ is principal.

Proof. Induction on n. If $n = 0$, then we are done—we have a field.

If $n = 1$, then a height one prime is maximal, hence principal, because regularity is equivalent to the ring's being a DVR.

Assume $n > 1$. The prime ideal $\mathfrak{p}$ has height one, so it is contained in a maximal ideal $\mathfrak{m}$. Note that $\mathfrak{m}^2 \subset \mathfrak{m}$ as well. I claim that there is an element x of $\mathfrak{m} - \mathfrak{p} - \mathfrak{m}^2$. This follows as an argument like prime avoidance. To see that x exists, choose $x_1 \in \mathfrak{m} - \mathfrak{p}$ and $x_2 \in \mathfrak{m} - \mathfrak{m}^2$. We are done unless $x_1 \in \mathfrak{m}^2$ and $x_2 \in \mathfrak{p}$ (or we could take x to be x_1 or x_2). In this case, we just take $x = x_1 + x_2$.

So choose $x \in \mathfrak{m} - \mathfrak{p} - \mathfrak{m}^2$. Let us examine the ring $R_x = R[1/x]$, which contains an ideal $\mathfrak{p}[x^{-1}]$. This is a proper ideal as $x \notin \mathfrak{p}$. Now $R[1/x]$ is regular (i.e. its localizations at primes are regular local). The dimension, however, is of dimension less than n since by inverting x we have removed $\mathfrak{m}$. By induction we can assume that R_x is locally factorial.

Now $\mathfrak{p}R_x$ is prime and of height one, so it is invertible as R_x is locally factorial. In particular it is projective.

But $\mathfrak{p}$ has a finite resolution by R-modules (by regularity), so $\mathfrak{p}R_x$ has a finite free resolution. In particular, $\mathfrak{p}R_x$ is stably free and invertible, hence free. Thus $\mathfrak{p}R_x$ is **principal**.

We want to show that $\mathfrak{p}$ is principal, not just after localization. We know that there is a $y \in \mathfrak{p}$ such that y generates $\mathfrak{p}R_x$. Choose y such that $(y) \subset \mathfrak{p}$ is as large as possible. We can do this since R is noetherian. This implies that $x \nmid y$ because otherwise we could use y/x instead of y.

We shall now show that

$$\mathfrak{p} = (y).$$

So suppose $z \in \mathfrak{p}$. We know that y generates $\mathfrak{p}$ **after** x **is inverted.** In particular, $z \in \mathfrak{p}R_x$. That is, $zx^a \in (y)$ for a large. That is, we can write

$$zx^a = yw, \quad \text{for some } w \in R.$$

We chose x such that $x \notin \mathfrak{m}^2$. In particular, $R/(x)$ is regular, hence an integral domain; i.e. x is a prime element. We find that x must divide one of y, w if $a > 0$. But we know that $x \nmid y$, so $x \mid w$. Thus $w = w'x$ for some x. We find that, cancelling x,

$$zx^{a-1} = yw'$$

and we can repeat this argument over and over until we find that

$$z \in (y).$$

□

V.3. Étale, unramified, and smooth morphisms

In this chapter, we shall introduce three classes of morphisms of rings defined by lifting properties and study their properties. Although in the case of morphisms of finite presentation, the three types of morphisms (unramified, smooth, and étale) can be defined directly (without lifting properties), in practice, in algebraic geometry, the functorial criterion given by lifts matter: if one wants to show an algebra is representable, then one can just study the *corepresentable functor*, which may be more accessible.

3.1. Unramified morphisms

Definition

Formal étaleness, smoothness, and unramifiedness all deal with the existence or uniqueness of liftings under nilpotent extensions. We start with formal unramifiedness.

3.1.1 Definition Let $R \to S$ be a ring map. We say S is **formally unramified over** R if for every commutative solid diagram

$$\begin{tikzcd} S \arrow[r] \arrow[dr, dashed] & A/I \\ R \arrow[u] \arrow[r] & A \arrow[u] \end{tikzcd} \tag{3.1.1}$$

where $I \subset A$ is an ideal of square zero, there exists at most one dotted arrow making the diagram commute.

We say that S is **unramified over** R if S is formally unramified over R and is a finitely generated R-algebra.

In other words, an R-algebra S is formally unramified if and only if whenever A is an R-algebra and $I \subset A$ an ideal of square zero, the map of sets

$$\hom_R(S, A) \to \hom_R(S, A/I)$$

is injective. Restated again, for such A, I, there is *at most one* lift of a given R-homomorphism $S \to A/I$ to $S \to A$. This is a statement purely about the associated "functor of points." Namely, let S be an R-algebra, and consider the functor $F : R$–**alg** → **Sets** given by $F(X) = \hom_R(S, X)$. This is the "functor of points." Then S is formally unramified over R if $F(A) \to F(A/I)$ is injective for each A, I as above.

The intuition is that maps from S into T are like "tangent vectors," and consequently the condition geometrically means something like that tangent vectors can be lifted uniquely: that is, the associated map is an immersion. More formally, if $R \to S$ is a morphism of algebras of finite type over $\mathbb{C}$, which corresponds to a map $\operatorname{Spec} S \to \operatorname{Spec} R$ of *smooth* varieties (this is a condition on R, S!), then $R \to S$ is unramified if and only if the associated map of complex manifolds is an immersion. (We are not proving this, just stating it for intuition.)

Note also that we can replace "I of square zero" with the weaker condition "I nilpotent." That is, the map $R \to S$ (if it is formally unramified) still has the same lifting property. This follows because one can factor $A \to A/I$ into the *finite* sequence $\cdots \to A/I^{n+1} \to A/I^n \to \cdots \to A/I$, and each step is a square-zero extension.

We now show that the module of Kähler differentials provides a simple criterion for an extension to be formally unramified.

3.1.2 Proposition *An R-algebra S is formally unramified if and only if $\Omega_{S/R} = 0$.*

Suppose R, S are both algebras over some smaller ring k. Then there is an exact sequence

$$\Omega_{R/k} \otimes_R S \to \Omega_{S/k} \to \Omega_{S/R} \to 0,$$

and consequently, we see that formal unramifiedness corresponds to surjectivity of the map on "cotangent spaces" $\Omega_{R/k} \otimes_R S \to \Omega_{S/k}$. This is part of the intuition that formally unramified maps are geometrically like immersions (since surjectivity on the cotangent spaces corresponds to injectivity on the tangent spaces).

Proof. Suppose first $\Omega_{S/R} = 0$. This is equivalent to the statement that *any* R-derivation of S into an S-module is trivial, because $\Omega_{S/R}$ is the recipient of the "universal" R-derivation. If given an R-algebra T with an ideal $I \subset T$ of square zero and a morphism

$$S \to T/I, \qquad \square$$

and two liftings $f, g : S \to T$, then we find that $f - g$ maps S into I. Since T/I is naturally an S-algebra, it is easy to see (since I has square zero) that I is naturally an S-module and $f - g$ is an R-derivation $S \to I$. Thus $f - g \equiv 0$ and $f = g$.

Conversely, suppose S has the property that liftings in (3.1.1) are unique. Consider the S-module $T = S \oplus \Omega_{S/R}$ with the multiplicative structure $(a, a')(b, b') = (ab, ab' + a'b)$ that makes it into an algebra. (This is a general construction one can do with an S-module M: $S \oplus M$ is an algebra where M becomes an ideal of square zero.)

Consider the ideal $\Omega_{S/R} \subset T$, which has square zero; the quotient is S. We will find two liftings of the identity $S \to S$. For the first, define $S \to T$ sending $s \to (s, 0)$. For the second, define $S \to T$ sending $s \to (s, ds)$; the derivation property of b shows that this is a morphism of algebras.

By the lifting property, the two morphisms $S \to T$ are equal. In particular, the map $S \to \Omega_{S/R}$ sending $s \to ds$ is trivial. This implies that $\Omega_{S/R} = 0$.

Here is the essential point of the above argument. Let $I \subset T$ be an ideal of square zero in the R-algebra T. Suppose given a homomorphism $g : S \to T/I$. Then the set of lifts $S \to T$ of g (which are R-algebra morphisms) is either empty or a torsor over $\mathrm{Der}_R(S, I)$ (by adding a derivation to a homomorphism). Note that I is naturally a T/I-module (because $I^2 = 0$), and hence an S-module by g.

This means that if the object $\mathrm{Der}_R(S, I)$ is trivial, then injectivity of the above map must hold. Conversely, if injectivity of the above map always holds (i.e. S is formally unramified), then we must have $\mathrm{Der}_R(S, I) = 0$ for all such $I \subset T$; since we can obtain any S-module in this manner, it follows that there is no such thing as a nontrivial R-derivation out of S.

We next show that formal unramifiedness is a local property.

3.1.3 Lemma *Let $R \to S$ be a ring map. The following are equivalent:*

1. *$R \to S$ is formally unramified,*
2. *$R \to S_{\mathfrak{q}}$ is formally unramified for all primes $\mathfrak{q}$ of S, and*
3. *$R_{\mathfrak{p}} \to S_{\mathfrak{q}}$ is formally unramified for all primes $\mathfrak{q}$ of S with $\mathfrak{p} = R \cap \mathfrak{q}$.*

Proof. We have seen in proposition 3.1.2 that (1) is equivalent to $\Omega_{S/R} = 0$. Similarly, since Kähler differentials localize, we see that (2) and (3) are equivalent to $(\Omega_{S/R})_{\mathfrak{q}} = 0$ for all $\mathfrak{q}$. As a result, the statement of this lemma is simply the fact that an S-module is zero if and only if all its localizations at prime ideals are zero. □

We shall now give the typical list of properties ("le sorite") of unramified morphisms.

3.1.4 Proposition *Any map $R \to R_f$ for $f \in R$ is unramified. More generally, a map from a ring to any localization is* formally *unramified, but not necessarily unramified.*

Proof. Indeed, we know that $\Omega_{R/R} = 0$ and $\Omega_{R_f/R} = (\Omega_{R/R})_f = 0$, and the map is clearly of finite type. □

3.1.5 Proposition *A surjection of rings is unramified. More generally, a categorical epimorphism of rings is formally unramified.*

Proof. Obvious from the lifting property: if $R \to S$ is a categorical epimorphism, then given any R-algebra T, there can be *at most one* map of R-algebras $S \to T$ (regardless of anything involving square-zero ideals). □

In the proof of proposition 3.1.5, we could have alternatively argued as follows. If $R \to S$ is an epimorphism in the category of rings, then $S \otimes_R S \to S$ is an isomorphism. This is a general categorical fact, the dual of which for monomorphisms is perhaps simpler: if $X \to Y$ is a monomorphism of objects in any category, then $X \to X \times_Y X$ is an isomorphism. See **??**. By the alternate construction of $\Omega_{S/R}$ (proposition 9.2.17), it follows that this must vanish.

3.1.6 Proposition *If $R \to S$ and $S \to T$ are unramified (resp. formally unramified), so is $R \to T$.*

Proof. Since morphisms of finite type are preserved under composition, we only need to prove the result about formally unramified maps. So let $R \to S, S \to T$ be formally unramified. We need to check that $\Omega_{T/R} = 0$. However, we have an exact sequence (see proposition 9.2.9):

$$\Omega_{S/R} \otimes_S T \to \Omega_{T/R} \to \Omega_{T/S} \to 0,$$

and since $\Omega_{S/R} = 0, \Omega_{T/S} = 0$, we find that $\Omega_{T/R} = 0$. This shows that $R \to T$ is formally unramified. ∎

More elegantly, we could have proved this by using the lifting property (and this is what we will do for formal étaleness and smoothness). Then this is simply a formal argument.

3.1.7 Proposition *If $R \to S$ is unramified (resp. formally unramified), so is $R' \to S' = S \otimes_R R'$ for any R-algebra R'.*

Proof. This follows from the fact that $\Omega_{S'/R'} = \Omega_{S/R} \otimes_S S'$ (see proposition 9.2.14). Alternatively, it can be checked easily using the lifting criterion. For instance, suppose given an R'-algebra T and an ideal $I \subset T$ of square zero. We want to show that a morphism of R'-algebras $S' \to T/I$ lifts in at most one way to a map $S' \to T$. But if we had two distinct liftings, then we could restrict to S to get two liftings of $S \to S' \to T/I$. These are easily seen to be distinct, a contradiction as $R \to S$ was assumed formally unramified. ∎

In fact, the question of what unramified morphisms look like can be reduced to the case where the ground ring is a *field* in view of the previous and the following result. Given $\mathfrak{p} \in \operatorname{Spec} R$, we let $k(\mathfrak{p})$ to be the residue field of $R_{\mathfrak{p}}$.

3.1.8 Proposition *Let $\phi : R \to S$ be a morphism of finite type. Then ϕ is unramified if and only if for every $\mathfrak{p} \in \operatorname{Spec} R$, we have $k(\mathfrak{p}) \to S \otimes_R k(\mathfrak{p})$ unramified.*

The classification of unramified extensions of a field is very simple, so this will be useful.

Proof. One direction is clear by proposition 3.1.7. For the other, suppose $k(\mathfrak{p}) \to S \otimes_R k(\mathfrak{p})$ unramified for all $\mathfrak{p} \subset R$. We then know that $\Omega_{S/R} \otimes_R k(\mathfrak{p}) = \Omega_{S \otimes_R k(\mathfrak{p})/k(\mathfrak{p})} = 0$ for all $\mathfrak{p}$. By localization, it follows that

$$\mathfrak{p}\Omega_{S_{\mathfrak{q}}/R_{\mathfrak{p}}} = \Omega_{S_{\mathfrak{q}}/R_{\mathfrak{p}}} = \Omega_{S_{\mathfrak{q}}/R} \tag{3.1.2}$$

for any $\mathfrak{q} \in \operatorname{Spec} S$ lying over $\mathfrak{p}$.

Let $\mathfrak{q} \in \operatorname{Spec} S$. We will now show that $(\Omega_{S/R})_{\mathfrak{q}} = 0$. Given this, we will find that $\Omega_{S/R} = 0$, which will prove the assertion of the corollary. Indeed, let $\mathfrak{p} \in \operatorname{Spec} R$ be the image of $\mathfrak{q}$, so that there is a *local* homomorphism $R_{\mathfrak{p}} \to S_{\mathfrak{q}}$. By (3.1.2), we find that

$$\mathfrak{q}\Omega_{S_{\mathfrak{q}}/R} = \Omega_{S_{\mathfrak{q}}/R}.$$

and since $\Omega_{S_{\mathfrak{q}}/R}$ is a finite $S_{\mathfrak{q}}$-module (proposition 9.3.15), Nakayama's lemma now implies that $\Omega_{S_{\mathfrak{q}}/R} = 0$, proving what we wanted. ∎

The following is simply a combination of the various results proved:

3.1.9 Corollary *Let $A \to B$ be a formally unramified ring map.*

1. *For $S \subset A$ a multiplicative subset, $S^{-1}A \to S^{-1}B$ is formally unramified.*
2. *For $S \subset B$ a multiplicative subset, $A \to S^{-1}B$ is formally unramified.*

Unramified extensions of a field

Motivated by proposition 3.1.8, we classify unramified morphisms out of a field; we are going to see that these are just finite products of separable extensions. Let us first consider the case when the field is *algebraically closed.*

3.1.10 Proposition *Suppose k is algebraically closed. If A is an unramified k-algebra, then A is a product of copies of k.*

Proof. Let us show first that A is necessarily finite-dimensional. If not,

So let us now assume that A is finite-dimensional over k, hence *artinian.* Then A is a direct product of artinian local k-algebras. Each of these is unramified over k. So we need to study what local, artinian, unramified extensions of k look like; we shall show that any such is isomorphic to k with:

3.1.11 Lemma *A finite-dimensional, local k-algebra which is unramified over k (for k algebraically closed) is isomorphic to k.*

Proof. First, if $\mathfrak{m} \subset A$ is the maximal ideal, then $\mathfrak{m}$ is nilpotent, and $A/\mathfrak{m} \simeq k$ by the Hilbert Nullstellensatz. Thus the ideal $\mathfrak{M} = \mathfrak{m} \otimes A + A \otimes \mathfrak{m} \subset A \otimes_k A$ is nilpotent and $(A \otimes_k A)/\mathfrak{M} = k \otimes_k k = k$. In particular, $\mathfrak{M}$ is maximal and $A \otimes_k A$ is also local. (We could see this as follows: A is associated to a one-point variety, so the fibered product $\operatorname{Spec} A \times_k \operatorname{Spec} A$ is also associated to a one-point variety. It really does matter that we are working over an algebraically closed field here!)

By assumption, $\Omega_{A/k} = 0$. So if $I = \ker(A \otimes_k A \to A)$, then $I = I^2$. But from **??**, we find that if we had $I \neq 0$, then $\operatorname{Spec} A \otimes_k A$ would be disconnected. This is clearly false (a local ring has no nontrivial idempotents), so $I = 0$ and $A \otimes_k A \simeq A$. Since A is finite-dimensional over k, necessarily $A \simeq k$. ☐

Now let us drop the assumption of algebraic closedness to get:

3.1.12 Theorem *An unramified k-algebra for k any field is isomorphic to a product $\prod k_i$ of finite separable extensions k_i of k.*

Proof. Let k be a field, and $\overline{k}$ its algebraic closure. Let A be an unramified k-algebra. Then $A \otimes_k \overline{k}$ is an unramified $\overline{k}$-algebra by proposition 3.1.7, so is a finite product of copies of $\overline{k}$. It is thus natural that we need to study tensor products of fields to understand this problem.

3.1.13 Lemma *Let E/k be a finite extension, and L/k any extension. If E/k is separable, then $L \otimes_k E$ is isomorphic (as a L-algebra) to a product of copies of separable extensions of L.*

Proof. By the primitive element theorem, we have $E = k(\alpha)$ for some $\alpha \in E$ satisfying a separable irreducible polynomial $P \in k[X]$. Thus

$$E = k[X]/(P),$$

so

$$E \otimes_k L = L[X]/(P).$$

But P splits into several irreducible factors $\{P_i\}$ in $L[X]$, no two of which are the same by separability. Thus by the Chinese remainder theorem,

$$E \otimes_k L = L(X)/(\prod P_i) = \prod L[X]/(P_i),$$

and each $L[X]/(P_i)$ is a finite separable extension of L. ∎

As a result of this, we can easily deduce that any k-algebra of the form $A = \prod k_i$ for the k_i separable over k is unramified. Indeed, we have

$$\Omega_{A/k} \otimes_k \overline{k} = \Omega_{A \otimes_k \overline{k}/\overline{k}},$$

so it suffices to prove that $A \otimes_k \overline{k}$ is unramified over $\overline{k}$. However, from lemma 3.1.13, $A \otimes_k \overline{k}$ is isomorphic as a $\overline{k}$-algebra to a product of copies of $\overline{k}$. Thus $A \otimes_k \overline{k}$ is obviously unramified over $\overline{k}$.

On the other hand, suppose A/k is unramified. We shall show it is of the form given as in the theorem. Then $A \otimes_k \overline{k}$ is unramified over $\overline{k}$, so it follows by proposition 3.1.10 that A is finite-dimensional over k. In particular, A is *artinian*, and thus decomposes as a product of finite-dimensional unramified k-algebras.

We are thus reduced to showing that a local, finite-dimensional k-algebra that is unramified is a separable extension of k. Let A be one such. Then A can have no nilpotents because then $A \otimes_k \overline{k}$ would have nilpotents, and could not be isomorphic to a product of copies of $\overline{k}$. Thus the unique maximal ideal of A is zero, and A is a field. We need only show that A is separable over k. This is accomplished by:

3.1.14 Lemma *Let E/k be a finite inseparable extension. Then $E \otimes_k \overline{k}$ contains nonzero nilpotents.*

Proof. There exists an $\alpha \in E$ which is inseparable over k, i.e. whose minimal polynomial has multiple roots. Let $E' = k(\alpha)$. We will show that $E' \otimes_k \overline{k}$ has nonzero nilpotents; since the map $E' \otimes_k \overline{k} \to E \otimes_k \overline{k}$ is an injection, we will be done. Let P be the minimal polynomial of α, so that $E' = k[X]/(P)$. Let $P = \prod P_i^{e_i}$ be the factorization of P in $\overline{k}$ for the $P_i \in \overline{k}[X]$ irreducible (i.e. linear). By assumption, one of the e_i is greater than one. It follows that

$$E' \otimes_k \overline{k} = \overline{k}[X]/(P) = \prod \overline{k}[X]/(P_i^{e_i}) \qquad \square$$

has nilpotents corresponding to the e_i's that are greater than one. ∎

3.1.15 Remark (comment) We now come to the result that explains why the present theory is connected with Zariski's Main Theorem.

3.1.16 Corollary *An unramified morphism $A \to B$ is quasi-finite.*

Proof. Recall that a morphism of rings is *quasi-finite* if the associated map on spectra is. Equivalently, the morphism must be of finite type and have finite fibers. But by assumption $A \to B$ is of finite type. Moreover, if $\mathfrak{p} \in \operatorname{Spec} A$ and $k(\mathfrak{p})$ is the residue field, then $k(\mathfrak{p}) \to B \otimes_A k(\mathfrak{p})$ is *finite* by the above results, so the fibers are finite. ∎

Conormal modules and universal thickenings

It turns out that one can define the first infinitesimal neighbourhood not just for a closed immersion of schemes, but already for any formally unramified morphism. This is based on the following algebraic fact.

3.1.17 Lemma *Let $R \to S$ be a formally unramified ring map. There exists a surjection of R-algebras $S' \to S$ whose kernel is an ideal of square zero with the following universal property: Given any commutative diagram*

$$\begin{tikzcd} S \arrow[r, "a"'] & A/I \\ R \arrow[u] \arrow[r, "b"] & A \arrow[u] \end{tikzcd}$$

where $I \subset A$ is an ideal of square zero, there is a unique R-algebra map $a' : S' \to A$ such that $S' \to A \to A/I$ is equal to $S' \to S \to A$.

Proof. Choose a set of generators $z_i \in S$, $i \in I$ for S as an R-algebra. Let $P = R[\{x_i\_i \in I]$ denote the polynomial ring on generators x_i, $i \in I$. Consider the R-algebra map $P \to S$ which maps x_i to z_i. Let $J = \mathrm{Ker}(P \to S)$. Consider the map

$$\mathrm{d} : J/J^2 \longrightarrow \Omega_{P/R} \otimes_P S$$

see **??**. This is surjective since $\Omega_{S/R} = 0$ by assumption, see **??**. Note that $\Omega_{P/R}$ is free on $\mathrm{d}x_i$, and hence the module $\Omega_{P/R} \otimes_P S$ is free over S. Thus we may choose a splitting of the surjection above and write

$$J/J^2 = K \oplus \Omega_{P/R} \otimes_P S$$

Let $J^2 \subset J' \subset J$ be the ideal of P such that J'/J^2 is the second summand in the decomposition above. Set $S' = P/J'$. We obtain a short exact sequence

$$0 \to J/J' \to S' \to S \to 0$$

and we see that $J/J' \cong K$ is a square zero ideal in S'. Hence

$$\begin{tikzcd} S \arrow[r, "1"'] & S \\ R \arrow[u] \arrow[r] & S' \arrow[u] \end{tikzcd}$$

is a diagram as above. In fact we claim that this is an initial object in the category of diagrams. Namely, let $(I \subset A, a, b)$ be an arbitrary diagram. We may choose an R-algebra map $\beta : P \to A$ such that

$$\begin{tikzcd} S \arrow[r, "1"'] & S \arrow[r, "a"'] & A/I \\ R \arrow[u] \arrow[r] \arrow[rr, bend right, "b"'] & P \arrow[u] \arrow[r, "\beta"] & A \arrow[u] \end{tikzcd}$$

is commutative. Now it may not be the case that $\beta(J') = 0$, in other words it may not be true that β factors through $S' = P/J'$. But what is clear is that $\beta(J') \subset I$ and since $\beta(J) \subset I$ and $I^2 = 0$

we have $\beta(J^2) = 0$. Thus the "obstruction" to finding a morphism from $(J/J' \subset S', 1, R \to S')$ to $(I \subset A, a, b)$ is the corresponding S-linear map $\overline{\beta} : J'/J^2 \to I$. The choice in picking β lies in the choice of $\beta(x_i)$. A different choice of β, say β', is gotten by taking $\beta'(x_i) = \beta(x_i) + \delta_i$ with $\delta_i \in I$. In this case, for $g \in J'$, we obtain

$$\beta'(g) = \beta(g) + \sum\nolimits_i \delta_i \frac{\partial g}{\partial x_i}.$$

Since the map $\mathrm{d}|_{J'/J^2} : J'/J^2 \to \Omega_{P/R} \otimes_P S$ given by $g \mapsto \frac{\partial g}{\partial x_i}\mathrm{d}x_i$ is an isomorphism by construction, we see that there is a unique choice of $\delta_i \in I$ such that $\beta'(g) = 0$ for all $g \in J'$. (Namely, δ_i is $-\overline{\beta}(g)$ where $g \in J'/J^2$ is the unique element with $\frac{\partial g}{\partial x_j} = 1$ if $i = j$ and 0 else.) The uniqueness of the solution implies the uniqueness required in the lemma. □

In the situation of Lemma 3.1.17 the R-algebra map $S' \to S$ is unique up to unique isomorphism.

3.1.18 Definition Let $R \to S$ be a formally unramified ring map.

1. The *universal first order thickening* of S over R is the surjection of R-algebras $S' \to S$ of Lemma 3.1.17.
2. The *conormal module* of $R \to S$ is the kernel I of the universal first order thickening $S' \to S$, seen as a S-module.

We often denote the conormal module $C_{S/R}$ in this situation.

3.1.19 Lemma *Let $I \subset R$ be an ideal of a ring. The universal first order thickening of R/I over R is the surjection $R/I^2 \to R/I$. The conormal module of R/I over R is $C_{(R/I)/R} = I/I^2$.*

Proof. Omitted. □

3.1.20 Lemma *Let $A \to B$ be a formally unramified ring map. Let $\varphi : B' \to B$ be the universal first order thickening of B over A.*

1. *Let $S \subset A$ be a multiplicative subset. Then $S^{-1}B' \to S^{-1}B$ is the universal first order thickening of $S^{-1}B$ over $S^{-1}A$. In particular $S^{-1}C_{B/A} = C_{S^{-1}B/S^{-1}A}$.*
2. *Let $S \subset B$ be a multiplicative subset. Then $S' = \varphi^{-1}(S)$ is a multiplicative subset in B' and $(S')^{-1}B' \to S^{-1}B$ is the universal first order thickening of $S^{-1}B$ over A. In particular $S^{-1}C_{B/A} = C_{S^{-1}B/A}$.*

Note that the lemma makes sense by Corollary 3.1.9.

Proof. With notation and assumptions as in (1). Let $(S^{-1}B)' \to S^{-1}B$ be the universal first order thickening of $S^{-1}B$ over $S^{-1}A$. Note that $S^{-1}B' \to S^{-1}B$ is a surjection of $S^{-1}A$-algebras whose kernel has square zero. Hence by definition we obtain a map $(S^{-1}B)' \to S^{-1}B'$ compatible with the maps towards $S^{-1}B$. Consider any commutative diagram

$$\begin{tikzcd} B \arrow[r] & S^{-1}B \arrow[r] & D/I \\ A \arrow[u] \arrow[r] & S^{-1}A \arrow[u] \arrow[r] & D \arrow[u] \end{tikzcd}$$

where $I \subset D$ is an ideal of square zero. Since B' is the universal first order thickening of B over A we obtain an A-algebra map $B' \to D$. But it is clear that the image of S in D is mapped to invertible elements of D, and hence we obtain a compatible map $S^{-1}B' \to D$. Applying this to $D = (S^{-1}B)'$ we see that we get a map $S^{-1}B' \to (S^{-1}B)'$. We omit the verification that this map is inverse to the map described above.

With notation and assumptions as in (2). Let $(S^{-1}B)' \to S^{-1}B$ be the universal first order thickening of $S^{-1}B$ over A. Note that $(S')^{-1}B' \to S^{-1}B$ is a surjection of A-algebras whose kernel has square zero. Hence by definition we obtain a map $(S^{-1}B)' \to (S')^{-1}B'$ compatible with the maps towards $S^{-1}B$. Consider any commutative diagram

$$\begin{tikzcd} B \arrow[r] & S^{-1}B \arrow[r] & D/I \\ A \arrow[u] \arrow[r] & A \arrow[u] \arrow[r] & D \arrow[u] \end{tikzcd}$$

where $I \subset D$ is an ideal of square zero. Since B' is the universal first order thickening of B over A we obtain an A-algebra map $B' \to D$. But it is clear that the image of S' in D is mapped to invertible elements of D, and hence we obtain a compatible map $(S')^{-1}B' \to D$. Applying this to $D = (S^{-1}B)'$ we see that we get a map $(S')^{-1}B' \to (S^{-1}B)'$. We omit the verification that this map is inverse to the map described above. $\square$

3.1.21 Lemma *Let $R \to A \to B$ be ring maps. Assume $A \to B$ formally unramified. Let $B' \to B$ be the universal first order thickening of B over A. Then B' is formally unramified over A, and the canonical map $\Omega_{A/R} \otimes_A B \to \Omega_{B'/R} \otimes_{B'} B$ is an isomorphism.*

Proof. We are going to use the construction of B' from the proof of Lemma 3.1.17 allthough in principle it should be possible to deduce these results formally from the definition. Namely, we choose a presentation $B = P/J$, where $P = A[x_i]$ is a polynomial ring over A. Next, we choose elements $f_i \in J$ such that $\mathrm{d}f_i = \mathrm{d}x_i \otimes 1$ in $\Omega_{P/A} \otimes_P B$. Having made these choices we have $B' = P/J'$ with $J' = (f_i) + J^2$, see proof of Lemma 3.1.17.

Consider the canonical exact sequence

$$J'/(J')^2 \to \Omega_{P/A} \otimes_P B' \to \Omega_{B'/A} \to 0$$

see **??**. By construction the classes of the $f_i \in J'$ map to elements of the module $\Omega_{P/A} \otimes_P B'$ which generate it modulo J'/J^2 by construction. Since J'/J^2 is a nilpotent ideal, we see that these elements generate the module alltogether (by Nakayama's **??**). This proves that $\Omega_{B'/A} = 0$ and hence that B' is formally unramified over A, see **??**.

Since P is a polynomial ring over A we have $\Omega_{P/R} = \Omega_{A/R} \otimes_A P \oplus \bigoplus P\mathrm{d}x_i$. We are going to use this decomposition. Consider the following exact sequence

$$J'/(J')^2 \to \Omega_{P/R} \otimes_P B' \to \Omega_{B'/R} \to 0$$

see **??**. We may tensor this with B and obtain the exact sequence

$$J'/(J')^2 \otimes_{B'} B \to \Omega_{P/R} \otimes_P B \to \Omega_{B'/R} \otimes_{B'} B \to 0$$

If we remember that $J' = (f_i) + J^2$ then we see that the first arrow annihilates the submodule $J^2/(J')^2$. In terms of the direct sum decomposition $\Omega_{P/R} \otimes_P B = \Omega_{A/R} \otimes_A B \oplus \bigoplus B dx_i$ given we see that the submodule $(f_i)/(J')^2 \otimes_{B'} B$ maps isomorphically onto the summand $\bigoplus B dx_i$. Hence what is left of this exact sequence is an isomorphism $\Omega_{A/R} \otimes_A B \to \Omega_{B'/R} \otimes_{B'} B$ as desired. $\square$

3.2. Smooth morphisms

Definition

The idea of a *smooth* morphism in algebraic geometry is one that is surjective on the tangent space, at least if one is working with smooth varieties over an algebraically closed field. So this means that one should be able to lift tangent vectors, which are given by maps from the ring into $k[\epsilon]/\epsilon^2$.

This makes the following definition seem more plausible:

3.2.1 Definition Let S be an R-algebra. Then S is **formally smooth** over R (or the map $R \to S$ is formally smooth) if given any R-algebra A and ideal $I \subset A$ of square zero, the map

$$\hom_R(S, A) \to \hom_R(S, A/I)$$

is a surjection. We shall say that S is **smooth** (over R) if it is formally smooth and of finite presentation.

So this means that in any diagram

$$\begin{tikzcd} S \arrow[r] \arrow[dr, dashed] & A/I \\ R \arrow[u] \arrow[r] & A, \arrow[u] \end{tikzcd}$$

with I an ideal of square zero in A, there exists a dotted arrow making the diagram commute. As with formal unramifiedness, this is a purely functorial statement: if F is the corepresentable functor associated to S, then we want $F(A) \to F(A/I)$ to be a *surjection* for each $I \subset A$ of square zero and each R-algebra A. Also, again we can replace "I of square zero" with "I nilpotent."

3.2.2 Example The basic example of a formally smooth R-algebra is the polynomial ring $R[x_1, \dots, x_n]$. For to give a map $R[x_1, \dots, x_n] \to A/I$ is to give n elements of A/I; each of these elements can clearly be lifted to A. This is analogous to the statement that a free module is projective.

More generally, if P is a projective R-module (not necessarily of finite type), then the symmetric algebra $\mathsf{S}P$ is a formally smooth R-algebra. This follows by the same reasoning.

We can state the usual list of properties of formally smooth morphisms:

3.2.3 Proposition *Smooth (resp. formally smooth) morphisms are preserved under base extension and composition. If R is a ring, then any localization is formally smooth over R.*

Proof. As usual, only the statements about *formal* smoothness are interesting. The statements about base extension and composition will be mostly left to the reader: they are an exercise in diagram-chasing. (Note that we cannot argue as we did for formally unramified morphisms, where we had a simple criterion in terms of the module of Kähler differentials and various properties of them.) For example, let $R \to S, S \to T$ be formally smooth. Given a diagram (with $I \subset A$ an ideal of square zero)

$$\begin{tikzcd} T \arrow[r] \arrow[ddr, dashed] & A/I \\ S \arrow[u] \arrow[dr, dashed] & \\ R \arrow[u] \arrow[r] & A, \arrow[uu] \end{tikzcd}$$

we start by finding a dotted arrow $S \to A$ by using formal smoothness of $R \to S$. Then we find a dotted arrow $T \to A$ making the top quadrilateral commute. This proves that the composite is formally smooth. ☐

Quotients of formally smooth rings

Now, ultimately, we want to show that this somewhat abstract definition of smoothness will give us something nice and geometric. In particular, in this case we want to show that B is *flat,* and the fibers are smooth varieties (in the old sense). To do this, we will need to do a bit of work, but we can argue in a fairly elementary manner. On the one hand, we will first need to give a criterion for when a quotient of a formally smooth ring is formally smooth.

3.2.4 Theorem *Let A be a ring, B an A-algebra. Suppose B is formally smooth over A, and let $I \subset B$ be an ideal. Then $C = B/I$ is a formally smooth A-algebra if and only if the canonical map*

$$I/I^2 \to \Omega_{B/A} \otimes_B C$$

has a section. In other words, C is formally smooth precisely when the conormal sequence

$$I/I^2 \to \Omega_{B/A} \otimes_B C \to \Omega_{C/A} \to 0$$

is split exact.

This result is stated in more generality for *topological* rings, and uses some functors on ring extensions, in **?**, 0-IV, 22.6.1.

Proof. Suppose first C is formally smooth over A. Then we have a map $B/I^2 \to C$ given by the quotient. The claim is that there is a section of this map. There is a diagram of A-algebras

$$\begin{tikzcd} B/I & B/I^2 \arrow[l] \\ C \arrow[u, "="] \arrow[ur, dashed] & \end{tikzcd}$$

and the lifting $s : C \to B/I^2$ exists by formal smoothness. This is a section of the natural projection $B/I^2 \to C = B/I$.

In particular, the combination of the natural inclusion $I/I^2 \to B/I^2$ and the section s gives an isomorphism of *rings* (even A-algebras) $B/I^2 \simeq C \oplus I/I^2$. Here I/I^2 squares to zero.

We are interested in showing that $I/I^2 \to \Omega_{B/A} \otimes_B C$ is a split injection of C-modules. To see this, we will show that any map out of the former extends to a map out of the latter. Now suppose given a map of C-modules

$$\phi : I/I^2 \to M$$

into a C-module M. Then we get an A-derivation

$$\delta : B/I^2 \to M$$

by using the splitting $B/I^2 = C \oplus I/I^2$. (Namely, we just extend the map by zero on C.) Since I/I^2 is imbedded in B/I^2 by the canonical injection, this derivation restricts on I/I^2 to ϕ. In other words there is a commutative diagram

$$\begin{tikzcd}
I/I^2 \arrow[r] \arrow[d, "\phi"] & B/I^2 \arrow[dl, "\delta"] \\
M &
\end{tikzcd}.$$

It follows thus that we may define, by pulling back, an A-derivation $B \to M$ that restricts on I to the map $I \to I/I^2 \stackrel{\phi}{\to} M$. By the universal property of the differentials, this is the same thing as a homomorphism $\Omega_{B/A} \to M$, or equivalently $\Omega_{B/A} \otimes_B C \to M$ since M is a C-module. Pulling back this derivation to I/I^2 corresponds to pulling back via $I/I^2 \to \Omega_{B/A} \otimes_B C$.

It follows that the map

$$\hom_C(\Omega_{B/A} \otimes_B C, M) \to \hom_C(I/I^2, M)$$

is a surjection. This proves one half of the result.

Now for the other. Suppose that there is a section of the conormal map. This translates, as above, to saying that any map $I/I^2 \to M$ (of C-modules) for a C-module M can be extended to an A-derivation $B \to M$. We must deduce from this formal smoothness.

Let E be any A-algebra, and $J \subset E$ an ideal of square zero. We suppose given an A-homomorphism $C \to E/J$ and would like to lift it to $C \to E$; in other words, we must find a lift in the diagram

$$\begin{tikzcd}
 & C \arrow[d] \arrow[dl, dashed] \\
E \arrow[r] & E/J
\end{tikzcd}.$$

Let us pull this map back by the surjection $B \twoheadrightarrow C$; we get a diagram

$$\begin{tikzcd} & B \arrow[d] \arrow[ddl, dashed, "\phi"'] \\ & C \arrow[d] \arrow[dl, dashed] \\ E \arrow[r] & E/J \end{tikzcd}.$$

In this diagram, we know that a lifting $\phi : B \to E$ does exist because B is formally smooth over A. So we can find a dotted arrow from $B \to E$ in the diagram. The problem is that it might not send $I = \ker(B \to C)$ into zero. If we can show that there *exists* a lifting that does factor through C (i.e. sends I to zero), then we are done.

In any event, we have a morphism of A-modules $I \to E$ given by restricting $\phi : B \to E$. This lands in J, so we get a map $I \to J$. Note that J is an E/J-module, hence a C-module, because J has square zero. Moreover I^2 gets sent to zero because $J^2 = 0$, and we have a morphism of C-modules $I/I^2 \to J$. Now by hypothesis, there is an A-derivation $\delta : B \to J$ such that $\delta|_I = \phi$. Since J has square zero, it follows that

$$\phi - \delta : B \to E \qquad \square$$

is an A-homomorphism of algebras, and it kills I. Consequently this factors through C and gives the desired lifting $C \to E$.

3.2.5 Corollary *If $A \to B$ is formally smooth, then $\Omega_{B/A}$ is a projective B-module.*

The intuition is that projective modules correspond to vector bundles over the Spec (unlike general modules, the rank is locally constant, which should happen in a vector bundle). But a smooth algebra is like a manifold, and for a manifold the cotangent bundle is very much a vector bundle, whose dimension is locally constant.

Proof. Indeed, we can write B as a quotient of a polynomial ring D over A; this is formally smooth. Suppose $B = D/I$. Then we know that there is a split exact sequence

$$0 \to I/I^2 \to \Omega_{D/A} \otimes_D B \to \Omega_{B/A} \to 0.$$

But the middle term is free as D/A is a polynomial ring; hence the last term is projective. $\square$

In particular, we can rewrite the criterion for formal smoothness of $C = B/I$, if B is formally smooth over A:

1. $\Omega_{C/A}$ is a projective C-module.
2. $I/I^2 \to \Omega_{B/A} \otimes_B C$ is a monomorphism.

Indeed, these two are equivalent to the splitting of the conormal sequence (since the middle term is always projective by corollary 3.2.5).

In particular, we can check that smoothness is *local*:

3.2.6 Corollary *Let A be a ring, B a finitely presented A-algebra. Then B is smooth over A if and only if for each $\mathfrak{q} \in \operatorname{Spec} B$ with $\mathfrak{p} \in \operatorname{Spec} A$ the inverse image, the map $A_{\mathfrak{p}} \to B_{\mathfrak{q}}$ is formally smooth.*

Proof. Indeed, we see that $B = D/I$ for a polynomial ring $D = A[x_1, \dots, x_n]$ in finitely many variables, and $I \subset D$ a finitely generated ideal. We have just seen that we just need to check that the conormal map $I/I^2 \to \Omega_{D/A} \otimes_D B$ is injective, and that $\Omega_{B/A}$ is a projective B-module, if and only if the analogs hold over the localizations. This follows by the criterion for formal smoothness just given above.

But both can be checked locally. Namely, the conormal map is an injection if and only if, for all $\mathfrak{q} \in \operatorname{Spec} B$ corresponding to $\mathfrak{Q} \in \operatorname{Spec} D$, the map $(I/I^2)_{\mathfrak{q}} \to \Omega_{D_{\mathfrak{Q}}/A_{\mathfrak{p}}} \otimes_{D_{\mathfrak{Q}}} B_{\mathfrak{q}}$ is an injection. Moreover, we know that for a finitely presented module over a ring, like $\Omega_{B/A}$, projectivity is equivalent to projectivity (or freeness) of all the stalks (**??**). So we can check projectivity on the localizations too. ☐

In fact, the method of proof of corollary 3.2.6 yields the following observation: *formal* smoothness "descends" under faithfully flat base change. That is:

3.2.7 Corollary *If B is an A-algebra, and A' a faithfully flat algebra, then B is formally smooth over A if and only if $B \otimes_A A'$ is formally smooth over A'.*

We shall not give a complete proof, except in the case when B is finitely presented over A (so that the question is of smoothness).

Proof. One direction is just the "sorite" (see **??**). We want to show that formal smoothness "descends." The claim is that the two conditions for formal smoothness above (that $\Omega_{B/A}$ be projective and the conormal map be a monomorphism) descend under faithfully flat base-change. Namely, the fact about the conormal maps is clear (by faithful flatness).

Now let $B' = B \otimes_A A'$. So we need to argue that if $\Omega_{B'/A'} = \Omega_{B/A} \otimes_B B'$ is projective as a B'-module, then so is $\Omega_{B/A}$. Here we use the famous result of Raynaud-Gruson (see **?**), which states that projectivity descends under faithfully flat extensions, to complete the proof.

If B is finitely presented over A, then $\Omega_{B/A}$ is finitely presented as a B-module. We can run most of the same proof as before, but we want to avoid using the Raynaud-Gruson theorem: we must give a separate argument that $\Omega_{B/A}$ is projective if $\Omega_{B'/A'}$ is. However, for a finitely presented module, projectivity is *equivalent* to flatness, by theorem 4.4.16. Moreover, since $\Omega_{B'/A'}$ is B'-flat, faithful flatness enables us to conclude that $\Omega_{B/A}$ is B-flat, and hence projective. ☐

The Jacobian criterion

Now we want a characterization of when a morphism is smooth. Let us motivate this with an analogy from standard differential topology. Consider real-valued functions $f_1, \dots, f_p \in C^\infty(\mathbb{R}^n)$. Now, if $f_1, f_2, \dots, f_p$ are such that their gradients ∇f_i form a matrix of rank p, then we can define a manifold near zero which is the common zero set of all the f_i. We are going to give a relative version of this in the algebraic setting.

Recall that a map of rings $A \to B$ is *essentially of finite presentation* if B is the localization of a finitely presented A-algebra.

3.2.8 Proposition *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be a local homomorphism of local rings such that B is essentially of finite presentation. Suppose $B = (A[X_1, \dots, X_n])_{\mathfrak{q}}/I$ for some finitely generated ideal $I \subset A[X_1, \dots, X_n]_{\mathfrak{q}}$, where $\mathfrak{q}$ is a prime ideal in the polynomial ring.*

Then I/I^2 is generated as a B-module by polynomials $f_1, \dots, f_k \in I \subset A[X_1, \dots, X_n]$ whose Jacobian matrix has maximal rank in $C/\mathfrak{q} = B/\mathfrak{n}$ if and only if B is formally smooth over A. In this case, I/I^2 is even freely generated by the f_i.

The Jacobian matrix $\frac{\partial f_i}{\partial X_j}$ is a matrix of elements of $A[X_1, \dots, X_n]$, and we can take the associated images in $B/\mathfrak{n}$.

3.2.9 Example Suppose A is an algebraically closed field k. Then I corresponds to some ideal in the polynomial ring $k[X_1, \dots, X_n]$, which cuts out a variety X. Suppose $\mathfrak{q}$ is a maximal ideal in the polynomial ring.

Then B is the local ring of the algebraic variety X at $\mathfrak{q}$. Then proposition 3.2.8 states that $\mathfrak{q}$ is a "smooth point" of the variety (i.e., the Jacobian matrix has maximal rank) if and only if B is formally smooth over k. We will expand on this later.

Proof. Indeed, we know that polynomial rings are formally smooth. In particular $D = A[X_1, \dots, X_n]_{\mathfrak{q}}$ is formally smooth over A, because localization preserves formal smoothness. Note also that $\Omega_{D/A}$ is a free D-module, because this is true for a polynomial ring and Kähler differentials commute with localization.

So theorem 3.2.4 implies that

$$I/I^2 \to \Omega_{D/A} \otimes_D B$$

is a split injection precisely when B is formally smooth over A. Suppose that this holds. Now I/I^2 is then a summand of the free module $\Omega_{D/A} \otimes_D B$, so it is projective, hence free as B is local. Let $K = B/\mathfrak{n}$. It follows that the map

$$I/I^2 \otimes_D K \to \Omega_{D/A} \otimes_D K = K^n$$

is an injection. This map sends a polynomial to its gradient (reduced modulo $\mathfrak{q}$, or $\mathfrak{n}$). Hence the assertion is clear: choose polynomials $f_1, \dots, f_k \in I$ that generate $(I/I^2)_{\mathfrak{q}}$, and their gradients in $B/\mathfrak{n}$ must be linearly independent.

Conversely, suppose that I/I^2 has such generators. Then the map

$$I/I^2 \otimes K \to K^n, \quad f \mapsto df$$

is a split injection. However, if a map of finitely generated modules over a local ring, with the target free, is such that tensoring with the residue field makes it an injection, then it is a split injection. (We shall prove this below.) Thus $I/I^2 \to \Omega_{D/A} \otimes_D B$ is a split injection. In view of the criterion for formal smoothness, we find that B is formally smooth. $\square$

Here is the promised lemma necessary to complete the proof:

3.2.10 Lemma *If $(A, \mathfrak{m})$ is a local ring with residue field k, M a finitely generated A-module, N a finitely generated projective A-module, then a map $\phi : M \to N$ is a split injection if and only if $M \otimes_A k \to N \otimes_A k$ is an injection.*

Proof. One direction is clear, so it suffices to show that $M \to N$ is a split injection if the map on fibers is an injection.

Let L be a "free approximation" to M, that is, a free module L together with a map $L \to M$ which is an isomorphism modulo k. By Nakayama's lemma, $L \to M$ is surjective. Then the map $L \to M \to N$ is such that the $L \otimes k \to N \otimes k$ is injective, so $L \to N$ is a split injection (by an elementary criterion). It follows that we can find a splitting $N \to L$, which when composed with $L \to M$ is a splitting of $M \to N$. ☐

The fiberwise criterion for smoothness

We shall now prove that a smooth morphism is flat. In fact, we will get a general "fiberwise" criterion for smoothness (i.e., a morphism is smooth if and only if it is flat and the fibers are smooth), which will enable us to reduce smoothness questions, in some cases, to the situation where the base is a field.

We shall need some lemmas on regular sequences. The first will give a useful criterion for checking M-regularity of an element by checking on the fiber. For our purposes, it will also give a criterion for when quotienting by a regular element preserves flatness over a smaller ring.

3.2.11 Lemma *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be a local homomorphism of local noetherian rings. Let M be a finitely generated B-module, which is flat over A.*

Let $f \in B$. Then the following are equivalent:

1. *M/fM is flat over A and $f : M \to M$ is injective.*
2. *$f : M \otimes_A k \to M \otimes_A k$ is injective where $k = A/\mathfrak{m}$.*

For instance, let us consider the case $M = B$. The lemma states that if multiplication by f is regular on $B \otimes_A k$, then the hypersurface cut out by f (i.e., corresponding to the ring B/fB) is flat over A.

Proof. All Tor functors here will be over A. If M/fM is A-flat and $f : M \to M$ is injective, then the sequence

$$0 \to M \xrightarrow{f} M \to M/fM \to 0$$

leads to a long exact sequence

$$\mathrm{Tor}_1(k, M/fM) \to M \otimes_A k \xrightarrow{f} M \otimes_A k \to (M/fM) \otimes_A k \to 0.$$

But since M/fM is flat, the first term is zero, and it follows that $M \otimes k \xrightarrow{f} M \otimes k$ is injective.

The other direction is more subtle. Suppose multiplication by f is a monomorphism on $M \otimes_A k$. Now write the exact sequence

$$0 \to P \to M \xrightarrow{f} M \to Q \to 0$$

where P, Q are the kernel and cokernel. We want to show that $P = 0$ and Q is flat over A.

We can also consider the image $I = fM \subset M$, to split this into two exact sequences

$$0 \to P \to M \to I \to 0$$

and

$$0 \to I \to M \to Q \to 0.$$

Here the map $M \otimes_A k \to I \otimes_A k \to M \otimes_A k$ is given by multiplication by f, so it is injective by hypothesis. This implies that $M \otimes_A k \to I \otimes_A k$ is injective. So $M \otimes k \to I \otimes k$ is actually an isomorphism because it is obviously surjective, and we have just seen it is injective. Moreover, $I \otimes_A k \to M \otimes_A k$ is isomorphic to the homothety $f : M \otimes_A k \to M \otimes_A k$, and consequently is injective. To summarize:

1. $M \otimes_A k \to I \otimes_A k$ is an isomorphism.
2. $I \otimes_A k \to M \otimes_A k$ is an injection.

Let us tensor these two exact sequences with k. We get

$$0 \to \mathrm{Tor}_1(k, I) \to P \otimes_A k \to M \otimes_A k \to I \otimes_A k \to 0$$

because M is flat. We also get

$$0 \to \mathrm{Tor}_1(k, Q) \to I \otimes_A k \to M \otimes_A k \to Q \otimes_A k \to 0.$$

We'll start by using the second sequence. Now $I \otimes_A k \to M \otimes_A k$ was just said to be injective, so that $\mathrm{Tor}_1(k, Q) = 0$. By the local criterion for flatness, it follows that Q is a flat A-module as well. But $Q = M/fM$, so this gives one part of what we wanted.

Now, we want to show finally that $P = 0$. Now, I is flat; indeed, it is the kernel of a surjection of flat maps $M \to Q$, so the long exact sequence shows that it is flat. So we have a short exact sequence

$$0 \to P \otimes_A k \to M \otimes_A k \to I \otimes_A k \to 0,$$

which shows now that $P \otimes_A k = 0$ (as $M \otimes_A k \to I \otimes_A k$ was just shown to be an isomorphism earlier). By Nakayama $P = 0$. This implies that f is M-regular. ∎

3.2.12 Corollary *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be a morphism of noetherian local rings. Suppose M is a finitely generated B-module, which is flat over A.*

Let $f_1, \dots, f_k \in \mathfrak{n}$. Suppose that $f_1, \dots, f_k$ is a regular sequence on $M \otimes_A k$. Then it is a regular sequence on M and, in fact, $M/(f_1, \dots, f_k)M$ is flat over A.

Proof. This is now clear by induction. ∎

3.2.13 Theorem *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be a morphism of local rings such that B is the localization of a finitely presented A-algebra at a prime ideal, $B = (A[X_1, \dots, X_n])_{\mathfrak{q}}/I$. Then if $A \to B$ is formally smooth, B is a flat A-algebra.*

The strategy is that B is going to be written as the quotient of a localization of a polynomial ring by a sequence $\{f_i\}$ whose gradients are independent (modulo the maximal ideal), i.e. modulo $B/\mathfrak{n}$. If we were working modulo a field, then we could use arguments about regular local rings to argue that the $\{f_i\}$ formed a regular sequence. We will use corollary 3.2.12 to bootstrap from this case to the general situation.

Proof. Let us first assume that A is *noetherian.*

Let $C = (A[X_1, \dots, X_n])_{\mathfrak{q}}$. Then C is a local ring, smooth over A, and we have morphisms of local rings

$$(A, \mathfrak{m}) \to (C, \mathfrak{q}) \twoheadrightarrow (B, \mathfrak{n}).$$

Moroever, C is a *flat* A-module, and we are going to apply the fiberwise criterion for regularity to C and a suitable sequence.

Now we know that I/I^2 is a B-module generated by polynomials $f_1, \dots, f_m \in A[X_1, \dots, X_n]$ whose Jacobian matrix has maximal rank in $B/\mathfrak{n}$ (by the Jacobian criterion, proposition 3.2.8). The claim is that the f_i are linearly independent in $\mathfrak{q}/\mathfrak{q}^2$. This will be the first key step in the proof. In other words, if $\{u_i\}$ is a family of elements of C, not all non-units, we do not have

$$\sum u_i f_i \in \mathfrak{q}^2.$$

For if we did, then we could take derivatives and find

$$\sum u_i \partial_j f_i \in \mathfrak{q}$$

for each j. This contradicts the gradients of the f_i being linearly independent in $B/\mathfrak{n} = C/\mathfrak{q}$.

Now we want to show that the $\{f_i\}$ form a regular sequence in C. To do this, we shall reduce to the case where A is a field. Indeed, let us make the base-change $A \to k = A/\mathfrak{m}, B \to \overline{B} = B \otimes_A k, C \to \overline{C} = C \otimes_A k$ where $k = A/\mathfrak{m}$ is the residue field. Then $\overline{B}, \overline{C}$ are formally smooth local rings over a field k. We also know that $\overline{C}$ is a *regular* local ring, since it is a localization of a polynomial ring over a field.

Let us denote the maximal ideal of $\overline{C}$ by $\overline{\mathfrak{q}}$; this is just the image of $\mathfrak{q}$.

Now the $\{f_i\}$ have images in $\overline{C}$ that are linearly independent in $\overline{\mathfrak{q}}/\overline{\mathfrak{q}}^2 = \mathfrak{q}/\mathfrak{q}^2$. It follows that the $\{f_i\}$ form a regular sequence in $\overline{C}$, by general facts about regular local rings (see, e.g. corollary 9.1.12); indeed, each of the successive quotients $\overline{C}/(f_1, \dots, f_i)$ will then be regular. It follows from the fiberwise criterion (C being flat) that the $\{f_i\}$ form a regular sequence in C itself, and that the quotient $C/(f_i) = B$ is A-flat. ∎

The proof in fact showed a bit more: we expressed B as the quotient of a localized polynomial ring by a regular sequence. In other words:

3.2.14 Corollary (Smooth maps are local complete intersections) *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be an essentially of finite presentation, formally smooth map. Then there exists a localization of a polynomial ring, C, such that B can be expressed as $C/(f_1, \dots, f_n)$ for the $\{f_i\}$ forming a regular sequence in the maximal ideal of C.*

We also get the promised result:

3.2.15 Theorem *Let $A \to B$ be a smooth morphism of rings. Then B is flat over A.*

Proof. Indeed, we immediately reduce to theorem 3.2.13 by checking locally at each prime (which gives formally smooth maps). □

In fact, we can get a general criterion now:

3.2.16 Theorem *Let $(A, \mathfrak{m}) \to (B, \mathfrak{n})$ be a (local) morphism of local noetherian rings such that B is the localization of a finitely presented A-algebra at a prime ideal, $B = (A[X_1, \dots, X_n])_{\mathfrak{q}}/I$. Then B is formally smooth over A if B is A-flat and $B/\mathfrak{m}B$ is formally smooth over $A/\mathfrak{m}$.*

Proof. One direction is immediate from what we have already shown. Now we need to show that if B is A-flat, and $B/\mathfrak{m}B$ is formally smooth over $A/\mathfrak{m}$, then B is itself formally smooth over A. This will be comparatively easy, with all the machinery developed. This will be comparatively easy, with all the machinery developed.

As before, write the sequence

$$(A, \mathfrak{m}) \to (C, \mathfrak{q}) \twoheadrightarrow (B, \mathfrak{n}),$$

where C is a localization of a polynomial ring at a prime ideal, and in particular is formally smooth over A. We know that $B = C/I$, where $I \subset \mathfrak{q}$.

To check that B is formally smooth over A, we need to show (C being formally smooth) that the conormal sequence

$$I/I^2 \to \Omega_{C/A} \otimes_C B \to \Omega_{C/B} \to 0. \tag{3.2.1}$$

is split exact.

Let $\overline{A}, \overline{C}, \overline{B}$ be the base changes of A, B, C to $k = A/\mathfrak{m}$; let $\overline{I}$ be the kernel of $\overline{C} \twoheadrightarrow \overline{B}$. Note that $\overline{I} = I/\mathfrak{m}I$ by flatness of B. Then we know that the sequence

$$\overline{I}/\overline{I}^2 \to \Omega_{\overline{C}/k}/\overline{I}\Omega_{\overline{C}/k} \to \Omega_{\overline{C}/\overline{B}} \to 0 \tag{3.2.2}$$

is split exact, because $\overline{C}$ is a formally smooth k-algebra (in view of theorem 3.2.4).

But (3.2.2) is the reduction of (3.2.1). Since the middle term of (3.2.1) is finitely generated and projective over B, we can check splitting modulo the maximal ideal (see lemma 3.2.10). □

In particular, we get the global version of the fiberwise criterion:

3.2.17 Theorem *Let $A \to B$ be a finitely presented morphism of rings. Then B is a smooth A-algebra if and only if B is a flat A-algebra and, for each $\mathfrak{p} \in \operatorname{Spec} A$, the morphism $k(\mathfrak{p}) \to B \otimes_A k(\mathfrak{p})$ is smooth.*

Here $k(\mathfrak{p})$ denotes the residue field of $A_{\mathfrak{p}}$, as usual.

Proof. One direction is clear. For the other, we recall that smoothness is *local*: $A \to B$ is smooth if and only if, for each $\mathfrak{q} \in \operatorname{Spec} B$ with image $\mathfrak{p} \in \operatorname{Spec} A$, we have $A_{\mathfrak{p}} \to B_{\mathfrak{q}}$ formally smooth (see corollary 3.2.6). But, by theorem 3.2.16, this is the case if and only if, for each such pair $(\mathfrak{p}, \mathfrak{q})$, the morphism $k(\mathfrak{p}) \to B_{\mathfrak{q}} \otimes_{A_{\mathfrak{p}}} k(\mathfrak{p})$ is formally smooth. Now if $k(\mathfrak{p}) \to B \otimes_A k(\mathfrak{p})$ is smooth for each $\mathfrak{p}$, then this condition is clearly satisfied. □

Formal smoothness and regularity

We now want to explore the connection between formal smoothness and regularity. In general, the intuition is that a variety over an algebraically closed field is *smooth* if and only if the local rings at closed points (and thus at all points by **??**) are regular local rings. Over a non-algebraically closed field, only one direction is still true: we want the local rings to be *geometrically regular.* So far we will just prove one direction, though.

3.2.18 Theorem *Let $(A, \mathfrak{m})$ be a noetherian local ring containing a copy of its residue field $A/\mathfrak{m} = k$. Then if A is formally smooth over k, A is regular.*

Proof. We are going to compare the quotients $A/\mathfrak{m}^m$ to the quotients of $R = k[x_1, \dots, x_n]$ where n is the *embedding dimension* of A. Let $\mathfrak{n} \subset k[x_1, \dots, x_n]$ be the ideal $(x_1, \dots, x_n)$. We are going to give surjections

$$A/\mathfrak{m}^m \twoheadrightarrow R/\mathfrak{n}^m$$

for each $m \geqslant 2$.

Let $t_1, \dots, t_n \in \mathfrak{m}$ be a k-basis for $\mathfrak{m}/\mathfrak{m}^2$. Consider the map $A \twoheadrightarrow R/\mathfrak{n}^2$ that goes $A \twoheadrightarrow A/\mathfrak{m}^2 \simeq k \oplus \mathfrak{m}/\mathfrak{m}^2 \simeq R/\mathfrak{n}^2$, where t_i is sent to x_i. This is well-defined, and gives a surjection $A \twoheadrightarrow R/\mathfrak{n}^2$. Using the infinitesimal lifting property, we can lift this map to k-algebra maps

$$A \to R/\mathfrak{n}^m$$

for each k, which necessarily factor through $A/\mathfrak{m}^m$ (as they send $\mathfrak{m}$ into $\mathfrak{n}$). They are surjective by Nakayama's lemma. It follows that

$$\dim_k A/\mathfrak{m}^m \geqslant \dim_k R/\mathfrak{n}^m,$$

and since $R_{\mathfrak{n}}$ is a regular local ring, the last term grows asymptotically like m^n. It follows that $\dim R \geqslant n$, and since $\dim R$ is always at most the embedding dimension, we are done. □

A counterexample

It is in fact true that a formally smooth morphism between *arbitrary* noetherian rings is flat, although we have only proved this in the case of a morphism of finite type. This is false if we do not assume noetherian hypotheses. A formally smooth morphism need not be flat.

3.2.19 Example Consider a field k, and consider $R = k[T^x]_{x \in \mathbb{Q}_{>0}}$. This is the filtered colimit of the polynomial rings $k[T^{1/n}]$ over all n. There is a natural map $R \to k$ sending each power of T to zero. The claim is that $R \to k$ is a formally smooth morphism which is not flat. It is a *surjection*, so it is a lot different from the intuitive idea of a smooth map.

Yet it turns out to be *formally* smooth. To see this, consider an R-algebra S and an ideal $I \subset S$ such that $S^2 = 0$. The claim is that an R-homomorphism $k \to S/I$ lifts to $k \to S$. Consider the diagram

$$\begin{tikzcd} & & S \arrow[d] \\ R \arrow[r] \arrow[urr] & k \arrow[r] \arrow[ur, dashed] & S/I, \end{tikzcd}$$

in which we have to show that a dotted arrow exists.

However, there can be at most one R-homomorphism $k \to S/I$, since k is a quotient of R. It follows that each $T^x, x \in \mathbb{Q}_{>0}$ is mapped to zero in S/I. So each $T^x, x \in I$ maps to elements of I (by the map $R \to S$ assumed to exist). It follows that $T^x = (T^{x/2})^2$ maps to zero in S, as $I^2 = 0$. Thus the map $R \to S$ annihilates each T^x, which means that there is a (unique) dotted arrow.

Note that $R \to k$ is not flat. Indeed, multiplication by T is injective on R, but it acts by zero on k.

This example was described by Anton Geraschenko on MathOverflow; see **?**. The same reasoning shows more generally:

3.2.20 Proposition *Let R be a ring, $I \subset R$ an ideal such that $I = I^2$. Then the projection $R \to R/I$ is formally étale.*

For a noetherian ring, if $I = I^2$, then we know that I is generated by an idempotent in R (see proposition 4.1.26), and the projection $R \to R/I$ is projection on the corresponding direct factor (actually, the complementary one). In this case, the projection is flat, and this is to be expected: as stated earlier, formally étale implies flat for noetherian rings. But in the non-noetherian case, we can get interesting examples.

3.2.21 Example We shall now give an example showing that formally étale morphisms do not necessarily preserve reducedness. We shall later see that this is true in the *étale* case (see proposition 3.3.19).

Let k be a field of characteristic $\neq 2$. Consider the ring $R = k[T^x]_{x \in \mathbb{Q}_{>0}}$ as before. Take $S = R[X]/(X^2 - T)$, and consider the ideal I generated by all the positive powers $T^x, x > 0$. As before, clearly $I = I^2$, and thus $S \to S/I$ is formally étale. The claim is that S is reduced; clearly $S/I = k[X]/(X^2)$ is not. Indeed, an element of S can be uniquely described by $\alpha = P(T) + Q(T)X$ where P, Q are "polynomials" in T—in actuality, they are allowed to have terms $T^x, x \in \mathbb{Q}_{>0}$. Then $\alpha^2 = P(T)^2 + Q(T)^2T + 2P(T)Q(T)X$. It is thus easy to see that if $\alpha^2 = 0$, then $\alpha = 0$.

3.3. Étale morphisms

Definition

The definition is just another nilpotent lifting property:

3.3.1 Definition Let S be an R-algebra. Then S is **formally étale** over R (or the map $R \to S$ is formally étale) if given any R-algebra A and ideal $I \subset A$ of square zero, the map

$$\hom_R(S, A) \to \hom_R(S, A/I)$$

is a bijection. A ring homomorphism is **étale** if and only if it is formally étale and of finite presentation.

So S is *formally étale over R* if for every commutative solid diagram

$$\begin{tikzcd} S \arrow[r] \arrow[dr, dashed] & A/I \\ R \arrow[u] \arrow[r] & A \arrow[u] \end{tikzcd}$$

where $I \subset A$ is an ideal of square zero, there exists a unique dotted arrow making the diagram commute. As before, the functor of points can be used to test formal étaleness. Moreover, clearly a ring map is formally étale if and only if it is both formally smooth and formally unramified.

We have the usual:

3.3.2 Proposition *Étale (resp. formally étale) morphisms are closed under composition and base change.*

Proof. Either a combination of the corresponding results for formal smoothness and formal unramifiedness (i.e. proposition 3.1.6, proposition 3.1.7, and proposition 3.2.3), or easy to verify directly. ☐

Filtered colimits preserve formal étaleness:

3.3.3 Lemma *Let R be a ring. Let I be a directed partially ordered set. Let $(S_i, \varphi_{ii'})$ be a system of R-algebras over I. If each $R \to S_i$ is formally étale, then $S = colim_{i \in I}\ S_i$ is formally étale over R*

The idea is that we can make the lifts on each piece, and glue them automatically.

Proof. Consider a diagram as in Definition 3.3.1. By assumption we get unique R-algebra maps $S_i \to A$ lifting the compositions $S_i \to S \to A/I$. Hence these are compatible with the transition maps $\varphi_{ii'}$ and define a lift $S \to A$. This proves existence. The uniqueness is clear by restricting to each S_i. ☐

3.3.4 Lemma *Let R be a ring. Let $S \subset R$ be any multiplicative subset. Then the ring map $R \to S^{-1}R$ is formally étale.*

Proof. Let $I \subset A$ be an ideal of square zero. What we are saying here is that given a ring map $\varphi : R \to A$ such that $\varphi(f) \mod I$ is invertible for all $f \in S$ we have also that $\varphi(f)$ is invertible in A for all $f \in S$. This is true because A^* is the inverse image of $(A/I)^*$ under the canonical map $A \to A/I$. □

We now want to give the standard example of an étale morphism; geometrically, this corresponds to a hypersurface in affine 1-space given by a nonsingular equation. We will eventually show that any étale morphism looks like this, locally.

3.3.5 Example Let R be a ring, $P \in R[X]$ a polynomial. Suppose $Q \in R[X]/P$ is such that in the localization $(R[X]/P)_Q$, the image of the derivative $P' \in R[X]$ is a unit. Then the map

$$R \to (R[X]/P)_Q$$

is called a **standard étale morphism.**

The name is justified by:

3.3.6 Proposition *A standard étale morphism is étale.*

Proof. It is sufficient to check the condition on the Kähler differentials, since a standard étale morphism is evidently flat and of finite presentation. Indeed, we have that

$$\Omega_{(R[X]/P)_Q/R} = Q^{-1}\Omega_{(R[X]/P)/R} = Q^{-1}\frac{R[X]}{(P'(X), P(X))R[X]}$$

by basic properties of Kähler differentials. Since P' is a unit after localization at Q, this last object is clearly zero. □

3.3.7 Example A separable algebraic extension of a field k is formally étale. Indeed, we just need to check this for a finite separable extension L/k, in view of lemma 3.3.3, and then we can write $L = k[X]/(P(X))$ for P a separable polynomial. But it is easy to see that this is a special case of a standard étale morphism. In particular, any unramified extension of a field is étale, in view of the structure theory for unramified extensions of fields (theorem 3.1.12).

3.3.8 Example The example of example 3.2.19 is a formally étale morphism, because we showed the map was formally smooth and it was clearly surjective. It follows that a formally étale morphism is not necessarily flat!

We also want a slightly different characterization of an étale morphism. This criterion will be of extreme importance for us in the sequel.

3.3.9 Theorem *An R-algebra S of finite presentation is étale if and only if it is flat and unramified.*

This is in fact how étale morphisms are defined in **?** and in **?**.

Proof. An étale morphism is smooth, hence flat (theorem 3.2.15). Conversely, suppose S is flat and unramified over R. We just need to show that S is smooth over R. But this follows by the fiberwise criterion for smoothness, theorem 3.2.16, and the fact that an unramified extension of a field is automatically étale, by example 3.3.7. ☐

Finally, we would like a criterion for when a morphism of *smooth* algebras is étale. We state it in the local case first.

3.3.10 Proposition *Let B, C be local, formally smooth, essentially of finite presentation A-algebras and let $f : B \to C$ be a local A-morphism. Then f is formally étale if and only if and only if the map $\Omega_{B/A} \otimes_B C \to \Omega_{C/A}$ is an isomorphism.*

The intuition is that f induces an isomorphism on the cotangent spaces; this is analogous to the definition of an *étale* morphism of smooth manifolds (i.e. one that induces an isomorphism on each tangent space, so is a local isomorphism at each point).

Proof. We prove this for A noetherian.

We just need to check that f is flat if the map on differentials is an isomorphism. Since B, C are flat A-algebras, it suffices (by the general criterion, proposition 1.4.12), to show that $B \otimes_A k \to C \otimes_A k$ is flat for k the residue field of A. We will also be done if we show that $B \otimes_A \overline{k} \to C \otimes_A \overline{k}$ is flat. Note that the same hypotheses (that

So we have reduced to a question about rings essentially of finite type over a *field*. Namely, we have local rings $\overline{B}, \overline{C}$ which are both formally smooth, essentially of finite-type k-algebras, and a map $\overline{B} \to \overline{C}$ that induces an isomorphism on the Kähler differentials as above.

The claim is that $\overline{B} \to \overline{C}$ is flat (even local-étale). Note that both $\overline{B}, \overline{C}$ are *regular* local rings, and the condition about Kähler differentials implies that they of the same dimension. Consequently, $\overline{B} \to \overline{C}$ is *injective*: if it were not injective, then the dimension of $\mathrm{im}(\overline{B} \to \overline{C})$ would be *less* than $\dim \overline{B} = \dim \overline{C}$. But since $\overline{C}$ is unramified over $\mathrm{im}(\overline{B} \to \overline{C})$, the dimension can only drop: $\dim \overline{C} \leqslant \dim \mathrm{im}(\overline{B} \to \overline{C})$.[1] This contradicts $\dim \overline{B} = \dim \overline{C}$. It follows that $\overline{B} \to \overline{C}$ is injective, and hence flat by **??** below (one can check that there is no circularity).

The local structure theory

We know two easy ways of getting an unramified morphism out of a ring R. First, we can take a standard étale morphism, which is necessarily unramified; next we can take a quotient of that. The local structure theory states that this is all we can have, locally.

Warning: this section will use Zariski's Main Theorem, which is not in this book yet.

For this we introduce a definition.

[1]This follows by the surjection of modules of Kähler differentials, in view of **??**.

3.3.11 Definition Let R be a commutative ring, S an R-algebra of finite type. Let $\mathfrak{q} \in \operatorname{Spec} S$ and $\mathfrak{p} \in \operatorname{Spec} R$ be the image. Then S is called **unramified at $\mathfrak{q}$** (resp. **étale at $\mathfrak{p}$**) if $\Omega_{S_\mathfrak{q}/R_\mathfrak{p}} = 0$ (resp. that and $S_\mathfrak{q}$ is $R_\mathfrak{p}$-flat).

Now when works with finitely generated algebras, the module of Kähler differentials is always finitely generated over the top ring. In particular, if $\Omega_{S_\mathfrak{q}/R_\mathfrak{p}} = (\Omega_{S/R})_\mathfrak{q} = 0$, then there is $f \in S - \mathfrak{q}$ with $\Omega_{S_f/R} = 0$. So being unramified at $\mathfrak{q}$ is equivalent to the existence of $f \in S - \mathfrak{q}$ such that S_f is unramified over R. Clearly if S is unramified over R, then it is unramified at all primes, and conversely.

3.3.12 Theorem *Let $\phi : R \to S$ be morphism of finite type, and $\mathfrak{q} \subset S$ prime with $\mathfrak{p} = \phi^{-1}(\mathfrak{q})$. Suppose ϕ is unramified at $\mathfrak{q}$. Then there is $f \in R - \mathfrak{p}$ and $g \in S - \mathfrak{q}$ (divisible by $\phi(f)$) such that the morphism*

$$R_f \to S_g$$

factors as a composite

$$R_f \to (R_f[x]/P)_h \twoheadrightarrow S_g$$

where the first is a standard étale morphism and the second is a surjection. Moreover, we can arrange things such that the fibers above $\mathfrak{p}$ are isomorphic.

Proof. We shall assume that R is *local* with maximal ideal $\mathfrak{p}$. Then the question reduces to finding $g \in S$ such that S_g is a quotient of an algebra standard étale over R. This reduction is justified by the following argument: if R is not necessarily local, then the morphism $R_\mathfrak{p} \to S_\mathfrak{p}$ is still unramified. If we can show that there is $g \in S_\mathfrak{p} - \mathfrak{q}S_\mathfrak{p}$ such that $(S_\mathfrak{p})_g$ is a quotient of a standard étale $R_\mathfrak{p}$-algebra, it will follow that there is $f \notin \mathfrak{p}$ such that the same works with $R_f \to S_{gf}$.

We shall now reduce to the case where S is a finite R-algebra. Let R be local, and let $R \to S$ be unramified at $\mathfrak{q}$. By assumption, S is finitely generated over R. We have seen by corollary 3.1.16 that S is quasi-finite over R at $\mathfrak{q}$. By Zariski's Main Theorem (**??**), there is a finite R-algebra S' and $\mathfrak{q}' \in \operatorname{Spec} S'$ such that S near $\mathfrak{q}$ and S' near $\mathfrak{q}'$ are isomorphic (in the sense that there are $g \in S - \mathfrak{q}$, $h \in S' - \mathfrak{q}'$ with $S_g \simeq S'_h$). Since S' must be unramified at $\mathfrak{q}'$, we can assume at the outset, by replacing S by S', that $R \to S$ is finite and unramified at $\mathfrak{q}$.

We shall now reduce to the case where S is generated by one element as R-algebra. This will occupy us for a few paragraphs.

We have assumed that R is a local ring with maximal ideal $\mathfrak{p} \subset R$; the maximal ideals of S are finite, say, $\mathfrak{q}, \mathfrak{q}_1, \dots, \mathfrak{q}_r$ because S is finite over R; these all contain $\mathfrak{p}$ by Nakayama. These are no inclusion relations among $\mathfrak{q}$ and the $\mathfrak{q}_i$ as $S/\mathfrak{p}S$ is an artinian ring.

Now $S/\mathfrak{q}$ is a finite separable field extension of $R/\mathfrak{p}$ by theorem 3.1.12; indeed, the morphism $R/\mathfrak{p} \to S/\mathfrak{p}S \to S/\mathfrak{q}$ is a composite of unramified extensions and is thus unramified. In particular, by the primitive element theorem, there is $x \in S$ such that x is a generator of the field extension $R/\mathfrak{p} \to S/\mathfrak{q}$. We can also choose x to lie in the other $\mathfrak{q}_i$ by the Chinese remainder theorem. Consider the subring $C = R[x] \subset S$. It has a maximal ideal $\mathfrak{s}$ which is the intersection of $\mathfrak{q}$ with C. We are going to show that locally, C and S look the same.

3.3.13 Lemma (Reduction to the monogenic case) *Let $(R, \mathfrak{p})$ be a local ring and S a finite R-algebra. Let $\mathfrak{q}, \mathfrak{q}_1, \dots, \mathfrak{q}_r \in \operatorname{Spec} S$ be the prime ideals lying above $\mathfrak{p}$. Suppose S is unramified at $\mathfrak{q}$.*

Then there is $x \in S$ such that the rings $R[x] \subset S$ and S are isomorphic near $\mathfrak{q}$: more precisely, there is $g \in R[x] - \mathfrak{q}$ with $R[x]_g = S_g$.

Proof. Choose x as in the paragraph preceding the statement of the lemma. Define $\mathfrak{s}$ in the same way. We have morphisms

$$R \to C_{\mathfrak{s}} \to S_{\mathfrak{s}}$$

where $S_{\mathfrak{s}}$ denotes S localized at $C - \mathfrak{s}$, as usual. The second morphism here is finite. However, we claim that $S_{\mathfrak{s}}$ is in fact a local ring with maximal ideal $\mathfrak{q}S_{\mathfrak{s}}$; in particular, $S_{\mathfrak{s}} = S_{\mathfrak{q}}$. Indeed, S can have no maximal ideals other than $\mathfrak{q}$ lying above $\mathfrak{s}$; for, if $\mathfrak{q}_i$ lay over $\mathfrak{s}$ for some i, then $x \in \mathfrak{q}_i \cap C = \mathfrak{s}$. But $x \notin \mathfrak{s}$ because x is not zero in $S/\mathfrak{q}$.

It thus follows that $S_{\mathfrak{s}}$ is a local ring with maximal ideal $\mathfrak{q}S_{\mathfrak{s}}$. In particular, it is equal to $S_{\mathfrak{q}}$, which is a localization of $S_{\mathfrak{s}}$ at the maximal ideal. In particular, the morphism

$$C_{\mathfrak{s}} \to S_{\mathfrak{s}} = S_{\mathfrak{q}}$$

is finite. Moreover, we have $\mathfrak{s}S_{\mathfrak{q}} = \mathfrak{q}S_{\mathfrak{q}}$ by unramifiedness of $R \to S$. So since the residue fields are the same by choice of x, we have $\mathfrak{s}S_{\mathfrak{q}} + C_{\mathfrak{s}} = S_{\mathfrak{q}}$. Thus by Nakyama's lemma, we find that $S_{\mathfrak{s}} = S_{\mathfrak{q}} = C_{\mathfrak{s}}$.

There is thus an element $g \in C - \mathfrak{r}$ such that $S_g = C_g$. In particular, S and C are isomorphic near $\mathfrak{q}$. ∎

We can thus replace S by C and assume that C has one generator.

With this reduction now made, we proceed. We are now considering the case where S is generated by one element, so a quotient $S = R[X]$ for some monic polynomial P. Now $\overline{S} = S/\mathfrak{p}S$ is thus a quotient of $k[X]$, where $k = R/\mathfrak{p}$ is the residue field. It thus follows that

$$\overline{S} = k[X]/(\overline{P})$$

for $\overline{P}$ a monic polynomial, as $\overline{S}$ is a finite k-vector space.

Suppose $\overline{P}$ has degree n. Let $x \in S$ be a generator of S/R. We know that $1, x, \dots, x^{n-1}$ has reductions that form a k-basis for $S \otimes_R k$, so by Nakayama they generate S as an R-module. In particular, we can find a monic polynomial P of degree n such that $P(x) = 0$. It follows that the reduction of P is necessarily $\overline{P}$. So we have a surjection

$$R[X]/(P) \twoheadrightarrow S$$

which induces an isomorphism modulo $\mathfrak{p}$ (i.e. on the fiber).

Finally, we claim that we can modify $R[X]/P$ to make a standard étale algebra. Now, if we let $\mathfrak{q}'$ be the preimage of $\mathfrak{q}$ in $R[X]/P$, then we have morphisms of local rings

$$R \to (R[X]/P)_{\mathfrak{q}'} \to S_{\mathfrak{q}}.$$

The claim is that $R[X]/(P)$ is unramified over R at $\mathfrak{q}'$.

To see this, let $T = (R[X]/P)_{\mathfrak{q}'}$. Then, since the fibers of T and $S_{\mathfrak{q}}$ are the same at $\mathfrak{p}$, we have that

$$\Omega_{T/R} \otimes_R k(\mathfrak{p}) = \Omega_{T \otimes_R k(\mathfrak{p})/k(\mathfrak{p})} = \Omega_{(S_{\mathfrak{q}}/\mathfrak{p}S_{\mathfrak{q}})/k(\mathfrak{p})} = 0$$

as S is R-unramified at $\mathfrak{q}$. It follows that $\Omega_{T/R} = \mathfrak{p}\Omega_{T/R}$, so a fortiori $\Omega_{T/R} = \mathfrak{q}\Omega_{T/R}$; since this is a finitely generated T-module, Nakayama's lemma implies that is zero. We conclude that $R[X]/P$ is unramified at $\mathfrak{q}'$; in particular, by the Kähler differential criterion, the image of the derivative P' is not in $\mathfrak{q}'$. If we localize at the image of P', we then get what we wanted in the theorem. □

We now want to deduce a corresponding (stronger) result for *étale* morphisms. Indeed, we prove:

3.3.14 Theorem *If $R \to S$ is étale at $\mathfrak{q} \in \operatorname{Spec} S$ (lying over $\mathfrak{p} \in \operatorname{Spec} R$), then there are $f \in R - \mathfrak{p}, g \in S - \mathfrak{q}$ such that the morphism $R_f \to S_g$ is a standard étale morphism.*

Proof. By localizing suitably, we can assume that $(R, \mathfrak{p})$ is local, and (in view of **??**), $R \to S$ is a quotient of a standard étale morphism

$$(R[X]/P)_h \twoheadrightarrow S$$

with the kernel some ideal I. We may assume that the surjection is an isomorphism modulo $\mathfrak{p}$, moreover. By localizing S enough[2] we may suppose that S is a *flat* R-module as well.

Consider the exact sequence of $(R[X]/P)_h$-modules

$$0 \to I \to (R[X]/P)_h/I \to S \to 0.$$

Let $\mathfrak{q}'$ be the image of $\mathfrak{q}$ in $\operatorname{Spec}(R[X]/P)_h$. We are going to show that the first term vanishes upon localization at $\mathfrak{q}'$. Since everything here is finitely generated, it will follow that after further localization by some element in $(R[X]/P)_h - \mathfrak{q}'$, the first term will vanish. In particular, we will then be done.

Everything here is a module over $(R[X]/P)_h$, and certainly a module over R. Let us tensor everything over R with $R/\mathfrak{p}$; we find an exact sequence

$$I \to S/\mathfrak{p}S \to S/\mathfrak{p}S \to 0;$$

we have used the fact that the morphism $(R[X]/P)_h \to S$ was assumed to induce an isomorphism modulo $\mathfrak{p}$.

However, by étaleness we assumed that S was *R-flat*, so we find that exactness holds at the left too. It follows that

$$I = \mathfrak{p}I,$$

so a fortiori

$$I = \mathfrak{q}'I,$$

which implies by Nakayama that $I_{\mathfrak{q}'} = 0$. Localizing at a further element of $(R[X]/P)_h - \mathfrak{q}'$, we can assume that $I = 0$; after this localization, we find that S looks *precisely* a standard étale algebra. □

[2]We are not assuming S finite over R here,

Permanence properties of étale morphisms

We shall now return to (more elementary) commutative algebra, and discuss the properties that an étale extension $A \to B$ has. An étale extension is not supposed to make B differ too much from A, so we might expect some of the same properties to be satisfied.

We might not necessarily expect global properties to be preserved (geometrically, an open imbedding of schemes is étale, and that does not necessarily preserve global properties), but local ones should be.

Thus the right definition for us will be the following:

3.3.15 Definition A morphism of local rings $(A, \mathfrak{m}_A) \to (B, \mathfrak{m}_B)$ is **local-unramified** $\mathfrak{m}_A B$ is the maximal ideal of B and $B/\mathfrak{m}_B$ is a finite separable extension of $A/\mathfrak{m}_A$.

A morphism of local rings $A \to B$ is **local-étale** if it is flat and local-unramified.

3.3.16 Proposition *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a local-étale morphism of noetherian local rings. Then* $\dim R = \dim S$.

Proof. Indeed, we know that $\mathfrak{m}S = \mathfrak{n}$ because $R \to S$ is local-unramified. Also $R/\mathfrak{m} \to S/\mathfrak{n}$ is a finite separable extension. We have a natural morphism

$$\mathfrak{m} \otimes_R S \to \mathfrak{n}$$

which is injective (as the map $\mathfrak{m} \otimes_R S \to S$ is injective by flatness) and consequently is an isomorphism. More generally, $\mathfrak{m}^n \otimes_R S \simeq \mathfrak{n}^n$ for each n. By flatness again, it follows that

$$\mathfrak{m}^n/\mathfrak{m}^{n+1} \otimes_{R/\mathfrak{m}} (S/\mathfrak{n}) = \mathfrak{m}^n/\mathfrak{m}^{n+1} \otimes_R S \simeq \mathfrak{n}^n/\mathfrak{n}^{n+1}. \tag{3.3.1}$$

Now if we take the dimensions of these vector spaces, we get polynomials in n; these polynomials are the dimensions of R, S, respectively. It follows that $\dim R = \dim S$. ∎

3.3.17 Proposition *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a local-étale morphism of noetherian local rings. Then* $\operatorname{depth} R = \operatorname{depth} S$.

Proof. We know that a non-zero-divisor in R maps to a non-zero-divisor in S. Thus by an easy induction we reduce to the case where $\operatorname{depth} R = 0$. This means that $\mathfrak{m}$ is an associated prime of R; there is thus some $x \in R$, nonzero (and necessarily a non-unit) such that the annihilator of x is all of $\mathfrak{m}$. Now x is a nonzero element of S, too, as the map $R \to S$ is an inclusion by flatness. It is then clear that $\mathfrak{n} = \mathfrak{m}S$ is the annilhilator of x in S, so $\mathfrak{n}$ is an associated prime of S too. ∎

3.3.18 Corollary *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a local-étale morphism of noetherian local rings. Then R is regular (resp. Cohen-Macaulay) if and only if S is.*

Proof. The results proposition 3.3.17 and proposition 3.3.16 immediately give the result about Cohen-Macaulayness. For regularity, we use (3.3.1) with $n = 1$ to see at once that the embedding dimensions of R and S are the same. ∎

Recall, however, that regularity of S implies that of R if we just assume that $R \to S$ is *flat* (by Serre's characterization of regular local rings as those having finite global dimension).

We shall next show that reducedness is preserved under étale extensions. We shall need another hypothesis, though, that the map of local rings be essentially of finite type. This will always be the case in situations of interest, when we are looking at the map on local rings induced by a morphism of rings of finite type.

3.3.19 Proposition *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a local-étale morphism of noetherian local rings. Suppose S is essentially of finite type over R. Then S is reduced if and only if R is reduced.*

Proof. As $R \to S$ is injective by (faithful) flatness, it suffices to show that if R is reduced, so is S. Now there is an imbedding $R \to \prod_{\mathfrak{p} \text{ minimal}} R/\mathfrak{p}$ of R into a product of local domains. We get an imbedding of S into a product of local rings $\prod S/\mathfrak{p}S$. Each $S/\mathfrak{p}S$ is essentially of finite type over $R/\mathfrak{p}$, and local-étale over it too.

We are reduced to showing that each $S/\mathfrak{p}S$ is reduced. So we need only show that a local-étale, essentially of finite type local ring over a local noetherian domain is reduced.

So suppose A is a local noetherian domain, B a local-étale, essentially of finite type local A-algebra. We want to show that B is reduced, and then we will be done. Now A imbeds into its field of fractions K; thus B imbeds into $B \otimes_A K$. Then $B \otimes_A K$ is formally unramified over K and is essentially of finite type over K. This means that $B \otimes_A K$ is a product of fields by the usual classification, and is in particular reduced. Thus B was itself reduced. □

To motivate the proof that normality is preserved, though, we indicate another proof of this fact, which does not even use the essentially of finite type hypothesis. Recall that a noetherian ring A is reduced if and only if for every prime $\mathfrak{p} \in \operatorname{Spec} A$ of height zero, $A_{\mathfrak{p}}$ is regular (i.e., a field), and for every prime $\mathfrak{p}$ of height > 0, $R_{\mathfrak{p}}$ has depth at least one. See **??**.

So suppose $R \to S$ is a local-étale and suppose R is reduced. We are going to apply the above criterion, together with the results already proved, to show that S is reduced.

Let $\mathfrak{q} \in \operatorname{Spec} S$ be a minimal prime, whose image in $\operatorname{Spec} R$ is $\mathfrak{p}$. Then we have a morphism

$$R_{\mathfrak{p}} \to S_{\mathfrak{q}}$$

which is locally of finite type, flat, and indeed local-étale, as it is formally unramified (as $R \to S$ was). We know that $\dim R_{\mathfrak{p}} = \dim S_{\mathfrak{q}}$ by proposition 3.3.16, and consequently since $R_{\mathfrak{p}}$ is regular, so is $S_{\mathfrak{q}}$. Thus the localization of S at any minimal prime is regular.

Next, if $\mathfrak{q} \in \operatorname{Spec} S$ is such that $S_{\mathfrak{q}}$ has height has positive dimension, then $R_{\mathfrak{p}} \to S_{\mathfrak{q}}$ (where $\mathfrak{p}$ is as above) is local-étale and consequently $\dim R_{\mathfrak{q}} = \dim S_{\mathfrak{q}} > 0$. Thus, $\operatorname{depth} R_{\mathfrak{p}} = \operatorname{depth} S_{\mathfrak{q}} > 0$ because R was reduced. It follows that the above criterion is valid for S.

Recall that a noetherian ring is a *normal* domain if it is integrally closed in its quotient field, and simply *normal* if all its localizations are normal domains; this equates to the ring being a product of normal domains. We want to show that this is preserved under étaleness. To do this, we shall use a criterion similar to that used at the end of the last section. We have the following important criterion for normality.

3.3.20 Theorem (Serre) *Let A be a noetherian ring. Then A is normal if and only if for all $\mathfrak{p} \in \operatorname{Spec} R$:*

1. *If* $\dim A_{\mathfrak{p}} \leqslant 1$, *then* $A_{\mathfrak{p}}$ *is regular.*
2. *If* $\dim A_{\mathfrak{p}} \geqslant 2$, *then* $\operatorname{depth} A_{\mathfrak{p}} \geqslant 2$.

This is discussed in **??**.

From this, we will be able to prove without difficulty the next result.

3.3.21 Proposition *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a local-étale morphism of noetherian local rings. Suppose S is essentially of finite type over R. Then S is normal if and only if R is normal.*

Proof. This is proved in the same manner as the result for reducedness was proved at the end of the previous subsec. For instance, suppose R normal. Let $\mathfrak{q} \in \operatorname{Spec} S$ be arbitrary, contracting to $\mathfrak{p} \in \operatorname{Spec} R$. If $\dim S_{\mathfrak{q}} \leqslant 1$, then $\dim R_{\mathfrak{p}} \leqslant 1$ so that $R_{\mathfrak{p}}$, hence $S_{\mathfrak{q}}$ is regular. If $\dim S_{\mathfrak{q}} \geqslant 2$, then $\dim R_{\mathfrak{p}} \geqslant 2$, so $\operatorname{depth} S_{\mathfrak{q}} = \operatorname{depth} R_{\mathfrak{p}} \geqslant 2$. ☐

We mention a harder result:

3.3.22 Theorem *If $f : (R, \mathfrak{m}) \to (S, \mathfrak{n})$ is local-unramified, injective, and essentially of finite type, with R normal and noetherian, then $R \to S$ is local-étale. Thus, an injective unramified morphism of finite type between noetherian rings, whose source is a normal domain, is étale.*

A priori, it is not obvious at all that $R \to S$ should be flat. In fact, proving flatness directly seems to be difficult, and we will have to use the local structure theory for *unramified* morphisms together with nontrivial facts about étale morphisms to establish this result.

Proof. We essentially follow **?** in the proof. Clearly, only the local statement needs to be proved.

We shall use the (non-elementary, relying on ZMT) structure theory of unramified morphisms, which implies that there is a factorization of $R \to S$ via

$$(R, \mathfrak{m}) \xrightarrow{g} (T, \mathfrak{q}) \xrightarrow{h} (S, \mathfrak{n}),$$

where all morphisms are local homomorphisms of local rings, $g : R \to T$ is local-étale and essentially of finite type, and $h : T \to S$ is surjective. This was established in **??**.

We are going to show that h is an isomorphism, which will complete the proof. Let K be the quotient field of R. Consider the diagram

$$\begin{tikzcd} R \arrow[r, "g"] \arrow[d] & T \arrow[r, "h"] \arrow[d] & S \arrow[d] \\ K \arrow[r, "g\otimes 1"] & T \otimes_R K \arrow[r, "h\otimes 1"] & S \otimes_R K. \end{tikzcd}$$

Now the strategy is to show that h is injective. We will prove this by chasing around the diagram.

Here $R \to S$ is formally unramified and essentially of finite type, so $K \to S \otimes_R K$ is too, and $S \otimes_R K$ is in particular a finite product of separable extensions of K. The claim is that it is nonzero; this follows because $f : R \to S$ is injective, and $S \to S \otimes_R K$ is injective because localization is exact. Consequently $R \to S \otimes_R K$ is injective, and the target must be nonzero.

As a result, the surjective map $h \otimes 1 : T \otimes_R K \to S \otimes_R K$ is nonzero. Now we claim that $T \otimes_R K$ is a field. Indeed, it is an étale extension of K (by base-change), so it is a product of fields. Moreover, T is a normal domain since R is (by proposition 3.3.21) and $R \to T$ is injective by flatness, so the localization $T \otimes_R K$ is a domain as well. Thus it must be a field. In particular, the map $h \otimes 1 : T \otimes_R K \to S \otimes_R K$ is a surjection from a field to a product of fields. It is thus an *isomorphism.*

Finally, we can show that h is injective. Indeed, it suffices to show that the composite $T \to T \otimes_R K \to S \otimes_R K$ is injective. But the first map is injective as it is a map from a domain to a localization, and the second is an isomorphism (as we have just seen). So h is injective, hence an isomorphism. Thus $T \simeq S$, and we are done. ☐

Note that this *fails* if the source is not normal.

3.3.23 Example Consider a nodal cubic C given by $y^2 = x^2(x-1)$ in $\mathbb{A}^2_k$ over an algebraically closed field k. As is well-known, this curve is smooth except at the origin. There is a map $\overline{C} \to C$ where $\overline{C}$ is the normalization; this is a finite map, and a local isomorphism outside of the origin.

The claim is that $\overline{C} \to C$ is unramified but not étale. If it were étale, then C would be smooth since $\overline{C}$ is. So it is not étale. We just need to see that it is unramified, and for this we need only see that the map is unramified at the origin.

We may compute: the normalization of C is given by $\overline{C} = \mathbb{A}^1_k$, with the map

$$t \mapsto (t^2+1, t(t^2+1)).$$

Now the two points ± 1 are both mapped to 0. We will show that

$$\mathcal{O}_{C,0} \to \mathcal{O}_{\mathbb{A}^1_k,1}$$

is local-unramified; the other case is similar. Indeed, any line through the origin which is not a tangent direction will be something in $\mathfrak{m}_{C,0}$ that is mapped to a uniformizer in $\mathcal{O}_{\mathbb{A}^1_k,1}$. For instance, the local function $x \in \mathcal{O}_{C,0}$ is mapped to the function $t \mapsto t^2+1$ on $\mathbb{A}^1_k$, which has a simple zero at 1 (or -1). It follows that the maximal ideal $\mathfrak{m}_{C,0}$ generates the maximal ideal of $\mathcal{O}_{\mathbb{A}^1_k,1}$ (and similarly for -1).

Application to smooth morphisms

We now want to show that the class of étale morphisms essentially determines the class of smooth morphisms. Namely, we are going to show that smooth morphisms are those that look étale-locally like étale morphisms followed by projection from affine space. (Here "projection from affine space" is the geometric picture: in terms of commutative rings, this is the embedding $A \hookrightarrow A[x_1, \dots, x_n]$.)

Here is the first goal:

3.3.24 Theorem *Let $f : (A, \mathfrak{m}) \to (B, \mathfrak{n})$ be an essentially of finite presentation, local morphism of local rings. Then f is formally smooth if and only if there exists a factorization*

$$A \to C \to B$$

where $(C, \mathfrak{q})$ is a localization of the polynomial ring $A[X_1, \dots, X_n]$ at a prime ideal with $A \to C$ the natural embedding, and $C \to B$ a formally étale morphism.

For convenience, we have stated this result for local rings, but we can get a more general criterion as well (see below). This states that smooth morphisms, étale locally, look like the imbedding of a ring into a polynomial ring. In **?**, this is in fact how smooth morphisms are *defined.*

Proof. First assume f smooth. We know then that $\Omega_{B/A}$ is a finitely generated projective B-module, hence free, say of rank n. There are $t_1, \dots, t_n \in B$ such that $\{dt_i\}$ forms a basis for $\Omega_{B/A}$: namely, just choose a set of such elements that forms a basis for $\Omega_{B/A} \otimes_B B/\mathfrak{n}$ (since these elements generate $\Omega_{B/A}$).

Now these elements $\{t_i\}$ give a map of rings $A[X_1, \dots, X_n] \to B$. We let $\mathfrak{q}$ be the pre-image of $\mathfrak{n}$ (so $\mathfrak{n}$ contains the image of $\mathfrak{m} \subset A$), and take $C = C = A[X_1, \dots, X_n]_{\mathfrak{q}}$. This gives local homomorphisms $A \to C, C \to B$. We only need to check that $C \to B$ is étale. But the map

$$\Omega_{C/A} \otimes_C B \to \Omega_{B/A}$$

is an isomorphism, by construction. Since C, B are both formally smooth over A, we find that $C \to B$ is étale by the characterization of étaleness via cotangent vectors (proposition 3.3.10).

The other direction, that f is formally smooth if it admits such a factorization, is clear because the localization of a polynomial algebra is formally smooth, and a formally étale map is clearly formally smooth. ☐

3.3.25 Corollary *Let $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ be a formally smooth, essentially of finite type morphism of noetherian rings. Then if R is normal, so is S. Ditto for reduced.*

Proof.

Lifting under nilpotent extensions

In this subsec, we consider the following question. Let A be a ring, $I \subset A$ an ideal of square zero, and let $A_0 = A/I$. Suppose B_0 is a flat A_0-algebra (possibly satisfying other conditions). Then, we ask if there exists a flat A-algebra B such that $B_0 \simeq B \otimes_A A_0 = B/IB$. If there is, we say that B can be *lifted* along the nilpotent thickening from B_0 to B—we think of B as the mostly the same as B_0, but with some additional "fuzz" (given by the additional nilpotents).

We are going to show that this can *always* be done for étale algebras, and that this always can be done *locally* for smooth algebras. As a result, we will get a very simple characterization of what finiteétale algebras over a complete (and later, henselian) local ring look like: they are the same as étale extensions of the residue field (which we have classified completely).

In algebraic geometry, one spectacular application of these ideas is Grothendieck's proof in **?** that a smooth projective curve over a field of characteristic p can be "lifted" to characteristic zero. The idea is to lift it successively along nilpotent thickenings of the base field, bit by bit (for instance, $\mathbb{Z}/p^n\mathbb{Z}$ of $\mathbb{Z}/p\mathbb{Z}$), by using the techniques of this subsec; then, he uses hard existence results in formal geometry to show that this compatible system of nilpotent thickenings comes from a curve over a DVR (e.g. the p-adic numbers). The application in mind is the (partial) computation of the étale fundamental group of a smooth projective curve over a field of positive characteristic. We will only develop some of the more basic ideas in commutative algebra.

Namely, here is the main result. For a ring A, let $\mathsf{Et}(A)$ denote the category of étale A-algebras (and A-morphisms). Given $A \to A'$, there is a natural functor $\mathsf{Et}(A) \to \mathsf{Et}(A')$ given by base-change.

3.3.26 Theorem *Let $A \to A_0$ be a surjective morphism whose kernel is nilpotent. Then $\mathsf{Et}(A) \to \mathsf{Et}(A_0)$ is an equivalence of categories.*

Spec A and Spec A_0 are identical topologically, so this result is sometimes called the topological invariance of the étale site. Let us sketch the idea before giving the proof. Full faithfulness is the easy part, and is essentially a restatement of the nilpotent lifting property. The essential surjectivity is the non-elementary part, and relies on the local structure theory. Namely, we will show that a standard étale morphism can be lifted (this is essentially trivial). Since an étale morphism is locally standard étale, we can *locally* lift an étale A_0-algebra to an étale A-algebra. We next "glue" the local liftings using the full faithfulness.

Proof. Without loss of generality, we can assume that the ideal defining A_0 has square zero. Let B, B' be étale A-algebras. We need to show that

$$\hom_A(B, B') = \hom_{A_0}(B_0, B_0'),$$

where B_0, B_0' denote the reductions to A_0 (i.e. the base change). But $\hom_{A_0}(B_0, B_0') = \hom_A(B, B_0')$, and this is clearly the same as $\hom_A(B, B')$ by the definition of an étale morphism. So full faithfulness is automatic.

The trickier part is to show that any étale A_0-algebra can be lifted to an étale A-algebra. First, note that a standard étale A_0-algebra of the form $(A_0[X]/(P(X))_Q$ can be lifted to A—just lift P and Q. The condition that it be standard étale is invertibility of P', which is unaffected by nilpotents.

Now the strategy is to glue these appropriately. The details should be added at some point, but they are not. **To be added: details** □

Notation

covering	$\mathcal{U}$, $\mathcal{V}$
field	$\mathbb{F}$, $\mathbb{K}$, $\Bbbk$
— , of complex numbers	$\mathbb{C}$
— , of rational numbers	$\mathbb{Q}$
— , of real numbers	$\mathbb{R}$
fiber bundle	(E, B, p, F), $p : E \to B$, or $\pi : E \to B$, where F denotes typical fiber
function	f, g, h, $\ldots$, φ, ψ, $\ldots$
group	G, H, K
— , general linear	$\mathsf{GL}(\mathbb{K}, n)$, where $\mathbb{K} = \mathbb{R}$ or $= \mathbb{C}$
interval, open	(a, b), where $a < b$
— , closed	$[a, b]$, where $a \leqslant b$
— , half-open	$(a, b]$ or $[a, b)$, where $a < b$
manifold	M, N, P, Q, $\ldots$
open set	U, V, W, O, $\ldots$
pair	(x, y), where $x \in X$, $y \in Y$, and X, Y are sets
ring	R, S
sheaf	$\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, $\mathcal{F}$, $\ldots$
— , continuous functions	$\mathcal{C}$
— , smooth functions	$\mathcal{C}^\infty$
— , real analytic functions	$\mathcal{C}^\omega$
— , holomorphic functions	$\mathcal{O}$
topological space	X, Y, Z, A, B, $\ldots$
topology	$\mathcal{T}$, $\mathcal{S}$
vector bundle	$\pi : E \to B$

Bibliography

Alexandroff, P. & Hopf, H. (1965). *Topologie.* Erster Band. Grundbegriffe der mengentheoretischen Topologie, Topologie der Komplexe, topologische Invarianzsätze und anschliessende Begriffsbildungen, Verschlingungen im n-dimensionalen euklidischen Raum, stetige Abbildungen von Polyedern. Chelsea Publishing Co., New York.

Bourbaki, N. (1989). *Algebra I. Chapters 1–3.* Elements of Mathematics (Berlin). Berlin: Springer-Verlag. Translated from the French, Reprint of the 1974 English translation.

Bourbaki, N. (1998). *General topology. Chapters 1–4.* Elements of Mathematics (Berlin). Springer-Verlag, Berlin. Translated from the French, Reprint of the 1989 English translation.

Bourbaki, N. (2004). *Theory of sets.* Elements of Mathematics (Berlin). Springer-Verlag, Berlin. Reprint of the 1968 English translation [Hermann, Paris].

Brown, R. (2006). *Topology and groupoids.* BookSurge, LLC, Charleston, SC. Third edition of ıt Elements of modern topology [McGraw-Hill, New York, 1968], With 1 CD-ROM (Windows, Macintosh and UNIX).

Cartan, H. & Eilenberg, S. (1999). *Homological Algebra.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ. With an appendix by David A. Buchsbaum, Reprint of the 1956 original.

Dedekind, R. (1893). *Was Sind Und Was Sollen Die Zahlen?* (second ed.). Braunschweig: Friedrich Vieweg und Sohn.

Dold, A. (1995). *Lectures on Algebraic Topology.* Springer-Verlag Berlin Heidelberg. reprint of the 2nd edition (November 1980), originally published as volume 200 in the series: Grundlehren der mathematischen Wissenschaften.

Grothendieck, A. (1957a). Classes de faisceaux et théorème de Riemann-Roch. (Mimeographed). Paris.

Grothendieck, A. (1957b). Sur quelques points d'algèbre homologique. *Tôhoku Math. J. (2)*, *9*, 119–221.

Harary, F. (1969). *Graph Theory.* Reading, MA - Menlo Park, CA - London: Addison-Wesley Publishing Co.

Hatcher, A. (2002). *Algebraic Topology.* Cambridge: Cambridge University Press. Available at `http://www.math.cornell.edu/~hatcher/AT/AT.pdf`.

Kashiwara, M. & Schapira, P. (2006). *Categories and Sheaves*, volume 332 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Berlin: Springer-Verlag.

Mac Lane, S. (1998). *Categories for the Working Mathematician* (2nd ed.)., volume 5 of *Graduate Texts in Mathematics*. New York: Springer-Verlag.

Mendelson, E. (2008). *Number Systems and the Foundations of Analysis*. Dover Books on Mathematics. Dover Publications, Inc. Reprint of the Academic Press, New York, 1973 edition.

Moschovakis, Y. (2006). *Notes on set theory* (Second ed.). Undergraduate Texts in Mathematics. New York: Springer.

Savage, A. (2006). Finite-dimensional algebras and quivers. In J.-P. Françoise, G. L. Naber, & S. T. Tsou (Eds.), *Encyclopedia of Mathematical Physics*, volume 2 (pp. 313–320). Oxford: Elsevier.

Steen, L. A. & Seebach, Jr., J. A. (1995). *Counterexamples in topology*. Dover Publications, Inc., Mineola, NY. Reprint of the second (1978) edition.

von Neumann, J. (1923). Zur Einführung der transfiniten Zahlen. *Acta litterarum ac scientiarum Ragiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum*, *1*, 199–208.

Licensing

GNU Free Documentation License
Version 1.3, 3 November 2008

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. C. State on the Title page the name of the publisher of the Modified Version, as the publisher. D. Preserve all the copyright notices of the Document. E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. H. Include an unaltered copy of this License. I. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. K. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version. N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section. O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (c)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with. . . Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.