The CRing Project

Founded by Akhil Mathew

October 11, 2016

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

Part I.

Foundations

1. Category Theory

Introduction

The language of categories is not strictly necessary to understand the basics of commutative algebra, ring theory 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.

1.1. Objects, morphisms, and categories

Definitions and first examples

1.1.1 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.

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

- a collection of sets called *objects*,
- for each pair of objects X, Y a set of *morphisms* $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ such that for every quadruple of objects X, X', Y, Y' the morphism sets $\mathrm{Mor}_{\mathsf{C}}(X, Y)$ and $\mathrm{Mor}_{\mathsf{C}}(X', Y')$ have no common element in case $X \neq X'$ or $Y \neq Y'$,
- 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:

(Cat1) 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.

(Cat2) 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.

1.1.3 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 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 the mentioned examples, 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 $\mathcal{O}_{\mathbb{R},\mathrm{e}}$ and the discrete topology $\mathcal{P}(\mathbb{R})$ (recall that $\mathcal{P}(Y)$ denotes the powerset of a set Y). The identity map $\mathrm{id}_{\mathbb{R}}$ then is continuos from $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ to $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ and continuous from $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ (but not vice versa). Hence, $\mathrm{id}_{\mathbb{R}}$ would be regarded as a morphism both from $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ to $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ and from $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{O}_{\mathbb{R},\mathrm{e}})$ in violation of the requirement that $\mathrm{Mor}_{\mathsf{C}}(X,Y) \cap \mathrm{Mor}_{\mathsf{C}}(X',Y') = \emptyset$ 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, 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) In this book we will almost always consider only locally small categories which means categories where the collection of morphisms between two objects forms a set, or in other words, using language by (Bourbaki, 2004, Chap. II), where the relation of being a morphism between two given objects is *collectivizing*. Unless stated differently, we therefore consider our categories to be locally small.

Here is a simple list of examples.

1.1.4 Example (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.

(f) 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.

1.1.5 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.

1.1.6 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.

1.1.7 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.

1.1.8 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).

1.1.9 Example (Posets as categories) Let $(P, \leq)$ 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

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

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

1.1.10 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.

1.1.11 Example In the category Top of topological spaces, one can in fact recover the "underlying set" of a topological space via the hom-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

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

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

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

1.1.12 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).

1.1.13 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.

1.1.14 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.

1.1.15 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

$$\begin{array}{ccc} X & \xrightarrow{f'} & W \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ Y & \xrightarrow[g']{} & Z \end{array}$$

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:

$$\begin{array}{ccc} \mathrm{Mor}(X,Y) \times \mathrm{Mor}(Y,Z) \times \mathrm{Mor}(Z,W) & \longrightarrow & \mathrm{Mor}(X,Z) \times \mathrm{Mor}(Z,W) \\ \downarrow & & \downarrow \\ \mathrm{Mor}(X,Y) \times \mathrm{Mor}(Y,W) & \longrightarrow & \mathrm{Mor}(X,W). \end{array}$$

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:

1.1.16 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.

1.1.17 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

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

1.1.18 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.

1.1.19 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.

1.1.20 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 preordered set, and make P into a category as in Example 1.1.9. Then P is a poset if and only if two isomorphic objects are equal.

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

1.1.22 Remark If C is a groupoid and A an object of C, the set $\mathrm{Mor}_{\mathsf{C}}(A, A)$ is a groups. A group is essentially the same as a groupoid with one object.

1.1.23 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.

1.1.24 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 whenever X' is another object in C, the map

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

is an injection (of sets). Epimorphisms in a category are defined similarly; note that neither definition makes any reference to *surjections* of sets.

The reader can easily check:

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

1.1.26 Remark Prove **??**.

1.1.27 Remark 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.

1.2. 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 Tor, 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*.

1.2.1 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.

(Fun1) If $X \in \mathsf{C}$, then $F(\mathrm{id}_X)$ is the identity morphism $\mathrm{id}_{F(X)} : F(X) \to F(X)$.

(Fun2) 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

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow^{h} & \downarrow{\scriptstyle g} \\ & & Z \end{array}$$

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

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ & \searrow^{F(h)} & \downarrow{\scriptstyle F(g)} \\ & & F(Z) \end{array}$$

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.

1.2.2 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 11.6.1.

1.2.3 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.

1.2.4 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 13.3) is associative, which is really a statement about composing functors. The following (mostly self-explanatory) definition elucidates this.

1.2.5 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 (Fun1) and (Fun2) are both fulfilled for $G \circ F$.

1.2.6 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 morphisms as functors. Composition is defined as in Definition 1.2.5.

1.2.7 Example (Group actions) Fix a group G. Let us understand what a functor $B_G \to \mathsf{Ens}$ is. Here B_G is the category of Example 1.1.7. 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 $g, h \in G$:

$$\begin{array}{ccc} X & \xrightarrow{\varphi_h} & X \\ & \searrow^{\varphi_{gh}} & \downarrow{\scriptstyle \varphi_g} \\ & & X\,. \end{array}$$

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

1.2.8 Example (Forgetful functors) An important example of functors is given by the following. Let C be a "category of structured sets", see Remark 1.1.3 (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:

1.2.9 Definition A *contravariant functor* $\mathsf{C} \xrightarrow{F} \mathsf{D}$ (between categories C and D) is similar data as in Definition 1.2.1 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 (Fun1) to hold true and

(Func 2)° 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

$$\begin{array}{ccc} A & \longrightarrow & C \\ \downarrow & \nearrow & \\ B & & \end{array}$$

is commutative in C, then the diagram

$$\begin{array}{ccc} F(A) & \longleftarrow & F(C) \\ \uparrow & \swarrow & \\ F(B) & & \end{array}$$

commutes in D.

1.2.10 Remark One can, of course, compose contravariant functors as in Definition 1.2.5. 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 1.2.7, we have:

1.2.11 Example A *contravariant* functor from B_G (defined as in Example 1.1.7) to Ens corresponds to a set with a *right* G-action.

1.2.12 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

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

1.2.13 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.

1.2.14 Example If we map $B_G \to B_G$ 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.

1.2.15 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 $\mathsf{C} \to \mathsf{C}^{op}$. 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.

1.2.16 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.

1.2.17 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 $F(g)$. □

1.2.18 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.

1.2.19 Example 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.

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

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.

1.3. Natural transformations

Definition and some examples

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

- 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

$$\begin{array}{ccc} FX & \xrightarrow{F(f)} & FY \\ {\scriptstyle \eta_X}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ GX & \xrightarrow[G(f)]{} & GY \end{array}$$

  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.

1.3.2 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 $\Bbbk$. 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 $V \mapsto V, V \mapsto V^\vee$ are both covariant functors on C.[1] The first is the identity functor; for the second, if $f : V \to W$ is an isomorphism, then there is induced a transpose map $f^t : W^\vee \to V^\vee$ (defined by sending a map $W \to k$ to the precomposition $V \xrightarrow{f} W \to k$), which is an isomorphism; we can take its inverse. So we have two functors from C to itself, the identity and the 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:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow{\scriptstyle T_V} & & \downarrow{\scriptstyle T_W} \\ V^\vee & \xrightarrow{(f^t)^{-1}} & W^\vee. \end{array}$$

Indeed, fix $d > 1$, and choose $V = k^d$. Identify $V^\vee$ with k^d, and so the map T_V is a d-by-d matrix M with coefficients in k. The requirement is that for each *invertible* d-by-d matrix N, we have

$$(N^t)^{-1} M = MN,$$

by considering the above diagram with $V = W = k^d$, and f corresponding to the matrix N. This is impossible unless $M = 0$, by elementary linear algebra.

Nonetheless, it *is* possible to choose a natural isomorphism

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

To do this, given V, recall that $V^{\vee\vee}$ is the collection of maps $V^\vee \to k$. To give a map $V \to V^{\vee\vee}$ is thus the same as giving linear functions $l_v, v \in V$ such that $l_v : V \to k$ is linear in v. We can do this by letting l_v be "evaluation at v." That is, l_v sends a linear functional $\ell : V \to k$ to $\ell(v) \in k$. We leave it to the reader to check (easily) that this defines a homomorphism $V \to V^{\vee\vee}$, and that everything is natural.

[1] Note that the dual $\vee$ was defined as a *contravariant* functor in **??**.

1.3.3 Remark Suppose there are two functors $B_G \to \mathsf{Ens}$, i.e. G-sets. What is a natural transformation between them?

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 $\mathrm{Fun}(\mathsf{C}, \mathsf{D})$ (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."

1.3.4 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.

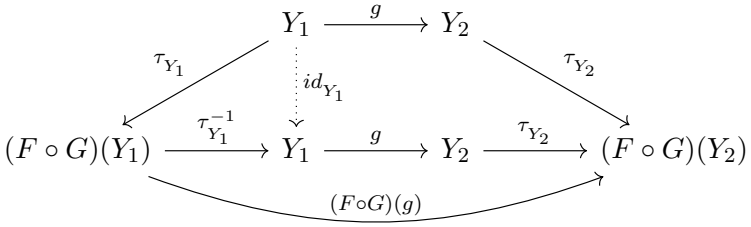
1.3.5 Definition A covariant functor $F : \mathsf{C} \to \mathsf{D}$ is called *fully faithful* if for each pair of objects $X, Y \in \mathsf{C}$ the map $F : \mathrm{Mor}_{\mathsf{C}}(X, Y) \to \mathrm{Mor}_{\mathsf{D}}(FX, FY)$ is a bijection. The functor F is called *essentially surjective* if every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

1.3.6 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.

1.3.7 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.*

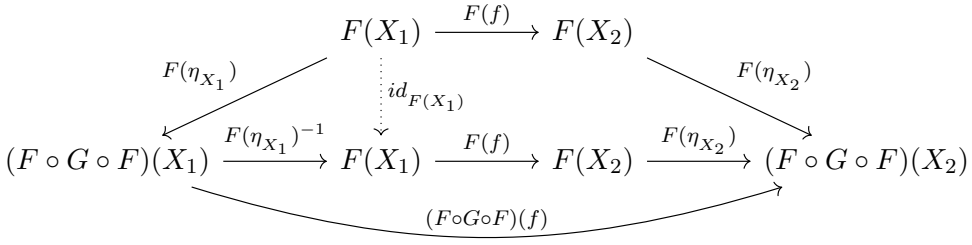
Proof. Let us first show that the condition is sufficient, and assume that F is fully faithful and essentially surjective. By essentially surjectivity we can then fix for any $Y \in \mathrm{Ob}(\mathsf{D})$ some $X_Y \in \mathrm{Ob}(\mathsf{C})$ and an isomorphism $\tau_Y : Y \to F(X)$. The fact that 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 $(F \circ G)(g) = \tau_{Y_2} \circ g \circ \tau_{Y_1}^{-1}$.



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 $(F \circ G \circ F)(f) = \tau_{F(X_2)} \circ F(g) \circ \tau_{F(X_1)}^{-1} =$

$F(\eta_{X_2}) \circ F(g) \circ F(\eta_{X_1})^{-1}$.



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 $f_1 = \eta_{X_2}^{-1} \circ (G \circ F)(f_1) \circ \eta_{X_1} = \eta_{X_2}^{-1} \circ (G \circ F)(f_2) \circ \eta_{X_1} = f_2$. 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 $f = \eta_{X_2}^{-1} \circ G(g) \circ \eta_{X_1}$. Using the diagram in C from Definition 1 again, we see that $(G \circ F)(f) = \eta_{X_2} \circ f \circ \eta_{X_1}^{-1}$ 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. □

1.3.8 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.

1.4. 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

Recall that if we have two sets X and Y, the product $X \times Y$ is the set of all elements of the form (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 they project to on X and Y. In fact, this is the case more generally; if we have an index set I and a product $X = \prod_{i \in I} X_i$, then an element $x \in X$ determined uniquely by where where the projections $p_i(x)$ land in X_i.

To get into the categorical spirit, we should speak not of elements but of maps to X. Here is the general observation: if we have any other set S with maps $f_i : S \to X_i$ then there is a unique map $S \to X = \prod_{i \in I} X_i$ given by sending $s \in S$ to the element $\{f_i(s)\}_{i \in I}$. This leads to the following characterization of a product using only "mapping properties."

1.4.1 Definition Let $\{X_i\}_{i \in I}$ be a collection of objects in some category C. Then an object $P \in \mathsf{C}$ with projections $p_i : P \to X_i$ is said to be the *product* $\prod_{i \in I} X_i$ if the following "universal property" holds: let S be any other object in C with maps $f_i : S \to X_i$. Then there is a unique morphism $f : S \to P$ such that $p_i f = f_i$.

In other words, to map into X is the same as mapping into all the $\{X_i\}$ at once. We have thus given a precise description of how to map into X. Note that, however, the product need not exist! If it does, however, we can express the above formalism by the following natural isomorphism of contravariant functors

$$\mathrm{Mor}(\cdot, \prod_I X_i) \simeq \prod_I \mathrm{Mor}(\cdot, X_i).$$

This is precisely the meaning of the last part of the definition. Note that this observation shows that products in the category of *sets* are really fundamental to the idea of products in any category.

1.4.2 Example One of the benefits of this construction is that an actual category is not specified; thus when we take C to be Ens, we recover the cartesian product notion of sets, but if we take C to be Grp, we achieve the regular notion of the product of groups (the reader is invited to check these statements).

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

1.4.3 Proposition (Uniqueness of products) *Any two products of the collection* $\{X_i\}$ *in* C *are isomorphic by a unique isomorphism commuting with the projections.*

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 $\{X_i\}$ with projections $p_i : X \to X_i$, and let Y be a product of them too, with projections $q_i : Y \to X_i$. Then the claim is that there is a *unique* isomorphism

$$f : X \to Y$$

such that the diagrams below commute for each $i \in I$:

(1.4.3.1)
$$\begin{array}{ccc} X & \xrightarrow{\quad f \quad} & Y \\ & {\scriptstyle p_i} \searrow \quad \swarrow {\scriptstyle q_i} & \\ & X_i. & \end{array}$$

Proof. This is a "trivial" result, and is part of a general fact that objects with the same universal property are always canonically isomorphic. Indeed, note that the projections $p_i : X \to X_i$ and the fact that mapping into Y is the same as mapping into all the X_i gives a unique map $f : X \to Y$ making the diagrams (1.4.3.1) commute. The same reasoning (applied to the $q_i : Y \to X_i$) gives a map $g : Y \to X$ making the diagrams

(1.4.3.2)
$$\begin{array}{ccc} Y & \xrightarrow{\quad g \quad} & X \\ & {\scriptstyle q_i} \searrow \quad \swarrow {\scriptstyle p_i} & \\ & X_i & \end{array}$$

commute. By piecing the two diagrams together, it follows that the composite $g \circ f$ makes the diagram

(1.4.3.3)
$$\begin{array}{ccc} X & \xrightarrow{\quad g \circ f \quad} & X \\ & {\scriptstyle p_i} \searrow \quad \swarrow {\scriptstyle p_i} & \\ & X_i & \end{array}$$

commute. But the identity $\mathrm{id}_X : X \to X$ also would make (1.4.3.3) commute, and 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. □

1.4.4 Remark 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 $(X_i)_{i \in I}$ is the same as mapping out of each of these. We shall later study this construction more generally.

1.4.5 Example Let P be a poset, and make P into a category as in Example 1.1.9. 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.

We shall, in this section, investigate this notion of "universality" more thoroughly.

Initial and terminal objects

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

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

So there is a unique map out of X into each $Y \in \mathsf{C}$. Note that this idea is faithful to the categorical spirit of describing objects in terms of their mapping properties. Initial objects are very easy to map *out* of.

1.4.7 Example If C is Ens, then the empty set $\emptyset$ is an initial object. There is a unique map from the empty set into any other set; one has to make no decisions about where elements are to map when constructing a map $\emptyset \to X$.

1.4.8 Example In the category Grp of groups, the group consisting of one element is an initial object.

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

1.4.9 Example Let P be a poset, and make it into a category as in **??**. 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).

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

1.4.10 Definition A *terminal object* in a category C is an object $Y \in \mathsf{C}$ such that $\mathrm{Mor}_{\mathsf{C}}(X, Y) = *$ for each $X \in \mathsf{C}$.

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.

1.4.11 Example The one point set is a terminal object in Ens.

The important thing about the next "theorems" is the conceptual framework.

1.4.12 Proposition (Uniqueness of the initial or terminal object) *Any two initial (resp. terminal) objects in* 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 maps $f : Y \to Y'$, $g : Y' \to Y$, whose composites must be the identities: we know that $\mathrm{Mor}(Y, Y)$ and $\mathrm{Mor}(Y', Y')$ are one-point sets, so the composites have no other choice to be the identities. This means that the maps $f : Y \to Y'$ and $g : Y' \to Y$ are isomorphisms. □

There is a philosophical point to be made here. We have characterized an object uniquely in terms of mapping properties. We have characterized it *uniquely up to unique isomorphism,* which is really the best one can do in mathematics. Two sets are not generally the "same," but they may be isomorphic up to unique isomorphism. They are different, but the sets are isomorphic up to unique isomorphism. Note also that the argument was essentially similar to that of Proposition 1.4.3.

In fact, we could interpret Proposition 1.4.3 as a special case of Proposition 1.4.12. If C is a category and $\{X_i\}_{i \in I}$ 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 $f_i : Y \to X_i$ for all $i \in I$. A morphism between objects $(Y, \{f_i : Y \to X_i\})$ and $(Z, \{g_i : Z \to X_i\})$ is a map $Y \to Z$ making the obvious diagrams commute. Then a product $\prod X_i$ in C is the same thing as a terminal object in D, as one easily checks from the definitions.

Pushouts and pullbacks

Like always in this chapter let C be a category.

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.

Suppose we have objects A, B, C, X of C.

1.4.13 Definition A commutative square

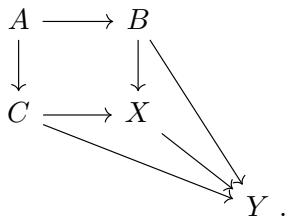
$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & X \end{array} .$$

is called *cocartesian* or a *pushout square* (and X is called the *pushout*) if it satisfies the following universal property:

- Given a commutative diagram

$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & Y \end{array}$$

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



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

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

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

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

$$\begin{array}{ccc} \mathbb{Z}/2 & \longrightarrow & \mathbb{Z}/4 \\ \downarrow & & \downarrow \\ \mathbb{Z}/6 & \longrightarrow & \mathbb{Z}/12 \end{array}$$

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 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.

1.4.15 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:

$$\begin{array}{ccc} R & \longrightarrow & S \\ \downarrow & & \downarrow \\ Q & \longrightarrow & X \end{array}$$

It turns out that the pushout in this case is the tensor product of algebras $S \otimes_R Q$ (see Section 13.3 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.

1.4.16 Proposition *Let* C *be any category. If the pushout of the diagram*

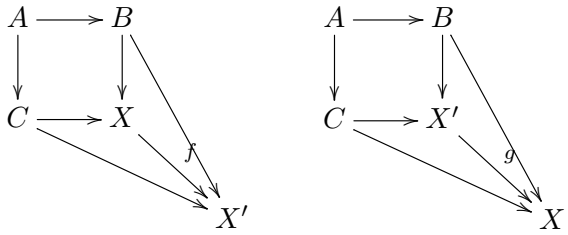
$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \\ C & & \end{array}$$

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:

$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & X \end{array} \qquad \begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & X' \end{array}$$

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



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,

$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & D \end{array},$$

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. □

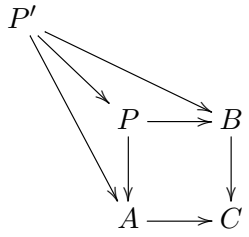
Often when studying categorical constructions, one can create a kind of "dual" construction by reversing the direction of the arrows. This is exactly the relationship between the pushout construction and the pullback construction to be described below. So suppose we have two morphisms $A \to C$ and $B \to C$, forming a diagram

$$\begin{array}{ccc} & & B \\ & & \downarrow \\ A & \longrightarrow & C. \end{array}$$

1.4.17 Definition The *pullback* or *fibered product* of the above diagram is an object P with two morphisms $P \to B$ and $P \to C$ such that the following diagram commutes:

$$\begin{array}{ccc} P & \longrightarrow & B \\ \downarrow & & \downarrow \\ A & \longrightarrow & C \end{array}$$

Moreover, the object P is required to be universal in the following sense: given any P' and maps $P' \to A$ and $P' \to B$ making the square commute, there is a unique map $P' \to P$ making the following diagram commute:



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

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

1.4.19 Example (Requires prerequisites not developed yet) The next example may be omitted without loss of continuity.

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:

$$\begin{array}{ccc} X \times_{\mathrm{Spec}(k)} Y & \longrightarrow & X \\ \downarrow & & \downarrow \\ Y & \longrightarrow & \mathrm{Spec}(k) \end{array}$$

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 $\mathrm{Spec}(A) \times_{\mathrm{Spec}(k)} \mathrm{Spec}(B) = \mathrm{Spec}(A \otimes_k B) = \mathrm{Spec}(k[x, y]) = \mathbb{A}^2_k$.

Colimits

We now want to generalize the pushout. Instead of a shape with A, B, C, we do something more general. Start with a small category I: recall that *smallness* means that the objects

of I form a set. I is to be called the *indexing category*. One is supposed to picture is that I is something like the category

$$\begin{array}{ccc} * & \longrightarrow & * \\ \downarrow & & \\ * & & \end{array}$$

or the category

$$* \rightrightarrows *.$$

We will formulate the notion of a *colimit* which will specialize to the pushout when I is the first case.

So we will look at functors

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

which in the case of the three-element category, will just correspond to diagrams

$$\begin{array}{ccc} A & \longrightarrow & B\ . \\ \downarrow & & \\ C & & \end{array}$$

We will call a *cone* on F (this is an ambiguous term) an object $X \in \mathsf{C}$ equipped with maps $F_i \to X, \forall i \in I$ such that for all maps $i \to i' \in I$, the diagram below commutes:

$$\begin{array}{ccc} F_i & \longrightarrow & X\ . \\ \downarrow & \nearrow & \\ F_{i'} & & \end{array}$$

An example would be a cone on the three-element category above: then this is just a commutative diagram

$$\begin{array}{ccc} A & \longrightarrow & B\ . \\ \downarrow & & \downarrow \\ C & \longrightarrow & D \end{array}$$

1.4.20 Definition The *colimit* of the diagram $F : I \to \mathsf{C}$, written as $\operatorname{colim} F$ or $\operatorname{colim}_I F$ or $\varinjlim_I F$, if it exists, is a cone $F \to X$ with the property that if $F \to Y$ is any other cone, then there is a unique map $X \to Y$ making the diagram

$$\begin{array}{ccc} F & \longrightarrow & X \\ & \searrow & \downarrow \\ & & Y \end{array}$$

commute. (This means that the corresponding diagram with F_i replacing F commutes for each $i \in I$.)

We could form a category D where the objects are the cones $F \to X$, and the morphisms from $F \to X$ and $F \to Y$ are the maps $X \to Y$ that make all the obvious diagrams commute. In this case, it is easy to see that a *colimit* of the diagram is just an initial object in D.

In any case, we see:

1.4.21 Proposition $\operatorname{colim} F$*, if it exists, is unique up to unique isomorphism.*

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

1.4.22 Example Consider the category I visualized as

$$*, *, *, *.$$

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

1.4.23 Example Suppose we had the same category I 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. Again, the direct sum is characterized by the same universal property.

1.4.24 Example Suppose we had the same I $(*, *, *, *)$ the functor took its value in the category of groups. Then the colimit is the free product of the four groups.

1.4.25 Example Suppose we had the same I and the category C was of commutative rings with unit. Then the colimit is the tensor product.

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

1.4.26 Example Take

$$I = * \rightrightarrows *.$$

So a functor $I \to \mathsf{Ens}$ is a diagram

$$A \rightrightarrows B.$$

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

This is the relation *generated* as above, not just as above. It can get tricky.

1.4.27 Definition When I is just a bunch of points $*, *, *, \ldots$ with no non-identity morphisms, then the colimit over I is called the *coproduct*.

We use the coproduct to mean things like direct sums, disjoint unions, and tensor products. If $\{A_i, i \in I\}$ is a collection of objects in some category, then we find the universal property of the coproduct can be stated succinctly:

$$\mathrm{Mor}_{\mathsf{C}}(\bigsqcup_I A_i, B) = \prod \mathrm{Mor}_{\mathsf{C}}(A_i, B).$$

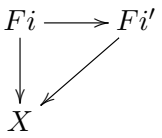
1.4.28 Definition When I is $* \rightrightarrows *$, the colimit is called the *coequalizer*.

1.4.29 Theorem *If* 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:



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 $\square$

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.

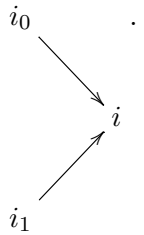
1.4.30 Example The product is an example of a limit where the indexing category is a small category I with no morphisms other than the identity. This example shows the power of universal constructions; by looking at colimits and limits, a whole variety of seemingly unrelated mathematical constructions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

1.4.31 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



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.

1.4.32 Example If I is the category

$$* \to * \to * \to \dots,$$

i.e. the category generated by the poset $\mathbb{Z}_{\geq 0}$, then that is filtered.

1.4.33 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.

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

1.4.35 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.

1.4.36 Proposition *If I is filtered*[2] *and $\mathsf{C} = \mathsf{Ens}, \mathsf{Ab}, \mathsf{Grp}$, etc., and $F : I \to \mathsf{C}$ is a functor, then $\operatorname{colim}_I F$ 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.

[2]Some people say filtering.

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

Alternatively, choose a sequence of numbers $m_1, m_2, \ldots$, such that for all p, n, we have $p^n \mid m_i$ for $i \gg 0$. Then we have a sequence of maps

$$\mathbb{Z} \overset{m_1}{\to} \mathbb{Z} \overset{m_2}{\to} \mathbb{Z} \to \ldots.$$

The colimit of this is $\mathbb{Q}$. There is a quick way of seeing this, which is left to the reader.

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.

1.4.38 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. □

1.4.39 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 $A \to X$.

1.4.40 Definition Suppose C has equalizers. If $A \in \mathsf{C}$ is such that $\mathrm{Mor}_{\mathsf{C}}(A, X) \neq \emptyset$ 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.

1.4.41 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:

1.4.42 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

$$A' \xrightarrow{i} A \xrightarrow{\mathrm{id}} A;$$ □

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 $si = \mathrm{id}_{A'}$. □

1.4.43 Theorem (Freyd) *Let C be a category admitting all small limits.*[3] *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 1.4.41 now implies that C has an initial object. □

Completeness and cocompleteness

1.4.44 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.

1.4.45 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*

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

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 $F : \emptyset \to \mathsf{C}$. 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.

Continuous and cocontinuous functors

1.5. 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.

1.5.1 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*

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

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

1.5.2 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 $X \to X'$.*

1.5.3 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 $Y \mapsto \mathrm{Mor}_{\mathsf{C}}(Y, X)$.

1.5.4 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 $f : Y \to X$.

1.5.5 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:

1.5.6 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[4] 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

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

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.

1.5.7 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 1.4.43. 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. □

1.6. Adjoint functors

According to MacLane, "Adjoint functors arise everywhere." We shall see several examples of adjoint functors in this book (such as Mor and the tensor product). The fact that a functor has an adjoint often immediately implies useful properties about it (for instance, that it commutes with either limits or colimits); this will lead, for instance, to conceptual arguments behind the right-exactness of the tensor product later on.

Definition

Suppose C, D are categories, and let $F : \mathsf{C} \to \mathsf{D}, G : \mathsf{D} \to \mathsf{C}$ be (covariant) functors.

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

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

whenever $c \in \mathsf{C}, d \in \mathsf{D}$. F is said to be the *right adjoint,* and G is the *left adjoint.*

Here "natural" means that the two quantities are supposed to be considered as functors $\mathsf{C}^{op} \times \mathsf{D} \to \mathsf{Ens}$.

1.6.2 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 11.6.1 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 (Lang, 2002, 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

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 $c \in \mathsf{C}$, we know that

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

so that the identity morphism $Fc \to Fc$ (which is natural in c!) corresponds to a map $c \to GFc$ that is natural in c, or equivalently a natural transformation

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

Similarly, we get a natural transformation

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

where the map $FGd \to d$ corresponds to the identity $Gd \to Gd$ under the adjoint correspondence. Here η is called the *unit*, and ϵ the *counit.*

These natural transformations η, ϵ are not simply arbitrary. We are, in fact, going to show that they determine the isomorphism determine the isomorphism $\mathrm{Mor}_{\mathsf{D}}(Fc, d) \simeq \mathrm{Mor}_{\mathsf{C}}(c, Gd)$. This will be a little bit of diagram-chasing.

We know that the isomorphism $\mathrm{Mor}_{\mathsf{D}}(Fc, d) \simeq \mathrm{Mor}_{\mathsf{C}}(c, Gd)$ 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 $\mathsf{C}^{op} \times \mathsf{D}$; by naturality of the adjoint isomorphism, we get a commutative square of sets

$$\begin{array}{ccc} \mathrm{Mor}_{\mathsf{D}}(Fc, Fc) & \xrightarrow{\text{adj}} & \mathrm{Mor}_{\mathsf{C}}(c, GFc) \\ \downarrow{\scriptstyle \phi_*} & & \downarrow{\scriptstyle G(\phi)_*} \\ \mathrm{Mor}_{\mathsf{D}}(Fc, d) & \xrightarrow{\text{adj}} & \mathrm{Mor}_{\mathsf{C}}(c, Gd) \end{array}$$

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 $\phi : Fc \to d$:

1.6.3 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

$$\psi = G(\phi) \circ \eta(c) : c \to Gd \tag{1.6.3.1}$$

(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

$$\epsilon(d) \circ F(\psi) : Fc \to FGd \to d. \tag{1.6.3.2}$$

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 $\epsilon \circ F$ gives a natural transformation $FGF \to F$. (These are slightly different forms of composition!)

1.6.4 Lemma *The composite natural transformation $F \to F$ given by $(\epsilon \circ F) \circ (F \circ \eta)$ is the identity. Similarly, the composite natural transformation $G \to GFG \to G$ given by $(G \circ \epsilon) \circ (\eta \circ G)$ 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 (1.6.3.1) $\psi = G(\phi) \circ \eta(c)$; the second step sends ψ to $\epsilon(d) \circ F(\psi)$, by (1.6.3.2). It follows that

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

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

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

This proves the claim. □

1.6.5 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

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

called the *unit* and *counit*, respectively, such that the composites $(\epsilon \circ F) \circ (F \circ \epsilon) : F \to F$ and $(G \circ \epsilon) \circ (\eta \circ G)$ 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 (1.6.3.1) and (1.6.3.2) 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.

1.6.6 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 (1.6.3.1). 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. □

1.6.7 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 1.6.6.

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).

1.6.8 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.

2. Number Systems

2.1. The natural numbers

Peano structures

2.1.1 Definition (Peano) A triple $(\mathbb{P}, 0, s)$ is called a *Peano structure*, if the following axioms hold true:

(P1) 0 is an element of $\mathbb{P}$.

(P2) $s : \mathbb{P} \to \mathbb{P}$ is a mapping.

(P3) 0 is not in the image of s.

(P4) s is injective.

(P5) (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 (P3), 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 $\mathbb{P}^* := \{n \in \mathbb{P} \mid n \neq 0\}$ of all non-zero elements, in signs $s(\mathbb{P}) = \mathbb{P}^*$.*

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 (P5), 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:

(P6) $f(0) = 0'$,

(P7) $f \circ s = s' \circ f$.

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

2.1.4 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 systems of natural numbers, 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 $g : (\mathbb{P}', 0', s') \to (\mathbb{P}'', 0'', s'')$.

2.1.5 Proposition *The Peano structures as objects together with their morphisms and the composition of morphisms form a category.*

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

2.1.6 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 (P5) 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.

2.1.7 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 $t := s'$. □

2.1.8 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 (P5), 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 (P5) 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 (P3), 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 (P4) 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. □

2.1.9 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 **?**. 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 $\mathfrak{I} \subset \mathcal{P}(I)$ the set of all inductive subsets of I. Now put

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

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*.

2.1.10 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 $\emptyset \in \mathbb{N}$, since $\emptyset$ 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 $\emptyset \in J \cap I$. 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. □

2.1.11 Remark The proposition entails in particular that the construction of $\mathbb{N}$ does not depend on the initial choice of the inductive set I.

2.1.12 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 2.1.10 will entail the claim. Clearly, $\emptyset \in J$, since $\emptyset$ 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. □

2.1.13 Theorem (von Neumann) *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, hence Axioms (P1) and (P2) hold true. Since $n \in s(n)$ for every element $n \in \mathbb{N}$, 0 is not in the image of s. This gives Axiom (P3). 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 (P4). In the first case we are done with this as well, since then $m \subset n$ and $n \subset m$ by Lemma 2.1.12. The Induction Axiom (P5) is an immediate consequence of Proposition 2.1.10. □

Addition of natural numbers

2.1.14 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.

2.1.15 Proposition *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:*

(AMon1) *Addition is* associative *that means*

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

(AMon2) *The element 0 is* neutral *with respect to addition which means that*

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

(AMon3) *Addition is* commutative *that means*

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

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

$$\alpha_{s(m)}(n) = \alpha_m(s(n))\ . \tag{2.1.15.1}$$

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 (2.1.15.1) 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 (2.1.15.1). 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 (2.1.15.1), $\alpha_{s(m)}(n) = \alpha_m(s(n)) = s(\alpha_m(n)) = s(\alpha_n(m)) = \alpha_n(s(m))$, hence commutativity of addition follows by the Induction Axiom.

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

2.1.16 Remark If one is given a triple $(M, +, 0)$ where M is a set, $+ : M \times M \to M$ a map and $0 \in M$ an element such that the above axioms (AMon1) to (AMon3) are fulfilled with $\mathbb{N}$ replaced by M, then one calls M (together with + and 0) an *abelian monoid*.

Multiplication of natural numbers

2.1.17 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.

2.1.18 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* semiring *that means the following axioms hold true:*

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

(SRing2) $\mathbb{N}$ *together with multiplication $\cdot$ and the element 1 is a monoid, id est*

(Mon1) *Multiplication is* associative *that means*

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

(Mon2) *The element 1 is* neutral *with respect to multiplication that means*

$$1 \cdot n = n \cdot 1 = n \quad \textit{for all } n \in \mathbb{N}\ .$$

(SRing3) *Multiplication* distributes *from the left and the right over addition that means*

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

(DistR) $(m+n)\cdot l = m\cdot l + n\cdot l$ *for all* $l,m,n\in\mathbb{N}$.

(SRing4) *Multiplication by* 0 annihilates $\mathbb{N}$ *that is*

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

(SRing5) *Multiplication is* commutative *that is*

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

Proof. By Proposition 2.1.15 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 (SRing4).

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.

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 right neutral,

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

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. □

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

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

2.1.20 Remark If one is given a quintuple $(R, +, \cdot, 0, 1)$ where R is a set, $+ : R \times R \to R$ and $\cdot : R \times R \to R$ are maps and $0, 1 \in R$ are elements such that the above axioms (SRing1) to (SRing4) are fulfilled with $\mathbb{N}$ replaced by R, then one calls R (together with $+$, $\cdot$, 0 and 1) a *semiring*. If in addition Axiom (SRing5) holds true, the semiring is called *commutative*.

2.1.21 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.

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

(2.1.22.1) $s(n) = n + 1 = 1 + n$,

(2.1.22.2) $\alpha_m(n) = m + n = n + m$,

(2.1.22.3) $\mu_m(n) = m \cdot n = n \cdot m$,

(2.1.22.4) $\alpha_m(s(n)) = m + (n + 1) = (m + 1) + n$,

(2.1.22.5) $\mu_m(s(n)) = m + (m \cdot n)$.

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 2.1.22.2 and 2.1.22.3 are consequences of the definitions of $+$ and $\cdot$ and commutativity of these operations.

Equation (2.1.22.4) follows from Equations 2.1.22.2, 2.1.22.1 and 2.1.15.1.

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

2.2. The integers

Construction of $\mathbb{Z}$

2.3. The real numbers

Complete ordered fields

2.3.1 Proposition *For an ordered field $(\mathbb{F}, +, \cdot, \leq)$ the following properties are equivalent:*

(i) *The* nested interval property *holds true in $\mathbb{F}$ that means*

(ii) *Every Cauchy sequence in $\mathbb{F}$ converges.*

(iii) *Every subset $X \subset \mathbb{F}$ which is bounded above has a supremum.*

(iv) *No Dedekind cut in $\mathbb{F}$ has a gap.*

Part II.

Fundamentals of Algebra

10. Group Theory

11. Rings and Modules

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 $n \geq 3$. 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.

11.1. Rings and their ideals

Definition of Rings

Even though we shall mostly just work with commutative rings in this book, we will introduce the general notion of rings which are allowed to be non-commutative.

11.1.1 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 0 is an *abelian group* under addition which means the following properties hold true:

- (Grp1) Addition is *associative* that is $(x + y) + z = x + (y + z)$ for all $x, y, z \in R$.
- (Grp2) The element 0 is a *zero* or *neutral element* with respect to addition that means $0 + x = x + 0 = x$ for all $x \in R$.
- (Grp3) For each $x \in R$ there exists an *additive inverse*, i.e. an element $-x \in R$ such that $x + (-x) = (-x) + x = 0$.
- (Grp4) Addition is *commutative* that is $x + y = y + x$ for all $x, y \in R$.

(Ring2) R together with 1 is a monoid under multiplication, i.e. the following axioms are satisfied:

- (Mon1) Multiplication is *associative* that is $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in R$.
- (Mon2) The element 1 is an *identity element* for multiplication that means $1 \cdot x = x \cdot 1 = x$ for all $x \in R$.

(Ring3) Multiplication *distributes* from the left and the right over addition which means that
$x \cdot (y + z) = x \cdot y + x \cdot z$ and $(x + y) \cdot z = x \cdot z + y \cdot z$ for all $x, y, z \in R$.

In addition, if the following axiom holds, the ring R is called *commutative*:

(Ring4) Multiplication is *commutative* that is $x \cdot y = y \cdot x$ for all $x, y \in R$.

We shall typically write xy for $x \cdot y$.

If R is a ring, an *invertible element* or *unit* is an element $x \in R$, such that there exists a $x^{-1} \in R$, called (*multiplicative*) *inverse of* x, with

$$x \cdot (x^{-1}) = 1 \quad \text{and} \quad (x^{-1}) \cdot x = 1 \ .$$

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 $x, y \in S$ the elements $x + y$, $-x$ and xy 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 $\cdot$ for which all above conditions (namely Axioms (Ring1), (Mon1) and (Ring3)) are satisfied besides the unitality requirement (Mon2). A pseudo-ring R is called *commutative* if Axiom (Ring4) is satisfied. 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.

If R is a pseudo-ring, the *center* of R is defined as the set of $x \in R$ commuting with all ring elements, i.e. as the set

$$\mathrm{Z}(R) := \big\{ x \in R \mid xy = yx \text{ for all } y \in R \big\} \ .$$

11.1.2 Proposition *Let R be a pseudo-ring. Then*

(i) $0 \cdot x = x \cdot 0 = 0$ *for all* $x \in R$.

(ii) $(-x)y = x(-y) = -(xy)$ *for all* $x, y \in R$.

(iii) *A multiplicative identity element in R is uniquely determined.*

(iv) *Assume that R possesses an identity element. Then the inverse for an invertible $x \in R$ is uniquely determined.*

Proof. *ad* (i). First compute using associativity, distributivity and that 0 is a zero element:

$$0 \cdot x = (0 + 0) \cdot x = (0 \cdot x) + (0 \cdot x) \ .$$

Adding $-(0 \cdot x)$ on both sides gives $0 = 0 \cdot x$. By an analogous argument we obtain $0 = x \cdot 0$.

ad (ii). By (i) we obtain

$$0 = 0 \cdot y = (x + (-x)) \cdot y = xy + (-x)y \ ,$$

which entails $(-x)y = -(xy)$. Similarly, one shows $x(-y) = -(xy)$.

ad (iii). Assume that 1 and $1'$ are two identity elements in R. Then

$$1 = 1 \cdot 1' = 1' \ .$$

ad (iv). Let R be a ring with identity 1 and $y, y' \in R$ be two inverses of x. Then

$$y = y \cdot 1 = y \cdot (x \cdot y') = (y \cdot x) \cdot y' = 1 \cdot y' = y' \ . \qquad \square$$

11.1.3 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 ring $\mathbb{Z}_2 = \{0, 1\}$ which is commutative as well. 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 main 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.

11.1.4 Example The center $\mathrm{Z}(R)$ of a pseudo-ring R is a sub-pseudo-ring of R by Proposition 11.1.2. It is commutative by definition. Moreover, if R possesses an identity element, then $\mathrm{Z}(R)$ is even a subring.

11.1.5 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 $x \in R$. 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

$$R^{(X)} := \{f \in R^X \mid f(x) \neq 0 \text{ for at most finitely many } x \in X\},$$

since $0_X \in R^{(X)}$, since the sum and the product of two elements $f, g \in R^{(X)}$ lie again in $R^{(X)}$, and since $R^{(X)}$ 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 function 1_X is continuous, and since the sum and product of two real-valued functions on X is 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.

11.1.6 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_k for every $k \in \mathbb{N}$ its k-th component, that means let $a = (a_k)_{k\in\mathbb{N}}$. Using this agreement, the sum and product of two elements $a, b \in R[x]$ are defined by

$$\begin{aligned}(a+b)_k &:= (a_k + b_k) \text{ for all } k \in \mathbb{N}, \text{ and}\\ (a \cdot b)_k &:= \sum_{i=0}^{k} a_{k-i} b_i \text{ for all } k \in \mathbb{N}\,.\end{aligned}$$

By Example 11.1.5 (b) , $(R[x], +, 0_{\mathbb{N}})$ is an abelian group with zero element $0_{\mathbb{N}} : \mathbb{N} \to R$, $n \mapsto 0$.

11.1.7 Example For any ring R, we can consider the polynomial ring $R[x_1, \ldots, x_n]$ which consists of the polynomials in n variables with coefficients in R. This can be defined inductively as $(R[x_1, \ldots, x_{n-1}])[x_n]$, where the procedure of adjoining a single variable comes from the previous **??**.

We shall see a more general form of this procedure in Example 11.1.15.

The category of rings

The class of rings forms a category. Its morphisms are called ring homomorphisms.

11.1.8 Definition A *morphism of pseudo-rings* is a map $f : R \to S$ between pseudo-rings R and S that respects addition and multiplication. That is

(Ring5) $f(x + y) = f(y) + f(y)$ for all $x, y \in R$.

(Ring6) $f(xy) = f(x)f(y)$ 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*. In other words, $f : R \to S$ is a *ring homomorphism* if it satisfies in addition to Axioms (Ring5) and (Ring6) the axiom

(Ring7) $f(1_R) = 1_S$, where 1_R and 1_S are the respective identity elements.

The composition of two ring homorphisms is obviously again a ring homomorphism. Moreover, the identity map $\mathrm{id}_R : R \to R$ on a ring R is a ring homomorphism, too. There is thus a category Ring whose objects are rings and whose morphisms are ring homomorphisms.

11.1.9 Proposition *Let $f : R \to S$ be a morphism of pseudo-rings. Then*

(i) *$f(0_R) = 0_S$, where 0_R and 0_S are the respective zero elements.*

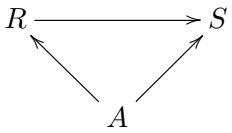
11.1.10 The philosophy of Grothendieck, as expounded in his EGA **?**, is that one should always do things in a relative context. This means that instead of working with objects, one should work with *morphisms* of objects. Motivated by this, we introduce:

11.1.11 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



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.

11.1.12 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.

11.1.13 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."

11.1.14 Example The field of complex numbers $\mathbb{C}$ is an $\mathbb{R}$-algebra.

Here is an example that generalizes the case of the polynomial ring.

11.1.15 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 $G = \mathbb{Z}_{\geq 0}$ 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 40.2.5 below.

11.1.16 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.

[1] That is, there is a commutative multiplication on G with an identity element, but not necessarily with inverses.

11.1.17 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.

11.1.18 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 **Rings** $\to$ **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 11.1.18, the forgetful functor **Rings** $\to$ **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.

11.1.19 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.

11.1.20 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.

As a result, we find:

11.1.21 Proposition *The category* **Rings** *is complete.*

As we said, we will not yet show that **Rings** 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 **Rings** $\to$ **Sets** commutes with *filtered* colimits (though not with general colimits).

11.1.22 Example (Filtered colimits) Let I be a filtering category, $F : I \to \mathbf{Rings}$ 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.

Ideals

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.

11.1.23 Definition Let R be a ring. An **ideal** in R is a subset $I \subset R$ that satisfies the following.

1. $0 \in I$.
2. If $x, y \in I$, then $x + y \in I$.
3. If $x \in I$ and $y \in R$, then $xy \in I$.

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 $\sum r_i x_i$, where $r_i \in R$. 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 $x \in R$.

11.1.24 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 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).

11.1.25 Example We saw in examples 11.1.5 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.

11.1.26 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.

11.1.27 Definition The sum $I + J$ of two ideals $I, J \subset R$ is defined as the set of sums

$$\{x + y : x \in I, y \in J\}.$$

11.1.28 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:

11.1.29 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.

11.1.30 Example In $\mathbb{Z}$, we have the following for any m, n.

1. $(m) + (n) = (\gcd\{m, n\})$,
2. $(m)(n) = (mn)$,
3. $(m) \cap (n) = (\mathrm{lcm}\{m, n\})$.

11.1.31 Proposition *For ideals $I, J, K \subset R$, we have the following.*

1. *Distributivity: $I(J + K) = IJ + IK$.*
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. □

11.1.32 Remark There is a *contravariant* functor **Rings** $\to$ **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.

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 $(a + I)(b + I) = ab + I$.

11.1.33 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.

11.1.34 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 $\mathbb{Z}[X]/(X - X^2)$.)

11.1.35 Remark Show that the functor assigning to each ring the set of elements annihilated by 2 is corepresentable.

11.1.36 Remark If $I \subset J \subset R$, then J/I is an ideal of R/I, and there is a canonical isomorphism

$$(R/I)/(J/I) \simeq R/J.$$

Zerodivisors

Let R be a commutative ring.

11.1.37 Definition If $r \in R$, then r is called a **zerodivisor** if there is $s \in R, s \neq 0$ with $sr = 0$. Otherwise r is called a **nonzerodivisor.**

As an example, we prove a basic result on the zerodivisors in a polynomial ring.

11.1.38 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 $r \in R \setminus \{0\}$ such that $f \cdot r = 0$.*

So all the coefficients are zerodivisors.

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. □

11.1.39 Remark The product of two nonzerodivisors is a nonzerodivisor, and the product of two zerodivisors is a zerodivisor. It is, however, not necessarily true that the *sum* of two zerodivisors is a zerodivisor.

11.2. 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).

11.2.1 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.

11.2.2 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.

11.2.3 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 $y_1, \ldots, y_n$. 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.

11.2.4 Example Consider the polynomial ring $\mathbb{C}[x_1, \ldots, x_n]$. Recall that before we thought of the ring $\mathbb{C}[t]$ as a ring of meromorphic functions. Similarly each element of the polynomial ring $\mathbb{C}[x_1, \ldots, x_n]$ 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 $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$; consider the collection of all functions $f \in \mathbb{C}[x_1, \ldots, x_n]$ 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.

11.2.5 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.

11.2.6 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.

11.2.7 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] Such a subset is called an algebraic variety.

11.3. Modules over a ring

We will now establish some basic terminology about modules. Throughout this section, R denotes always a ring.

Definitions

11.3.1 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)° Scalar multiplication is associative, i.e. $n(xy) = (nx)y$ for all $x, y \in R$ and $n \in N$.

(Mod 2)° The unit $1 \in R$ acts as identity that means $n \cdot 1 = n$ for all $n \in N$.

(Mod 3)° 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.

11.3.2 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.

11.3.3 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.

11.3.4 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:

$$\begin{array}{ccc} R \times M_0 & \longrightarrow & M_0 \\ \downarrow & & \downarrow \\ R \times M & \longrightarrow & M \end{array} \quad \text{respectively} \quad \begin{array}{ccc} N \times R & \longrightarrow & N \\ \downarrow & & \downarrow \\ M \times R & \longrightarrow & M \end{array},$$

depending on whether M is a left or right R-module. Here the horizontal maps are multiplication by scalars.

11.3.5 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.

11.3.6 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.

11.3.7 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.

11.3.8 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**.

11.3.9 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).

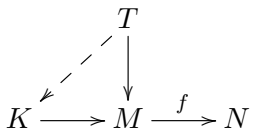
In the case of the category of modules, the zero object is clearly the zero module, and the zero morphism $M \to N$ sends $m \mapsto 0$ for each $m \in M$.

11.3.10 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 $N/\operatorname{Im}(f)$.

11.3.11 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)



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 $X \xrightarrow{0} Y$.

11.3.12 Remark What is the universal property of the cokernel?

11.3.13 Remark On the category of modules, the functor assigning to each module M its underlying set is corepresentable (cf. remark 11.1.18). 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.

11.3.14 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.

11.3.15 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.

11.3.16 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 11.3.16 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 1.4.44 for the definition). To see this, it suffices to show that (by theorem 1.4.29 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:

11.3.17 Proposition *If R is a ring, the category of R-modules is complete and cocomplete.*

11.3.18 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 11.6.3), which implies it preserves limits (by proposition 1.6.8).

Exactness

Finally, we introduce the notion of *exactness*.

11.3.19 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 11.3.19. 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.

11.3.20 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'.

11.3.21 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:

11.3.22 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).

11.3.23 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.

11.3.24 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 $g \circ f = 1_A$.

11.3.25 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.

11.3.26 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.

11.3.27 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.

11.3.28 Theorem *Suppose given a commutative diagram*

$$\begin{array}{ccccccccc} A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & D & \longrightarrow & E \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & D' & \longrightarrow & E' \end{array}$$

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

$$\begin{array}{ccccccccc} A & \xrightarrow{k} & B & \xrightarrow{l} & C & \xrightarrow{m} & D & \xrightarrow{n} & E \\ \downarrow{\scriptstyle a} & & \downarrow{\scriptstyle b} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle e} \\ F & \xrightarrow[p]{} & G & \xrightarrow[q]{} & H & \xrightarrow[r]{} & I & \xrightarrow[s]{} & J \end{array}$$

□

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.

11.4. 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.

11.4.1 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 $y \in I$.

11.4.2 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.

11.4.3 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.

11.4.4 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).

11.4.5 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.

11.4.6 Remark Find the maximal ideals in $\mathbb{C}[t]$.

11.4.7 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.

11.4.8 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. □

11.4.9 Corollary *Let R be a nonzero commutative ring. Then R has a maximal ideal.*

Proof. Apply the lemma to the zero ideal. □

11.4.10 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 11.4.8 finishes the argument. □

Fields and integral domains

Recall:

11.4.11 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.

11.4.12 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.

11.4.13 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. □

There is a similar characterization of prime ideals.

11.4.14 Definition A commutative ring R is an **integral domain** if for all $x, y \in R$, $x \neq 0$ and $y \neq 0$ imply $xy \neq 0$.

11.4.15 Proposition *An ideal $I \subset R$ is prime iff R/I is a domain.*

11.4.16 Remark Prove proposition 11.4.15.

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.

11.4.17 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 $\tilde{\phi}(a_0+\cdots+a_nx^n) = (a_0+P)+\cdots+(a_n+P)x^n$. This is clearly a homomorphism because $\phi : R \to R/P$ is, and its kernel consists of those polynomials $a_0 + \cdots + a_nx^n$ 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 $P[x]$.

11.4.18 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!

11.4.19 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 zerodivisors on a module M is contained in associated prime.

11.4.20 Theorem (Prime avoidance) *Let $I_1, \ldots, I_n \subset R$ 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 $A \subset I_1 \cup \cdots \cup I_n$, 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 $n \geq 3$. We may assume that for each j, $A \not\subset I_1 \cup \cdots \cup \hat{I}_j \cup \cdots I_n$.[3] Fix an element $a_j \in A \smallsetminus (I_1 \cup \cdots \cup \hat{I}_j \cup \cdots I_n)$. Then this a_j must be contained in I_j since $A \subset \bigcup I_j$. Since $n \geq 3$, 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 $j \geq 2$. 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:

11.4.21 Theorem (Chinese remainder theorem) *Let $I_1, \ldots I_n$ be ideals in a ring R which satisfy $I_i + I_j = R$ for $i \neq j$. Then we have $I_1 \cap \cdots \cap I_n = I_1 \ldots I_n$ and the morphism of rings*

$$R \to \bigoplus R/I_i$$

is an epimorphism with kernel $I_1 \cap \cdots \cap I_n$.

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 $n = 2$.

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 $1 \leq i \leq n-1$, 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 $\prod_{i=1}^{n} I_i$.

[3] The hat means omit I_j.

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 $R/I_1 \cdots I_n \simeq \bigoplus_{i=1}^n R/I_i$.

11.5. Some special classes of domains

Principal ideal domains

11.5.1 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.

11.5.2 Example $\mathbb{Z}$ is a PID. The only nontrivial fact to check here is that:

11.5.3 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 $0 \leq r < n$, 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:

11.5.4 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.

11.5.5 Definition An element of a domain R is **irreducible** if it cannot be written as the product of two non-unit elements of R.

11.5.6 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.

11.5.7 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.

11.5.8 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 11.5.13 and remark 11.5.14, we saw that $\mathbb{Z}[i]$ and $F[t]$ are PIDs, so they also have unique factorization.

11.5.9 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 $k \geq n$. 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 $m \geq n$. 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.

11.5.10 Definition An integral domain R is a **euclidean domain** if there is a function $|\cdot| : R \to Z_{\geq 0}$ (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.

11.5.11 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 $0 \leq r < |b|$, 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.*

11.5.12 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:* □

11.5.13 Remark Prove that $\mathbb{Z}[i]$ is principal. (Define the norm as $N(a + ib) = a^2 + b^2$.)

11.5.14 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 $d : R \to \mathbb{Z}_{\geq 0}$ such that

1. $d(a) = 0$ iff $a = 0$.
2. $d(ab) \geq d(a)$ 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:

11.5.15 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 11.5.9), 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 $d(y) \leq d(b)$. 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 $d(y) < d(b)$.

11.5.16 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 11.5.15 actually goes back to Hasse (and these norms are sometimes called "Dedekind-Hasse norms").

11.6. Basic properties of modules

Free modules

We now describe a simple way of constructing modules over a ring, and an important class of modules.

11.6.1 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.

11.6.2 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:

11.6.3 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 11.4.4) and corollary 11.4.13. 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.

11.6.4 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.

11.6.5 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 11.3.13).

The next claim is that free modules form a reasonably large class of the category of R-modules.

11.6.6 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.

11.6.7 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.

11.6.8 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

$$\begin{array}{ccc} M & \xrightarrow{\phi_M} & M \\ \downarrow{\scriptstyle f} & & \downarrow \\ N & \xrightarrow{\phi_N} & N. \end{array}$$

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

$$\begin{array}{ccc} F & \xrightarrow{\phi_F} & F \\ \downarrow & & \downarrow \\ M & \xrightarrow{\phi_M} & M. \end{array}$$

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.

11.6.9 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.

11.6.10 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 41, this does not happen over *noetherian* rings.

11.6.11 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.

11.6.12 Remark Show that a quotient of a finitely generated module is finitely generated.

11.6.13 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.

11.6.14 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.

11.6.15 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

$$\begin{array}{ccccccc} & & F' & \longrightarrow & F & \longrightarrow & M & \longrightarrow & 0 \\ & & & & \vdots & & \downarrow & & \\ 0 & \longrightarrow & K & \longrightarrow & R^m & \longrightarrow & M & \longrightarrow & 0 \end{array}$$

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

$$\begin{array}{ccccccc} & & F' & \longrightarrow & F & \longrightarrow & M & \longrightarrow & 0\,, \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow & & \\ 0 & \longrightarrow & K & \longrightarrow & R^m & \longrightarrow & M & \longrightarrow & 0 \end{array}$$

to which we can apply the snake lemma. There is an exact sequence

$$0 \to \operatorname{coker}(f) \to \operatorname{coker}(g) \to 0,$$

which gives an isomorphism $\operatorname{coker}(f) \simeq \operatorname{coker}(g)$. However, $\operatorname{coker}(g)$ is finitely generated, as a quotient of R^m. Thus $\operatorname{coker}(f)$ is too. Since we have an exact sequence

$$0 \to \operatorname{Im}(f) \to K \to \operatorname{coker}(f) \to 0,$$

and $\operatorname{Im}(f)$ is finitely generated (as the image of a finitely generated object, F'), we find by proposition 11.6.10 that $\operatorname{coker}(f)$ is finitely generated. □

11.6.16 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 41 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

$$\begin{array}{ccccccccc} & & F_1' & & & & F_1'' & & \\ & & \downarrow & & & & \downarrow & & \\ & & F_0' & & & & F_0'' & & \\ & & \downarrow & & & & \downarrow & & \\ 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \end{array}$$

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. □

11.6.17 Corollary *The (finite) direct sum of finitely presented modules is finitely presented.*

Proof. Immediate from proposition 11.6.16 □

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.

11.6.18 Definition M is **simple** if $M \neq 0$ and M has no nontrivial submodules.

11.6.19 Remark A torsion-free abelian group is never a simple $\mathbb{Z}$-module.

11.6.20 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. □

11.6.21 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.

11.6.22 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.

11.6.23 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 $M_1 \subset M_2 \subset \ldots$ of submodules stabilizes, and it is *artinian* if every descending chain stabilizes.

11.6.24 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 41.1.7 and proposition 41.4.5 of chapter 41).

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. □

11.6.25 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:

11.6.26 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. □

11.6.27 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.*

11.6.28 Definition This number is called the **length** of M and is denoted $\ell(M)$.

Proof of theorem 11.6.27. 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 $l(M) \geq 2$. 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

$$0 = N_0 \cap K \subset M_1 \subset K \subset \dots N_{l(M)-1} \cap K = K \subset N_{l(M)-1}.$$

These filtrations have simple or zero quotients by lemma 11.6.26, 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. □

11.6.29 Remark Prove that the successive quotients M_i/M_{i-1} are also determined (up to permutation).

12. Fields and Extensions

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.

12.1. Fields

Recall once again that:

12.1.1 Definition A *field* is an integral domain where every non-zero element is invertible. Alternatively, it is a set $\Bbbk$, endowed with binary operations of addition $+$ and multiplication $\cdot$, which satisfy the usual axioms of associativity of $+$ and $\cdot$, commutativity of $+$ and $\cdot$, 0 and 1 being the neutral elements with respect to $+$ and $\cdot$, respectively, the requirement $1 \neq 0$, distributivity of $\cdot$ over $+$, existence of additive inverses, and existence of multiplicative inverses for non-zero elements.

A *subfield* is a subset closed under these operations: equivalently, it is a subring that is itself a field.

For a field $\Bbbk$, we write $\Bbbk^*$ for the subset $\Bbbk \setminus \{0\}$. This generalizes the usual notation **??** R^* that 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 11.4.13) that if R is a ring and $I \subset R$ an ideal, then R/I is a field precisely when I is maximal.

12.1.2 Example One of the most familiar examples of a field is the rational numbers $\mathbb{Q}$.

12.1.3 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.

12.1.4 Example (Quotients of the polynomial ring) In a principal ideal domain, every prime ideal is principal. Now, by 11.5.12, 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 $\mathbb{R}[x]/(x^2+1)$.

12.1.5 Remark What is $\mathbb{C}[x]/(x^2+1)$?

12.1.6 Example (Quotient fields) Recall from remark 11.4.18 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 $\mathbb{F}_p$!).

12.1.7 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.

12.1.8 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:

12.1.9 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. □

12.1.10 Corollary *Every exact sequence of modules over a field splits.*

Proof. This follows from **??** and proposition 12.1.9, 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 12.1.10 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 11.1.16. 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 11.4.15. Hence the kernel of f is either (0) or (p) for some prime number p, see 11.4.2.

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.

12.1.11 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.

12.1.12 Example The characteristic of $\mathbb{Z}/p$ is p, and that of $\mathbb{Q}$ is 0. This is obvious from the definitions.

12.2. 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.

12.2.1 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. □

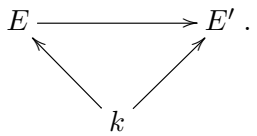
12.2.2 Definition If F is a field contained in a field G, then G is said to be a **field extension** of F. We shall write G/F to indicate that G is an extension of F.

So if F, F' are fields, and $F \to F'$ is any ring-homomorphism, we see by proposition 12.2.1 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



12.2.3 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.

12.2.4 Example Let k be a field, and $P \in k[x]$ an irreducible polynomial. We have seen that $k[x]/(P)$ is a field (12.1.7). Since it is also a k-algebra in the obvious way, it is an extension of k.

12.2.5 Example If X is a Riemann surface, then the field of meromorphic functions $\mathbb{C}(X)$ (see example 12.1.8) 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.

12.2.6 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 $\mathbb{R}$.

12.2.7 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.

12.2.8 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).

12.2.9 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.

12.2.10 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 12.2.8, in the next two examples.

12.2.11 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 **??**.)

12.2.12 Example (Degree of a simple algebraic extension) Consider a monogenic field extension E/k of the form in 12.1.7, 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

$$P = t^n + a_1 t^{n-1} + \cdots + a_0. \tag{12.2.12.1}$$

It is then easy to see that the images of $1, t, \ldots, t^{n-1}$ 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 $S = \{1, t, \ldots, t^{n-1}\}$ (or more properly their images) spans $k[t]/(P)$ as a vector space. Indeed, we have by (12.2.12.1) 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 $\{1, t, \ldots, t^n\}$—by what was just shown, thus in the span of S. Working upward inductively, we find that the image of t^M for $M \geq n$ 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.*

12.2.13 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 $\{\alpha_i\beta_j, 1 \leq i \leq n, 1 \leq j \leq m\}$ 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:

12.2.14 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.

12.2.15 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 12.2.8, 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.

12.2.16 Example $\mathbb{C}$ is algebraic over $\mathbb{R}$.

12.2.17 Example Let X be a compact Riemann surface, and $f \in \mathbb{C}(X) - \mathbb{C}$ any non-constant meromorphic function on X (see example 12.1.8). 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.

12.2.18 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 $\{1, \alpha, \ldots, \alpha^n\}$ 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 12.2.11 and example 12.2.12. 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 12.2.11 and example 12.2.12): a monogenic extension is finite if and only if it is algebraic. We shall use this observation in the next result.

12.2.19 Corollary *Let k be a field, and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be elements of some extension field such that each α_i is finite over k. Then the extension $k(\alpha_1, \ldots, \alpha_n)/k$ is finite. That is, a finitely generated algebraic extension is finite.*

Proof. Indeed, each $k(\alpha_1, \ldots, \alpha_{i+1})/k(\alpha_1, \ldots, \alpha_i)$ 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).

12.2.20 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 12.2.19. It follows that $k(\alpha + \beta) \subset k(\alpha, \beta)$ is a finite extension, which implies that $\alpha + \beta$ is algebraic by proposition 12.2.18. □

Many nice properties of field extensions, like those of rings, will have the property that they will be preserved by towers and composita.

12.2.21 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 12.2.19 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.

12.2.22 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.

12.2.23 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 12.2.12). □

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.

12.2.24 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.

12.2.25 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 $n \geq k$ (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 $n \geq 1$, 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)

$$P = c(x-\alpha_1)\ldots(x-\alpha_n),$$

where c is the constant term and the $\{\alpha_i\} \subset k$ are the roots of P. In particular:

12.2.26 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.

12.2.27 Definition If k is any field, we say that two polynomials in $k[x]$ are **relatively prime** if they generate the unit ideal in $k[x]$.

12.2.28 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. $\square$

12.3. 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.

12.3.1 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 12.2.28, 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.

12.3.2 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

12.3.3 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.

12.3.4 Definition If $K = F(\alpha)$ for some α with minimal polynomial $q(x) \in F[x]$, then by 12.4.3, $q(x) = r(x^{p^d})$, where $p = \operatorname{char} F$ (or 1 if $\operatorname{char} F = 0$) and r is separable; in this case we also denote $\deg_s(K/F) = \deg(r), \deg_i(K/F) = p^d$.

12.4. 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.

12.4.1 Definition $\operatorname{Aut}(K/F)$ 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.

12.4.2 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 12.4.13 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

12.4.3 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* $d \geq 0$ *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 12.3.2.

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. □

12.4.4 Corollary *In the statement of 12.4.3,* 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 $x^{p^d} - \beta$. □

12.4.5 Lemma *The correspondence which to each* $g \in \operatorname{Emb}(K/F)$ *assigns the* n*-tuple* $(g(\alpha_1), \ldots, g(\alpha_n))$ *of elements of* M *is a bijection from* $\operatorname{Emb}(K/F)$ *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 $g_1 \in \operatorname{Emb}(K_1/F)$ 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. □

12.4.6 Corollary $|\operatorname{Emb}(K/F)| = \prod_{i=1}^{n} \deg_s(q_i)$.

Proof. This follows immediately by induction from 12.4.5 by 12.4.4. □

12.4.7 Lemma *For any* $f \in \operatorname{Emb}(K/F)$*, the map* $\operatorname{Aut}(K/F) \to \operatorname{Emb}(K/F)$ *given by* $\sigma \mapsto f \circ \sigma$ *is injective.*

Proof. This is immediate from the injectivity of f. □

12.4.8 Corollary $\operatorname{Aut}(K/F)$ *is finite.*

Proof. By 12.4.7, $\operatorname{Aut}(K/F)$ injects into $\operatorname{Emb}(K/F)$, which by 12.4.6 is finite. □

12.4.9 Proposition *The inequality*

$$|\operatorname{Aut}(K/F)| \leq |\operatorname{Emb}(K/F)|$$

is an equality if and only if the q_i *all split in* K*.*

Proof. The inequality follows from 12.4.7 and from 12.4.8. Since both sets are finite, equality holds if and only if the injection of 12.4.7 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 12.4.5, and extract an embedding $g\colon K \to M$ with $g(\alpha_i) = \beta_i$. Since the correspondence $f \mapsto f \circ \sigma$ ($\sigma \in \mathrm{Aut}(K/F)$) 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 12.4.5. 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 $f^{-1} \circ g \in \mathrm{Aut}(K/F)$. Thus $g = f \circ (f^{-1} \circ g)$ shows that the map of 12.4.7 is surjective. □

12.4.10 Corollary *Define*

$$D(K/F) = \prod_{i=1}^{n} \deg_s(K_i/K_{i-1}).$$

Then the chain of equalities and inequalities

$$|\mathrm{Aut}(K/F)| \leq |\mathrm{Emb}(K/F)| = D(K/F) \leq \deg(K/F)$$

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 12.4.9; the interior equality is just 12.4.6; the latter inequality is obvious from the multiplicativity of the degrees of field extensions; and the deduction for equality follows from the definition of $\deg_s$. □

12.4.11 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 12.4.10 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. □

12.4.12 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 12.4.10 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 $|\mathrm{Emb}(K/F)| = 1$ by 12.4.10, 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. □

12.4.13 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 12.4.3, 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 12.4.12, 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:

$$\deg(K/K') = \prod_{i=1}^{n} p^{d_i} = \prod_{i=1}^{n} \frac{\deg(K_i/K_{i-1})}{\deg_s(K_i/K_{i-1})} = \frac{\deg(K/F)}{D(K/F)}.$$

□

12.4.14 Theorem *The following inequality holds:*

$$|\operatorname{Aut}(K/F)| \leq |\operatorname{Emb}(K/F)| = \deg_s(K/F) \leq \deg(K/F).$$

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 12.4.10 and 12.4.13. The deductions combine 12.4.11 and 12.4.12. □

Definitions

Throughout, we will denote as before K/F a finite field extension, and $G = \operatorname{Aut}(K/F)$, H a subgroup of G. L/F is a subextension of K/F.

12.4.15 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.

12.4.16 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

12.4.17 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. □

12.4.18 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 12.4.17. □

12.4.19 Lemma *If K/F is Galois, so is K/L, and* $\mathrm{Gal}(K/L) = G_L$..

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. $\mathrm{Gal}(K/L) = \mathrm{Aut}(K/L)$ consists of those automorphisms σ of K which fix L; since $F \subset L$ we have *a fortiori* that σ fixes F, hence $\mathrm{Gal}(K/L) \subset G$ and consists of the subgroup which fixes L; i.e. G_L. □

12.4.20 Corollary *If K/F and L/F are Galois, then the action of G on elements of L defines a surjection of G onto* $\mathrm{Gal}(L/F)$. *Thus G_L is normal in G and* $\mathrm{Gal}(L/F) \cong G/G_L$. *Conversely, if $N \subset G$ is normal, then K^N/F is Galois.*

Proof. L/F is splitting, so by 12.4.17 the elements of G act as endomorphisms (hence automorphisms) of L/F, and the kernel of this action is G_L. By 12.4.19, we have $G_L = \mathrm{Gal}(K/L)$, so $|G_L| = |\mathrm{Gal}(K/L)| = [K : L] = [K : F]/[L : F]$, or rearranging and using that K/F is Galois, we get $|G|/|G_L| = [L : F] = |\mathrm{Gal}(L/F)|$. Thus the map $G \to \mathrm{Gal}(L/F)$ is surjective and thus the induced map $G/G_L \to \mathrm{Gal}(L/F)$ 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. □

12.4.21 Proposition *If K/F is Galois and $H = G_L$, then $K^H = L$.*

Proof. By 12.4.19, K/L and K/K^H are both Galois. By definition, $\mathrm{Gal}(K/L) = G_L = H$; since H fixes K^H we certainly have $H < \mathrm{Gal}(K/K^H)$, but since $L \subset K^H$ we have *a fortiori* that $\mathrm{Gal}(K/K^H) < \mathrm{Gal}(K/L) = H$, so $\mathrm{Gal}(K/K^H) = H$ as well. It follows from 12.4.14 that $\deg(K/L) = |H| = \deg(K/K^H)$, so that $K^H = L$. □

12.4.22 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. □

12.4.23 Corollary *If K is a finite field, then* $\mathrm{Gal}(K/F)$ *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 12.4.22, 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 12.4.21 we have $\mathrm{Gal}(K/F) = \mathrm{Gal}(K/\mathbb{F}_p)_F$, and every subgroup of a cyclic group is cyclic. □

12.4.24 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 $F(\alpha) = K$. □

12.4.25 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 $K = F(\alpha_1 + t_1\alpha_2)$. □

12.4.26 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 12.4.21 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 12.4.24 we may assume F is infinite. The result now follows from 12.4.25. □

12.4.27 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. □

12.4.28 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 12.4.21, 12.4.27, and 12.4.20. □

12.5. Transcendental Extensions

There is a distinguished type of transcendental extension: those that are "purely transcendental."

12.5.1 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 $b_1, \ldots, b_n \in B$.

12.5.2 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.

12.5.3 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 $trdeg(E'/E)$.

12.5.4 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.

12.5.5 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 $\#(B') \leq \#(B)$. 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 $\beta \in B \setminus B^*$. We know β is algebraic over $E(B')$ which is algebraic over $E(B^*)$. Therefor β is algebraic over $E(B^*)$, a contradiction. So $\#(B) \leq \sum_{\alpha \in B'} \#(B_\alpha)$. Now if B' is finite, then so is B so we can assume B' is infinite; this means

$$\#(B) \leq \sum_{\alpha \in B'} \#(B_\alpha) = \#(\coprod B_\alpha) \leq \#(B' \times \mathbb{Z}) = \#(B') \tag{12.5.5.1}$$

with the inequality $\#(\coprod B_\alpha) \leq \#(B' \times \mathbb{Z})$ 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 $m \leq n$. 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, $m = n$. □

12.5.6 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.

12.5.7 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.

12.5.8 Remark If we have a tower of fields $G/F/E$, then $trdeg(G/E) = trdeg(F/E) + trdeg(G/F)$.

12.5.9 Example Let X be a compact Riemann surface. Then the function field $\mathbb{C}(X)$ (see example 12.1.8) 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.

12.5.10 Definition The compositum of K and L written KL is $k(K \cup L) = L(K) = K(L)$.

12.5.11 Definition K and L are said to be linearly disjoint over k if the following map is injective:

$$\theta : K \otimes_k L \to KL \tag{12.5.11.1}$$

defined by $x \otimes y \mapsto xy$.

13. 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.

13.1. 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:

13.1.1 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.

13.1.2 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.

13.1.3 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 $S^{-1}M \overset{S^{-1}f}{\to} S^{-1}N$.

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.

13.1.4 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.

13.1.5 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.

13.1.6 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.

13.1.7 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.

13.1.8 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 $S^{-1}R$.

13.1.9 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.

13.1.10 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.*

13.1.11 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 11.4.8, 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.

13.1.12 Example Any field is a local ring because the unique maximal ideal is (0).

13.1.13 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.

13.1.14 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. □

13.1.15 Remark (exercise) Any local ring is of the form $R_\mathfrak{p}$ for some ring R and for some prime ideal $\mathfrak{p} \subset R$.

13.1.16 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.

13.1.17 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.

13.1.18 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. □

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.

13.1.19 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:

13.1.20 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:

13.1.21 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.

13.1.22 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.

13.1.23 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.

13.1.24 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 13.1.22) to M/N. □

We shall prove more generally:

13.1.25 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 $\{x_1, \ldots, x_n\} \subset M$. 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:

13.1.26 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 $(y) = I$. □

13.1.27 Remark (exercise) 13.1.26 fails if the ideal is not finitely generated.

13.1.28 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. $Q(f)f = 1_M$.

13.1.29 Remark (exercise) Give a counterexample to the conclusion of Nakayama's lemma when the module is not finitely generated.

13.1.30 Remark (exercise) Let M be a finitely generated module over the ring R. Let $\mathfrak{I}$ be the Jacobson radical of R (cf. 11.4.19). If $\mathfrak{I}M = M$, then $M = 0$.

13.1.31 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.

13.1.32 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 $x_1, \ldots, x_n$. Then write $x_1 = a_1x_1 + \cdots + a_nx_n$ where each $a_i \in \mathfrak{m}$. Deduce that x_1 is in the submodule generated by the $x_i, i \geq 2$, 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.

13.1.33 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 $e_1, \ldots, e_r$ be a "basis" for F, and let $f_1, \ldots, f_r$ be the images under ϕ in F'. Then the reductions $\overline{f_1}, \ldots, \overline{f_r}$ 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 $\overline{g_1}, \ldots, \overline{g_s} \in F'/\mathfrak{m}F'$, which we can lift to elements $g_1, \ldots, g_s \in F'$. It is clear that F' has rank $r+s$ since its reduction $F'/\mathfrak{m}F'$ does.

We claim that the set $\{f_1, \ldots, f_r, g_1, \ldots, g_s\}$ 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 $f_1, \ldots, f_r, g_1, \ldots, g_s$ 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.

13.1.34 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 $\overline{x_1}, \ldots, \overline{x_n}$ for $M/\mathfrak{m}M$ (as an $R/\mathfrak{m}$-vector space) and lift this to elements $x_1, \ldots, x_n \in M$. 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 13.1.33). It follows that we can find a splitting $F \to L$, which when composed with $L \to M$ is a splitting of $M \to F$. □

13.1.35 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.)

13.2. 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.

13.2.1 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.

13.2.2 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 11.3.11; 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''. □

[1] Strictly speaking, this may depend on your set-theoretic foundations.

This functor $\hom_R(M, \cdot)$ is not exact in general. Indeed:

13.2.3 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.

13.2.4 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 13.2.2, 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. □

13.2.5 Remark (exercise) Come up with an example where $\hom_R(-, M)$ is not exact.

13.2.6 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.

13.2.7 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

13.2.8 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 11.6.6. 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 $P \simeq \ker(P \to M) \oplus \mathrm{im}(M \to P) \simeq \ker(P \to M) \oplus M$, 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 $M \to N/N'$. □

[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.

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

$$\begin{array}{ccccc} & & P & & \\ & & \downarrow & & \\ M & \longrightarrow & M'' & \longrightarrow & 0 \end{array}$$

where the bottom row is exact. Then, if P is projective, there is a lifting $P \to M$ making commutative the diagram

$$\begin{array}{ccccc} & & P & & \\ & \swarrow & \downarrow & & \\ M & \longrightarrow & M'' & \longrightarrow & 0 \end{array}$$

13.2.9 Corollary *Let M be a module. Then there is a surjection $P \twoheadrightarrow M$, where P is projective.*

Proof. Indeed, we know (11.6.6) that we can always get a surjection from a free module. Since free modules are projective by 13.2.8, we are done. □

13.2.10 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.

13.2.11 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:

13.2.12 Proposition $\Gamma(X, E)$ *is a finitely generated projective $C(X)$-module.*

13.2.13 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 $C(X)$.*

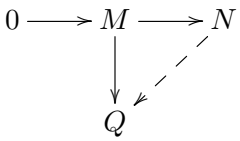
Injective modules

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:

13.2.14 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



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.

13.2.15 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 "$(\tilde{f}, M') \leq (\tilde{f}', M'')$ 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. □

13.2.16 Remark (exercise) theorem 13.2.15 works over any principal ideal domain.

13.2.17 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:

13.2.18 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 13.2.4 and theorem 13.2.15, 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:

13.2.19 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 11.6.6, 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.

13.2.20 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

$$\hom_R(B, R^\vee) = \hom_{\mathbb{Z}}(B, \mathbb{Q}/\mathbb{Z}). \tag{13.2.20.1}$$

In particular, $\hom_R(-, R^\vee)$ is an exact functor because $\mathbb{Q}/\mathbb{Z}$ is an injective abelian group. The proof of eq. (13.2.20.1) 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 13.2, it is also more general. Apparently this method originates with Baer, and was revisited by Cartan & Eilenberg (1999) and by Grothendieck (1957). 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

$$\varinjlim \hom(A, B_\alpha) \to \hom(A, \varinjlim B_\alpha) \tag{13.2.20.2}$$

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. (13.2.20.2) is neither injective or surjective.

13.2.21 Example Consider the category of sets. Let $A = \mathbb{N}$ and $B_n = \{1, \ldots, n\}$ be the inductive system indexed by the natural numbers (where $B_n \to B_m, n \leq m$ 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.

13.2.22 Example Next we give an example where the map fails to be injective. Let $B_n = \mathbb{N}/\{1, 2, \ldots, n\}$, 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 $n \leq m$, 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 $\mathbb{N} \to \mathbb{N}/\{1, 2, \ldots, n\}$ 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 $B_1 \to B_2 \to \ldots$, 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 $c_1, \ldots, c_r \in \varinjlim B_n$. 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 $N \geq m$ 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 $B_1 \to B_2 \to \ldots$, so that it has to factor through a finite step.

Let us generalize this.

13.2.23 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

$$\varinjlim \hom(A, B_\alpha) \to \hom(A, \varinjlim B_\alpha) \tag{13.2.23.1}$$

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:

13.2.24 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] There are, incidentally, categories, such as the category of rings, where a categorical epimorphism may not be a surjection of sets.

13.2.25 Example A limit ordinal (e.g. the natural numbers ω_0) is κ-filtered for any finite cardinal κ.

13.2.26 Proposition *Let κ be a cardinal. Then there exists a κ-filtered ordinal ω.*

Proof. If κ is finite, example 13.2.25 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. □

13.2.27 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. (13.2.23.1) 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.

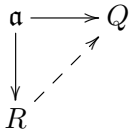
So suppose to the contrary that all of the $f^{-1}(B_\alpha)$ were proper subobjects of A. However, we know that

$$\bigcup f^{-1}(B_\alpha) = f^{-1}\left(\bigcup B_\alpha\right) = A.$$

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 13.2.17): a module Q is injective if and only if in every commutative diagram



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*

$$\begin{array}{ccc} \mathfrak{a} & \longrightarrow & Q \\ \downarrow & & \downarrow \\ R & \longrightarrow & R \oplus_{\mathfrak{a}} Q \end{array} .$$

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.

13.2.28 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

(13.2.28.1)
$$\begin{array}{ccc} \bigoplus \mathfrak{a} & \longrightarrow & N \\ \downarrow & & \downarrow \\ \bigoplus R & \longrightarrow & N \oplus_{\bigoplus \mathfrak{a}} \bigoplus R \end{array}$$

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. (13.2.28.1), 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.

13.2.29 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 13.2.17), 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 13.2.27, 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.

13.3. 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 $M \to \hom_R(N, P)$.

13.3.1 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.

13.3.2 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.

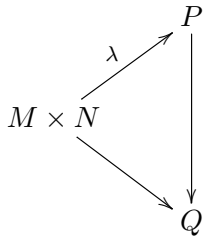
Let us keep the notation of the previous discussion: in particular, M, N, P will be modules over a commutative ring R.

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 $M \times N$.

13.3.3 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**

$$\hom_R(P, Q) \simeq \{\text{bilinear maps } M \times N \to Q\}$$

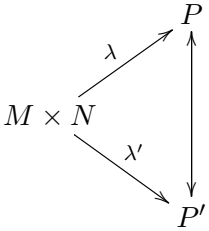
So, given a bilinear map $M \times N \to Q$, there is a *unique* map $P \to Q$ making the diagram



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:



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.

13.3.4 Proposition *Given M, N, a universal bilinear map out of $M \times N$ exists.*

Before proving it we make:

13.3.5 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 13.3.4. 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:

$$\begin{array}{ccc} M \times N & \longrightarrow & M \otimes_R N \\ & \searrow^{\lambda} & \downarrow \\ & & P \end{array}.$$

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

$$M \otimes_R N \to M' \otimes_R N'. \tag{13.3.5.1}$$

To obtain eq. (13.3.5.1), 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.

13.3.6 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. □

13.3.7 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. □

13.3.8 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

$$\begin{aligned} \hom_R(M \otimes (N \otimes P), Q) &\simeq \text{bilinear } M \times (N \otimes P) \to Q \\ &\simeq \hom(N \otimes P, \hom(M, Q)) \\ &\simeq \text{bilinear } N \times P \to \hom(M, Q) \\ &\simeq \hom(N, \hom(P, \hom(M, Q)) \\ &\simeq \text{trilinear maps.} \end{aligned}$$

□

[5] Easy to define.

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:

13.3.9 Proposition *For R-modules M, N, P, there is a natural bijection*

$$\hom_R(M, \hom_R(N, P)) \simeq \hom_R(M \otimes_R N, P).$$

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.

13.3.10 Proposition *Tensoring commutes with colimits.*

In particular, it follows that if $\{N_\alpha\}$ is a family of modules, and M is a module, then

$$M \otimes_R \bigoplus N_\alpha = \bigoplus M \otimes_R N_\alpha.$$

13.3.11 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 13.3.10, 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:

13.3.12 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.

13.3.13 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. □

13.3.14 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.

13.3.15 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

$$\begin{array}{ccc} \mathbb{Z}/(10) \times \mathbb{Z}/(12) & \xrightarrow{\otimes} & \mathbb{Z}/(10) \otimes \mathbb{Z}/(12) \\ & \searrow^{g} & \downarrow f \\ & & \mathbb{Z}/(2). \end{array}$$

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.

13.3.16 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.

13.3.17 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 $\{a_1, \ldots, a_k\}$. 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.

13.3.18 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 $\phi_0 : R \to R_0, \quad \phi_1 : R \to R_1$.

13.3.19 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

$$(R_0 \otimes_R R_1) \times (R_0 \otimes R_1) \to R_0 \otimes_R R_1,$$

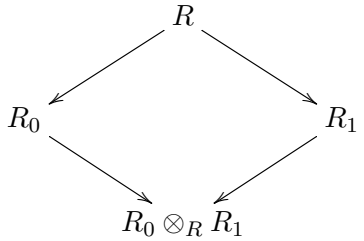
i.e. an R-linear map

$$(R_0 \otimes_R R_1) \otimes_R (R_0 \otimes R_1) \to R_0 \otimes_R R_1$$

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:



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,

13.3.20 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).

13.4. 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.

13.4.1 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. □

13.4.2 Remark Formation of tensor products is, in general, **not** exact.

13.4.3 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.

13.4.4 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.

13.4.5 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

$$\begin{array}{ccccccc} T_{F(R)}(R^I) & \longrightarrow & T_{F(R)}(R^J) & \longrightarrow & T_{F(R)}(M) & \longrightarrow & 0 \\ \downarrow & & \downarrow & & \downarrow & & \\ F(R^I) & \longrightarrow & F(R^J) & \longrightarrow & F(M) & \longrightarrow & 0 \end{array}$$

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. □

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]

13.4.6 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.

13.4.7 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.

[6] Recall that an additive functor commutes with finite direct sums.

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.

13.4.8 Example $\mathbb{Z}/2\mathbb{Z}$ is not flat as a $\mathbb{Z}$-module by 13.4.3.

13.4.9 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.

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

$$\begin{array}{ccc} P \otimes_R M & \longrightarrow & \bigoplus_A M \\ \downarrow & & \downarrow \\ P \otimes_R M' & \longrightarrow & \bigoplus_A M' \end{array} .$$

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. □

13.4.10 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.

13.4.11 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.

13.4.12 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:

13.4.13 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

$$\begin{array}{ccccccc} 0 & \longrightarrow & T_1(M) & \longrightarrow & T_1(F) & \longrightarrow & T_1(F') \\ & & \downarrow & & \downarrow\simeq & & \downarrow\simeq \\ 0 & \longrightarrow & T_2(M) & \longrightarrow & T_2(F) & \longrightarrow & T_2(F'). \end{array}$$

□

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. □

13.4.14 Example Let us now consider finitely generated flat modules over a principal ideal domain R. By 11.5.4, 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 zerodivisor on M. However, we know no element of $R - \{0\}$ can be a zerodivisor on M. It follows that all the $a_i = 0$. In particular, we have proved:

13.4.15 Proposition *A finitely generated module over a PID is flat if and only if it is free.*

Finitely presented flat modules

In example 13.4.9, 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:

13.4.16 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

$$M^* \otimes_R F \to \hom_R(F, M)^*. \tag{13.4.16.1}$$

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 (13.4.16.1) 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 (13.4.16.1) 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:

13.4.17 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). □

Part III.

Fundamentals of Topology

20. General Topology

20.1. The category of topological spaces

Topologies and continuous maps

20.1.1 Definition Let X be a set. By a topology on X on understands a set $\mathcal{O}$ of subsets of X such that:

(Top0) $X \in \mathcal{O}$ and $\emptyset \in \mathcal{O}$.

(Top1) The union of any collection of elements of $\mathcal{O}$ is again in $\mathcal{O}$ that means for each family $(U_i)_{i\in I}$ of $U_i \in \mathcal{O}$ one has $\bigcup_{i\in I} U_i \in \mathcal{O}$.

(Top2) The intersection of finitely many elements of $\mathcal{O}$ is again in $\mathcal{O}$ that means for $U_1, \ldots, U_n \in \mathcal{O}$ with $n \in \mathbb{N}$ one has $\bigcap_{i=1}^{n} U_i \in \mathcal{O}$.

A pair $(X, \mathcal{O})$ is a called a *topological space* when X is a set and $\mathcal{O}$ a topology on X. Moreover, a subset U of X is called *open* if $U \in \mathcal{O}$ and *closed* if $\complement_X U \in \mathcal{O}$.

20.1.2 Remark 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 $\emptyset$, 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) as well to prove that a given set of subsets of set X is a topology, indeed.

20.1.3 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 $\{\emptyset, X\}$ 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 $\{\emptyset, \{1\}, \{0, 1\}\}$ 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 $\emptyset$, $\{0\}$ and S.

(d) The euclidean topology $\mathcal{O}_{\mathbb{R},\mathrm{e}}$ on the set $\mathbb{R}$ of real numbers consists of all sets $U \subset \mathbb{R}$ such that for each $x \in U$ there exist real numbers a, b satisfying $a < x < b$ and $]a, b[\subset U$.

Let us show that $\mathcal{O}_{\mathbb{R},\mathrm{e}}$ is a topology on $\mathbb{R}$ indeed. Obviously $\emptyset$ and $\mathbb{R}$ are elements of $\mathcal{O}_{\mathbb{R},\mathrm{e}}$. Let $U, V \in \mathcal{O}_{\mathbb{R},\mathrm{e}}$ 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 $U \cap V \in \mathcal{O}_{\mathbb{R},\mathrm{e}}$. If $\mathcal{U} \subset \mathcal{O}_{\mathbb{R},\mathrm{e}}$ and $x \in \bigcup \mathcal{U}$, then there exists an $U \in \mathcal{U}$ with $x \in U$. Choose $a, b \in \mathbb{R}$ such that $x \in]a,b[\subset U$. Then $x \in]a,b[\subset U \subset \bigcup \mathcal{U}$, which proves $\bigcup \mathcal{U} \in \mathcal{O}_{\mathbb{R},\mathrm{e}}$. If not mentioned differently, we always assume the set of real numbers to be equipped with the euclidean topology. One therefore sometimes calls $\mathcal{O}_{\mathbb{R},\mathrm{e}}$ the *standard topology* on $\mathbb{R}$.

(e) The euclidean topology $\mathcal{O}_{\mathbb{Q},\mathrm{e}}$ on the set $\mathbb{Q}$ of rational numbers is defined analogously to the previous example as the set 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 $\mathcal{O}_{\mathbb{Q},\mathrm{e}}$ is a topology on $\mathbb{Q}$. Unless mentioned differently it is also always assumed that $\mathbb{Q}$ comes equipped with the euclidean topology.

(f) Let X be a set, and let $\mathcal{O}_{\mathrm{cof}}$ denote the set of all subset of X which are either empty or have finite complement in X. Then $\mathcal{O}_{\mathrm{cof}}$ is a topology on X called the *cofinite topology* on X.

(g) Let X be a (nonempty) set, $(Y, \mathcal{O})$ be a topological space, and $f : X \to Y$ a function. Define

$$f^*\mathcal{O} := f^{-1}\mathcal{O} := \{f^{-1}(U) \mid U \in \mathcal{O}\} .$$

Then $(X, f^*\mathcal{O})$ is a topological space. One calls $f^*\mathcal{O}$ the *initial topology on X with respect to f* or the *topology induced by f*.

Let us verify that $f^*\mathcal{O}$ is a topology on X indeed. By construction, $f^{-1}(Y) = X$ and $f^{-1}(\emptyset) = \emptyset$, so $\emptyset, X \in f^*\mathcal{O}$. Now let $(V_i)_{i\in I}$ be a family of elements of $f^*\mathcal{O}$. In other words we have, for each $i \in I$, $V_i = f^{-1}(U_i)$ for some $U_i \in \mathcal{O}$. Then $U := \bigcup_{i\in I} U_i \in \mathcal{O}$ and

$$\bigcup_{i\in I} V_i = \bigcup_{i\in I} f^{-1}(U_i) = f^{-1}\Big(\bigcup_{i\in I} U_i\Big) = f^{-1}(U) \in f^*\mathcal{O} .$$

Finally, let $V_1, \ldots, V_n \in f^{-1}\mathcal{O}$. Then, by definition, there exist $U_1, \ldots, U_n \in \mathcal{O}$ such that $V_i = f^{-1}(U_i)$ for $i = 1, \ldots, n$. Thus $U := \bigcap_{i=1}^n U_i \in \mathcal{O}$ and

$$\bigcap_{i=1}^{n} V_i = \bigcap_{i=1}^{n} f^{-1}(U_i) = f^{-1}\Big(\bigcap_{i=1}^{n} U_i\Big) = f^{-1}(U) \in f^*\mathcal{O} .$$

Section 20.2 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.

20.1.4 Definition Let $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ be two topological spaces and assume that $f : X \to Y$ is a function. One says that f is *continuous* if for all $U \in \mathcal{O}_Y$ the preimage $f^{-1}(U)$ is open in X. The map f is called *open* if $f(V)$ is open in Y for all $V \in \mathcal{O}_X$.

20.1.5 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.

20.1.6 Proposition and Definition (a) *The identity map* id_X *on a topological space* $(X, \mathcal{O}_X)$ *is continuous and open.*

(b) *Let* $(X, \mathcal{O}_X)$, $(Y, \mathcal{O}_Y)$ *and* $(Z, \mathcal{O}_Z)$ *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 $U \in \mathcal{O}_Z$. Then $g^{-1}(U) \in \mathcal{O}_Y$ by continuity of g. Hence $f^{-1}(g^{-1}(U)) \in \mathcal{O}_X$ by continuity of f. So $g \circ f$ is continuous. If f and g are open maps, and $V \in \mathcal{O}_X$, then $f(V) \in \mathcal{O}_Y$ and $g \circ f(V) = g(f(V)) \in \mathcal{O}_Z$. Hence the composition of two open maps is open, too. The rest of the claim follows immediately. □

Comparison of topologies

The initial topology $f^*\mathcal{O}$ 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.

20.1.7 Definition Let X be a set. Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be two topologies on X. One says that $\mathcal{O}_1$ is *finer* or *stronger* than $\mathcal{O}_2$ and $\mathcal{O}_2$ is *coarser* or *weaker* than $\mathcal{O}_1$ when $\mathcal{O}_2 \subset \mathcal{O}_1$.

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.

20.1.8 Theorem *Let* X *be a set. Let* $\mathcal{T}$ *be a nonempty set of topologies on* E*. Then the set*

$$\mathcal{O}_\mathcal{T} := \bigcap \mathcal{T} = \{U \in \mathcal{P}(X) \mid U \in \mathcal{O} \text{ for all } \mathcal{O} \in \mathcal{T}\}$$

is a topology on X *and it is the greatest lower bound of* $\mathcal{T}$*, where the order between topologies is given by inclusion. In other words,* $\mathcal{O}_\mathcal{T}$ *is the finest topology contained in each topology from* $\mathcal{T}$*.*

Proof. We first show that $\mathcal{O}_\mathcal{T}$ is a topology. Since each $\mathcal{O} \in \mathcal{T}$ is a topology on X, we have $\emptyset, X \in \mathcal{O}$ for all $\mathcal{O} \in \mathcal{T}$. Hence $\emptyset, X \in \mathcal{O}_\mathcal{T}$.

Let $(U_i)_{i \in I}$ be a family of elements $U_i \in \mathcal{O}_\mathcal{T}$. Let $\mathcal{O} \in \mathcal{T}$ be arbitrary. By definition of $\mathcal{O}_\mathcal{T}$, we have $U_i \in \mathcal{O}$ for all $i \in I$. Since $\mathcal{O}$ is a topology, $\bigcup_{i \in I} U_i \in \mathcal{O}$. Hence, as $\mathcal{O}$ was arbitrary, $\bigcup_{i \in I} U_i \in \mathcal{O}_\mathcal{T}$.

Now, let $U_1, \ldots, U_n \in \mathcal{O}_\mathcal{T}$. Let $\mathcal{O} \in \mathcal{T}$ be arbitrary. By definition of $\mathcal{O}_\mathcal{T}$, we have $U_1, \ldots, U_n \in \mathcal{O}$. Therefore, $U_1 \cap \ldots \cap U_n \in \mathcal{O}$ since $\mathcal{O}$ is a topology. Since $\mathcal{O}$ was arbitrary in $\mathcal{T}$, we conclude that $U_1 \cap \ldots \cap U_n \in \mathcal{O}_\mathcal{T}$ by definition.

So $\mathcal{O}_\mathcal{T}$ is a topology on X. By construction, $\mathcal{O}_\mathcal{T} \subset \mathcal{O}$ for all $\mathcal{O} \in \mathcal{T}$, so $\mathcal{O}_\mathcal{T}$ is a lower bound for $\mathcal{T}$. Assume given a new topology $\mathcal{Q}$ on X such that $\mathcal{Q} \subset \mathcal{O}$ for all $\mathcal{O} \in \mathcal{T}$. Let $U \in \mathcal{Q}$. Then we have $U \in \mathcal{O}$ for all $\mathcal{O} \in \mathcal{T}$. Hence by definition $U \in \mathcal{O}_\mathcal{T}$. So $\mathcal{Q} \subset \mathcal{O}_\mathcal{T}$ and thus $\mathcal{O}_\mathcal{T}$ is the greatest lower bound of $\mathcal{T}$. □

20.1.9 Corollary *Let X be a set and $(Y, \mathcal{O}_Y)$ be a topological space. The coarsest topology on X which makes a function $f : X \to Y$ continuous is the initial topology $f^*\mathcal{O}$.*

Proof. Let $\mathcal{T}$ be the set of all topologies on X such that f is continuous. By definition, $f^*\mathcal{O}$ is a lower bound of $\mathcal{T}$. Moreover, $f^*\mathcal{O} \in \mathcal{T}$. Hence $f^*\mathcal{O}$ is the coarsest topology making the function $f : X \to Y$ continuos. □

We can use Theorem 20.1.8 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:

20.1.10 Proposition and Definition *Let X be a set, and $\mathcal{S}$ a subset of $\mathcal{P}(X)$. The greatest lower bound of the set*

$$\mathcal{T} = \{\mathcal{O} \in \mathcal{P}(\mathcal{P}(X)) \mid \mathcal{O} \text{ is a topology on } X \,\&\, \mathcal{S} \subset \mathcal{O}\}$$

is the coarsest topology on X containing $\mathcal{S}$. We call it the topology generated by $\mathcal{S}$ on X *and denote it by $\mathcal{O}_\mathcal{S}$. The topology $\mathcal{O}_\mathcal{S}$ consists of unions of finite intersections of elements of $\mathcal{S}$ that means*

$$\mathcal{O}_\mathcal{S} = \left\{U \in \mathcal{P}(X) \mid \exists J\, \forall j \in J\, \exists n_j \in \mathbb{N}\, \exists U_{j,1}, \ldots, U_{j,n_j} \in \mathcal{S} :\ U = \bigcup_{j\in J} \bigcap_{k=1}^{n_j} U_{j,k}\right\} .$$

Proof. By definition of $\mathcal{T}$ and Theorem 20.1.8, $\mathcal{O}_\mathcal{T} = \bigcap \mathcal{T}$ is a topology on X which contains $\mathcal{S}$. Hence $\mathcal{O}_\mathcal{T}$ is an element of $\mathcal{T}$ as well and a subset of any element of $\mathcal{T}$. The first claim follows. To verify the second, observe that it suffices to show that

$$\mathcal{Q} := \left\{U \in \mathcal{P}(X) \mid \exists J\, \forall j \in J\, \exists n_j \in \mathbb{N}\, \exists U_{j,1}, \ldots, U_{j,n_j} \in \mathcal{S} :\ U = \bigcup_{j\in J} \bigcap_{k=1}^{n_j} U_{j,k}\right\}$$

is a topology. The set $\mathcal{Q}$ being a topology namely entails $\mathcal{O}_\mathcal{S} \subset \mathcal{Q}$, because $\mathcal{S} \subset \mathcal{Q}$, and $\mathcal{Q} \subset \mathcal{O}_\mathcal{S}$ is clear by definition, since $\mathcal{O}_\mathcal{S}$ is a topology containing $\mathcal{S}$. The second claim $\mathcal{Q} = \mathcal{O}_\mathcal{S}$ then follows. So let us show that $\mathcal{Q}$ is a topology. Obviously $\emptyset$ and X are elements of $\mathcal{Q}$ because $\bigcup_{i\in\emptyset} U_i = \emptyset$ and $\bigcap_{k=1}^{0} U_k = X$. Now assume that $(U_i)_{i\in I}$ is a family of elements of $\mathcal{Q}$. 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

$$U := \bigcup_{i\in I} U_i = \bigcup_{i\in I} \bigcup_{j\in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} = \bigcup_{(i,j)\in J} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} \in \mathcal{Q} .$$

Last assume $U_1, \dots U_n \in \mathcal{O}$ where $n \in \mathbb{N}$. Then one can find for each $i \in \{1, \dots, n\}$ a set J_i and for every $j \in J_i$ a natural number $n_{i,j}$ together with elements $U_{i,j,1}, \dots, U_{i,j,n_{i,j}} \in \mathcal{S}$ 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

$$U := \bigcap_{i=1}^{n} U_i = \bigcap_{i=1}^{n} \bigcup_{j \in J_i} \bigcap_{k=1}^{n_{i,j}} U_{i,j,k} = \bigcup_{(j_1,\dots,j_n) \in J} \bigcap_{k_1=1}^{n_{1,j_1}} U_{1,j_1,k_1} \cap \ldots \cap \bigcap_{k_n=1}^{n_{n,j_n}} U_{n,j_n,k_n} \in \mathcal{Q} \ .$$

Hence $\mathcal{Q}$ is a topology, indeed, and the proposition is proved. □

20.1.11 Definition Let X be a set, and $\mathcal{O}$ a topology on X. One calls a subset $\mathcal{S} \subset \mathcal{O}$ a *subbase* or *subbasis* of the topology $\mathcal{O}$ if $\mathcal{O}$ coincides with $\mathcal{O}_{\mathcal{S}}$. If in addition $X = \bigcup_{S \in \mathcal{S}} S$, the subbase $\mathcal{S}$ is said to be *adequate*.

Bases of topologies

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 $\mathcal{O}_{\mathcal{B}}$ generated by $\mathcal{B}$.

20.1.12 Definition Let X be a set. A (*topological*) *base* 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 $B, B' \in \mathcal{B}$ and all $x \in B \cap B'$ there exists a $B'' \in \mathcal{B}$ such that $x \in B''$ and $B'' \subset B \cap B'$.

The main purpose for this definition stems from the following theorem:

20.1.13 Theorem *Let X be some set. Let $\mathcal{B}$ be a topological basis on E. Then the topology generated by $\mathcal{B}$ coincides with the set of unions of elements of $\mathcal{B}$ that means*

$$\mathcal{O}_{\mathcal{B}} = \Big\{ \bigcup_{B \in \mathcal{U}} B \mid \mathcal{U} \subset \mathcal{B} \Big\} \ .$$

Proof. Denote, for this proof, the set $\{\bigcup \mathcal{U} : \mathcal{U} \subseteq \mathcal{B}\}$, by σ, and let us abbreviate $\mathcal{O}(\mathcal{B})$ by $\mathcal{O}$. We wish to prove that $\mathcal{O} = \sigma$. First, note that $\mathcal{B} \subseteq \sigma$ by construction. By definition, $\mathcal{B} \subseteq \mathcal{O}$ and since $\mathcal{O}$ is a topology, it is closed under arbitrary unions. Hence $\sigma \subseteq \mathcal{O}$. To prove the converse, it is sufficient to show that σ is a topology. As it contains $\mathcal{B}$ and $\mathcal{O}$ is the smallest such topology, this will provide us with the inverse inclusion. By definition, $\bigcup \emptyset = \emptyset$ and thus $\emptyset \in \sigma$. By assumption, since $\mathcal{B}$ is a basis, $E = \bigcup \mathcal{B}$ so $E \in \sigma$. As the union of unions of elements in $\mathcal{B}$ is a union of elements in $\mathcal{B}$, σ is closed under abritrary unions. Now, let U, V be elements of $\mathcal{B}$. If $U \cap V = \emptyset$ then $U \cap V \in \sigma$. Assume that U and

V are not disjoints. Then by definition of a basis, for all $x \in U \cap V$ there exists $W_x \in \mathcal{B}$ such that $x \in W_x$ and $W_x \subseteq U \cap V$. So:

$$U \cap V = \bigcup_{x \in U \cap V} W_x$$

and therefore, by definition, $U \cap V \in \sigma$. We conclude that the intersection of two arbitary elements in σ is again in σ by using the distributivity of the union with respect to the intersection. □

20.1.14 Definition We shall say that a base $\mathcal{B}$ on a set X is *a base for a topology* $\mathcal{O}$ on X when the smallest topology containing $\mathcal{B}$ coincides with $\mathcal{O}$, in other words when $\mathcal{O} = \mathcal{O}_{\mathcal{B}}$.

The typical usage of the preceding theorem comes from the following result.

20.1.15 Corollary *Let $\mathcal{B}$ be a topological base for a topology $\mathcal{O}$ on X. A subset U of X is in $\mathcal{O}$ if and only if for any $x \in U$ 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 $\mathcal{O}$ is a union of elements in $\mathcal{B}$: hence if $x \in U$ for $U \in \mathcal{O}$ then there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq U$. Conversely, if U is some subset of E such that for all $x \in U$ there exists $B_x \in \mathcal{B}$ such that $x \in B_x$ and $B_x \subseteq U$ then $U = \bigcup_{x \in U} B_x$ and thus $U \in \mathcal{O}$. □

As a basic application, we show that:

20.1.16 Corollary *Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be two topological spaces. Let $\mathcal{B}$ be a basis for the topology $\mathcal{O}_E$. Let $f : E \to F$. Then f is continuous on E if and only if:*

$$\forall V \in \mathcal{O}_F \;\; \forall x \in f^{-1}(V) \;\; \exists B \in \mathcal{B} \;\; x \in B \wedge B \subset f^{-1}(V).$$

20.1.17 Corollary *Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be two topological spaces. Let $\mathcal{B}$ be a basis for the topology $\mathcal{O}_F$. Let $f : E \to F$. Then f is continuous on E if and only if:*

$$\forall V \in \mathcal{B} \;\; f^{-1}(V) \in \mathcal{O}_E.$$

Proof. By definition, continuity of f implies 20.1.17. Conversely, assume 20.1.17 holds. Let $V \in \mathcal{O}_F$. Then there exists $\mathcal{U} \subseteq \mathcal{B}$ such that $V = \bigcup \mathcal{U}$. Now by assumption, $f^{-1}(B) \in \mathcal{O}_E$ for all $B \in \mathcal{U}$ and thus $f^{-1}(V) = \bigcup_{B \in \mathcal{U}} f^{-1}(B) \in \mathcal{O}_E$ since $\mathcal{O}_E$ is a topology. □

We leave to the reader to write the statement when both E and F have a basis.

20.2. Fundamental examples of topologies

This section provides various examples of topological spaces which will be used all along this book.

The order topology

20.2.1 Proposition *Let $(X, \leq)$ be a totally ordered set, and assume that $\infty, -\infty$ are two symbols not in X. Define $[-\infty, \infty] = X \cup \{-\infty, \infty\}$ and extend $\leq$ to $[-\infty, \infty]$ by requiring $x \leq y$ for $x, y \in [-\infty, \infty]$ to hold when $x, y \in X$ and $x \leq y$, when $x = -\infty$, or when $y = \infty$. Then $[-\infty, \infty]$ together with the relation $\leq$ becomes a totally ordered set, as well, and the embedding $X \hookrightarrow [-\infty, \infty]$ is order-preserving.*

Proof. By definition, the relation $\leq$ on $[-\infty, \infty]$ is reflexive. Assume that $x \leq y$ and $y \leq x$. Then □

20.2.2 Remark For the rest of this paragraph we always assume that an ordered set $(X, \leq)$ does not contain the symbols $\infty, -\infty$, and that $[-\infty, \infty]$ and the extended order relation $\leq$ are defined as in the preceding proposition.

20.2.3 Definition For a totally ordered set $(X, \leq)$, define *intervals* with boundaries $x, y \in [-\infty, \infty]$, where $x \leq y$ is required, as follows:

$$\begin{aligned}]x, y[&:= \{z \in [-\infty, \infty] \mid x < z < y\} , \\ [x, y[&:= \{z \in [-\infty, \infty] \mid x \leq z < y\} , \\ ]x, y] &:= \{z \in [-\infty, \infty] \mid x < z \leq y\} , \\ [x, y] &:= \{z \in [-\infty, \infty] \mid x \leq z \leq y\} . \end{aligned}$$

20.2.4 Definition Let $(X, \leq)$ be a totally ordered set. Then the topology generated by the set

$$\mathcal{I}_X = \{\,]x, y[\in \mathcal{P}(X) \mid x, y \in [-\infty, \infty] \,\&\, x \leq y\}$$

is called the *order topology* on X.

20.2.5 Proposition *Let $(X, \leq)$ 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 $]x, \infty[$ *and* $]-\infty, y[$*, where x, y run through the elemants 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 $x \leq y$, the set $\mathcal{S}_X$ is a subbase of the order topology. □

20.2.6 Example The standard topology on $\mathbb{R}$ from Example 20.1.3 (d) is the order topology. Likewise, the standard topology on $\mathbb{Q}$ coincides with the order topology.

20.2.7 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 $(\mathbb{R}, \leq)$ in that way gives the so-called *extended real number system* $\overline{\mathbb{R}}$.

The subspace topology

20.2.8 Proposition and Definition *Let $(X, \mathcal{O})$ be a topological space. Let $S \subset X$ and $\iota : S \hookrightarrow X$ the canonical embedding. Then initial topology $\iota^*\mathcal{O}$ coincides with*

$$\mathcal{O}_S^X := \{U \cap S \in \mathcal{P}(S) \mid U \in \mathcal{O}\} \ .$$

One calls $\mathcal{O}_Y^X$ the subspace *or* trace topology *on S. Sometimes one says that $\mathcal{O}_Y^X$ is the* topology induced by $(X, \mathcal{O})$.

Proof. The claim follows immediately from the definition of the initial topology $\iota^*\mathcal{O}$. □

Just as easy is the following observation:

20.2.9 Proposition *Let $(X, \mathcal{O})$ be a topological space, and $S \subset X$ a subset. Let $\mathcal{B}$ be a basis for $\mathcal{O}$. 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 $(X, \mathcal{O})$.

Proof. Trivial exercise. □

20.2.10 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

20.2.11 Definition Let I be some nonempty set. Let us assume given a family $(X_i, \mathcal{O}_i)_{i \in I}$ 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 $U_i \in \mathcal{O}_i$.

20.2.12 Definition Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{O}_i)_{i \in I}$ of topological spaces. The product topology on $\prod_{i \in I} E_i$ is the smallest topology containing all the basic open sets.

20.2.13 Proposition *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{O}_i)_{i \in I}$ 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. □

20.2.14 Remark The product topology is not just the basic open sets on the cartesian products: there are many more open sets!

20.2.15 Proposition *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{O}_i)_{i\in I}$ 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 $V \in \mathcal{O}_{E_i}$. By definition, $p_i^{-1}(V) = \prod_{j\in I} U_j$ where $U_j = E_j$ for $j \in I \setminus \{i\}$, 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 $\{p_i^{-1}(V) : i \in I, V \in \mathcal{O}_{E_i}\}$, so it is coarser than the initial topology for $\{p_i : i \in I\}$. This concludes this proof. □

20.2.16 Corollary *Let I be some nonempty set. Let us assume given a family $(E_i, \mathcal{O}_i)_{i\in I}$ of topological spaces. Let $\mathcal{O}$ be the product topology on $F = \prod_{i\in I} E_i$. Let $(D, \mathcal{O}_D)$ be a topological space. Then $f : D \to F$ is continuous if and only if $p_i \circ f$ is continuous from $(D, \mathcal{O}_D)$ to $(E_i, \mathcal{O}_{E_i})$ 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. □

20.2.17 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

20.2.18 Definition Let E be a set. A function $d : E \times E \to [0, \infty)$ is a distance on E when:

1. For all $x, y \in E$, we have $d(x, y) = 0$ if and only if $x = y$,
2. For all $x, y \in E$ we have $d(x, y) = d(y, x)$,
3. For all $x, y, z \in E$ we have $d(x, y) \leq d(x, z) + d(z, y)$.

20.2.19 Definition A pair (E, d) is a metric space when E is a set and d a distance on E.

The following is often useful:

20.2.20 Proposition *Let (E,d) be a metric space. Let $x, y, z \in E$. Then:*

$$|d(x,y) - d(x,z)| \leq d(y,z).$$

Proof. Since $d(x,y) \leq d(x,z) + d(z,y)$ we have $d(x,y) - d(x,z) \leq d(z,y) = d(y,z)$. Since $d(x,z) \leq d(x,y) + d(y,z)$ we have $d(x,z) - d(x,y) \leq d(y,z)$. Hence the proposition holds.□

20.2.21 Definition Let (E,d) be a metric space. Let $x \in E$ and $r \in (0,\infty) \subseteq \mathbb{R}$. The open ball of center x and radius r in (E,d) is the set:

$$B(x,r) = \{y \in E : d(x,y) < r\}.$$

20.2.22 Definition Let (E,d) be a metric space. The metric topology on E induced by d is the smallest topology containing all the open balls of E.

20.2.23 Theorem *Let (E,d) be a metric space. The set of all open balls on E is a basis for the metric topology on E induced by d.*

Proof. It is enough to show that the set of all open balls is a basis. By definition, $E = \bigcup_{x \in E} B(x,1)$. Now, let us be given $B(x,r_x)$ and $B(y,r_y)$ for some $x, y \in E$ and $r_x, r_y > 0$. If the intersection of these two balls is empty, we are done; let us assume that there exists $z \in B(x,r_x) \cap B(y,r_y)$. Let ρ be the smallest of $r_x - d(x,z)$ and $r_y - d(y,z)$. Let $w \in B(z,\rho)$. Then:

$$d(x,w) \leq d(x,z) + d(z,w) < d(x,z) + r_x - d(x,z) = r_x$$

so $w \in B(x,r_x)$. Similarly, $w \in B(y,r_y)$. Hence, $B(z,\rho) \subseteq B(x,r_x) \cap B(y,r_y)$ as desired.□

The following theorem shows that metric topologies are minimal in the sense of making the distance functions continuous.

20.2.24 Theorem *Let (E,d) be a metric space. For all $x \in E$, the function $y \in E \mapsto d(x,y)$ is continuous on E for the metric topology. Moreover, the metric topology is the smallest topology such that all the functions in the set $\{y \mapsto d(x,y) : x \in E\}$ are continuous.*

Proof. Fix $x \in E$. It is sufficient to show that the preimage of $[0,r)$ and (r,∞) by $d_x : y \in E \mapsto d(x,y)$ is open in the metric topology of E, where $r \geq 0$ is arbitrary. Indeed, these intervals form a basis for the topology of $[0,\infty)$. Let $r \geq 0$ be given. Then $d_x^{-1}([0,r)) = B(x,r)$ by definition, so it is open. Moreover, it shows that the minimal topology making all these maps continuous must indeed contain the metric topology. Now, let $y \in E$ such that $d(x,y) > r$. Let $\rho = d(x,y) - r > 0$. Then if $d(w,y) < \rho$ for some $w \in E$ then:

$$d(x,y) \leq d(x,w) + d(w,y) \text{ so } d(x,y) - d(w,y) \leq d(x,w)$$

so $d(x,w) > r$. Hence

$$B(y,\rho) \subset d_x^{-1}((r,\infty))$$

for all $y \in d_x^{-1}((r,\infty))$. Therefore, $d_x^{-1}((r,\infty))$ is open, as desired, and our proposition is proven. □

20.2.25 Remark The topology on $[0, \infty)$ is the trace topology on $[0, \infty)$ induced by the usual, i.e. the order topology on $\mathbb{R}$.

20.2.26 Remark The metric topology is the default topology on a metric space.

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:

20.2.27 Definition Let (E, d_E), (F, d_F) be metric spaces. A function $f : E \to F$ is k-Lipschitz for $k \in [0, \infty)$ if:

$$\forall x, y \in E \;\; d_F(f(x), f(y)) \leq k d_E(x, y).$$

20.2.28 Definition Let (E, d_E), (F, d_F) be metric spaces. Let $f : E \to F$ be a Lipschitz function. Then the Lipschitz constant of f is defined by:

$$\mathrm{Lip}(f) = \sup \left\{ \frac{d_F(f(x), f(y))}{d_E(x, y)} : x, y \in E, x \neq y \right\}.$$

20.2.29 Remark $\mathrm{Lip}(f) = 0$ if and only if f is constant.

20.2.30 Proposition *Let (E, d_E), (F, d_F) be metric spaces. If $f : E \to F$ is a Lipschitz function, then it is continuous.*

Proof. Assume f is nonconstant (otherwise the result is trivial). Let k be the Lipschitz constant for f. Let $y \in F$ and $\epsilon > 0$. Let $x \in f^{-1}(B(y, \epsilon))$. Let $z \in E$ such that $d_E(x, z) < \delta_x = \frac{\epsilon - d(f(x), y)}{k}$ (note that the upper bound is nonzero).

$$d_F(f(z), y) \leq d_F(f(z), f(x)) + d_F(f(x), y) \tag{20.2.30.1}$$
$$\leq k d_E(x, z) + d_F(f(x), y) \tag{20.2.30.2}$$
$$< \epsilon - d_F(f(x), y) + d_F(f(x), y) = \epsilon. \tag{20.2.30.3}$$

Hence $f^{-1}(B(y, \epsilon)) = \bigcup_{x \in f^{-1}(B(y,\epsilon))} B(x, \delta_x)$. So f is continuous. □

20.2.31 Remark The proof of continuity for Lipshitz maps can be simplified: it is a consequence of the squeeze theorem. We refer to the chapter on metric spaces for this.

20.2.32 Remark Using Lipshitz maps as morphisms for a category of metric spaces is natural. Another, more general type of morphisms, would be uniform continuous maps, which are discussed in the compact space chapter.

Co-Finite Topologies

A potential source for counter-examples, the family of cofinite topologies is easily defined:

20.2.33 Proposition *Let E be a set. Let:*

$$\mathcal{O}_{\text{cof}}(E) = \{\emptyset\} \cup \{U \subset E : \complement_E U \text{ is finite }\}.$$

Then $\mathcal{O}_{\text{cof}}(E)$ is a topology on E.

Proof. By definition, $\emptyset \in \mathcal{O}_{\text{cof}}(E)$. Moreover, $\complement_E E = \emptyset$ which is finite, so $E \in \mathcal{O}_{\text{cof}}(E)$. Let $U, V \in \mathcal{O}_{\text{cof}}(E)$. If U or V is empty then $U \cap V = \emptyset$ so $U \cap V \in \mathcal{O}_{\text{cof}}(E)$. 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 $U \cap V \in \mathcal{O}_{\text{cof}}(E)$. Last, let $\mathcal{U} \subseteq \mathcal{O}_{\text{cof}}(E)$. Again, if $\mathcal{U} = \{\emptyset\}$ then $\bigcup \mathcal{U} = \emptyset \in \mathcal{O}_{\text{cof}}(E)$. Let us now assume that $\mathcal{U}$ contains at least one nonempty set V. Then:

$$\complement_E \bigcup \mathcal{U} = \bigcap \{\complement_E U : U \in \mathcal{U}\} \subseteq \complement_E V.$$

Since $\complement_E V$ is finite by definition, so is $\bigcup \mathcal{U}$, which is therefore in $\mathcal{O}_{\text{cof}}(E)$. 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.

20.2.34 Definition Let ∞ be some symbol not found in $\mathbb{N}$. We define $\overline{\mathbb{N}}$ to be $\mathbb{N} \cup \{\infty\}$.

20.2.35 Proposition *The set:*

$$\mathcal{O}_{\overline{\mathbb{N}}} = \{U \subseteq \overline{\mathbb{N}} : (U \subseteq \mathbb{N}) \vee (\infty \in U \wedge \complement_{\mathbb{N}} U \text{ is finite})\}$$

is a topology on $\overline{\mathbb{N}}$.

Proof. By definition, $\emptyset \subseteq \mathbb{N}$ so $\emptyset \in \mathcal{O}_{\overline{\mathbb{N}}}$. Moreover $\complement_{\overline{\mathbb{N}}} \overline{\mathbb{N}} = \emptyset$ which has cardinal 0 so $\overline{\mathbb{N}} \in \mathcal{O}_{\overline{\mathbb{N}}}$. Let $U, V \in \mathcal{O}_{\overline{\mathbb{N}}}$. If either U or V is a subset of $\mathbb{N}$ then $U \cap V$ is a subset of $\mathbb{N}$ so $U \cap V \in \mathcal{O}_{\overline{\mathbb{N}}}$. Othwiwse, $\infty \in U \cap V$. Yet $\complement_{\overline{\mathbb{N}}}(U \cap V) = \complement_{\overline{\mathbb{N}}} U \cup \complement_{\overline{\mathbb{N}}} V$ which is finite as a finite union of finite sets. Hence $U \cap V \in \mathcal{O}_{\overline{\mathbb{N}}}$ again.

Last, assume that $\mathcal{U} \subseteq \mathcal{O}_{\overline{\mathbb{N}}}$. 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 $\bigcup \mathcal{U} \in \mathcal{O}_{\overline{\mathbb{N}}}$ by definition. If, on the other hand, $\infty \in \bigcup \mathcal{U}$, then there exists $U \in \mathcal{U}$ with $\complement_{\overline{\mathbb{N}}} U$ finite. Now, $\complement_{\overline{\mathbb{N}}} \bigcup \mathcal{U} = \bigcap \{\complement_{\overline{\mathbb{N}}} V : V \in \mathcal{U}\} \subseteq \complement_{\overline{\mathbb{N}}} U$ so it is finite, and thus again $\bigcup \mathcal{U} \in \mathcal{O}_{\overline{\mathbb{N}}}$. □

20.3. Separation properties

20.3.1 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

20.3.2 Definition (The Separation Axioms) Recall that two subsets A, B of a topological space $(X, \mathcal{O})$ are called *disjoint* if $A \cap B = \emptyset$. The two sets are called *separated* if $\overline{A} \cap B = A \cap \overline{B} = \emptyset$. The topological space $(X, \mathcal{O})$ 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 $y \in V$,

($\mathsf{T2}_{\frac{1}{2}}$) 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 $A \subset V$,

($\mathsf{T3}_{\frac{1}{2}}$) 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 (T3) , 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) .

20.4. Filters and convergence

Filters and ultrafilters

20.4.1 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 $\emptyset$ 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 $\bigcap_{A\in\mathcal{F}} A = \emptyset$ otherwise it is called *fixed*.

20.4.2 Examples (a) For every 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 $(X, \mathcal{O})$ is a topological space and $x \in X$ an element, then the *neigborhood filter* $\mathcal{U}_x := \{V \in \mathcal{P}(X) \mid \exists U \in \mathcal{O} : \ x \in U \subset V\}$ 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).

20.4.3 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 $Y_1 \cap \ldots \cap Y_n \neq \emptyset$. 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. □

20.4.4 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. □

20.4.5 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 $X \setminus A \in \mathcal{F}$.*

Convergence of filters

20.4.6 Definition

20.5. Nets

Directed sets

Let us first recall that by a *preorderd set* one understands a set P together with a binary relation which is reflexive, i.e. $x \leq x$ for all $x \in P$, and transitive, i.e. for all $x, y, z \in P$ the relation $x \leq y$ and $y \leq z$ implies $x \leq z$.

20.5.1 Definition (Directed sets) By a directed set one understands a set $(P,)$ together with a binary relation $\leq$ that is

(Dir1) *directed*, i.e. for all $x, y \in D$ exists a $z \in D$ such that $x \leq z$ and $y \leq z$.

20.5.2 Remark The property that $(P, \leq)$ is directed is the same as saying that any two elements of the preordered set P have an upper bound.

20.6. Compactness

Quasi-compact topological spaces

20.6.1 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 $(X, \mathcal{O})$ 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.

20.6.2 Proposition *Let be a topological spaces $(X, \mathcal{O})$. 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 $\bigcap_{i\in I} A_i = \emptyset$ there exist finitely many elements $A_{i_1}, \ldots, A_{i_n}$ such that $A_{i_1} \cap \ldots \cap A_{i_n} = \emptyset$.*

(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 $\bigcap_{i\in I} A_i = \emptyset$. Put $U_i := X \setminus A_i$ 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 $A_{i_1} \cap \ldots \cap A_{i_n} = \emptyset$ the follows, hence (ii) follows.

Next assume (ii), and let $\mathcal{F}$ be a filter on X. Then $\overline{A_1} \cap \ldots \cap \overline{A_n} \neq \emptyset$ for all $n \in \mathbb{N}^*$ and $A_1, \ldots, A_n \in \mathcal{F}$, since $\mathcal{F}$ is a filter. Hence $\bigcap_{A\in\mathcal{F}} \overline{A} \neq \emptyset$ by (ii). Every element of $\bigcap_{A\in\mathcal{F}} \overline{A}$ 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 $B_J := X \setminus \bigcup_{i\in J} U_i$ 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 $X \setminus U_i \in \mathcal{B} \subset \mathcal{F}$ by construction. This is a contradiction, so $\mathcal{U}$ must have a finite subcover. □

20.6.3 Definition (?) A topological space $(X, \mathcal{O})$ is called *quasi-compact*, if every filter on X has an accummulation point.

20.6.4 Theorem (Alexander Subbase Theorem) *Let $(X, \mathcal{O})$ be a topological space, and $\mathcal{S}$ an* adequate *subbase of the topology that is a subbase of $\mathcal{O}$ 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 $(X, \mathcal{O})$ is quasi-compact.*

Compact topological spaces

20.7. 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*.

20.7.1 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.

20.7.2 Definition A topology $\mathcal{O}$ 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.

20.7.3 Proposition *The compact open topology is coarser than any admissable topology on* $\mathcal{C}(Y,X)$.

Proof. Let $\mathcal{O}$ 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 $g \in T(K,U)$ □

20.7.4 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

$$\Psi : Z^{Y^X} \to Z^{X\times Y} g \mapsto ((x,y) \mapsto g(x)(y))$$

20.7.5 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 $y_1, y_2, \dots y_k \subset Y$ such that $K \subset V_{y_1} \cup V_{y_2} \cup \dots \cup V_{y_k}$. Put $W = W_{y_1} \cap W_{y_2} \cap \dots \cap W_{y_k}$ 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

$$\Psi(g)(O\times V) \subset g(O)(V) \subset g(O)(\bar{V}) \subset U$$

20.7.6 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.

$$T\left(K, \bigcup_I W_i\right) = \bigcup_{n\in\mathbb{N}^+} \bigcup_{\substack{K_1\times\dots\times K_n \subset K^n \\ K_1\cup\dots\cup K_n = K \\ K_i = \bar{K_i} \forall i}} \bigcup_{(i_1,\dots,i_n)\in I^n} \bigcap_{l=1}^{n} T(K_{i_l}, W_{i_l})$$

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

$$T(K,W) = \bigcup_{n\in\mathbb{N}^+} \bigcup_{\substack{K_1\times\dots\times K_n \subset K^n \\ K_1\cup\dots\cup K_n = K \\ K_i = \bar{K_i} \forall i}} \bigcup_{(i_1,\dots,i_n)\in J^n} \bigcap_{l=1}^{n} \bigcap_{j\in J_{i_l}} T(K_{i_l}, T(L_{i_l j}, U_{i_l j}))$$

20.7.7 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 20.7.6, there is an open $U \subset Z$ and compact subsets $L \subset Y$ and $K \subset X$ such that $\bar{phi}(f) \in T(K, T(L, U)) \subset W$. $T(K \times L, U)$ 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 proof in email 9/27/10

21. Sheaves

21.1. Presheaves

The category of open sets of a topological space

21.1.1 Before we define presheaves and sheaves on a topological space $(X, \mathcal{O})$, we briefly introduce the category $\mathsf{Ouv}(X)$ of open sets of $(X, \mathcal{O})$. By definition, its object class coincides with the set of open sets $\mathcal{O}$, 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 $V \not\subset U$ and $V \not\subset U$ and consists of the canonical (identical) embedding $i_{U,V} : V \hookrightarrow U$ when $V \subset U$. 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

$$i_{U,V} \circ i_{V,W} = i_{U,W} : W \hookrightarrow U \quad \text{for } U, V, W \in \mathcal{O} \text{ with } W \subset V \subset U.$$

This observation entails that $\mathsf{Ouv}(X)$ is a category indeed; it is called the *category of open sets* on the topological space $(X, \mathcal{O})$.

21.1.2 Remarks (a) The topology $\mathcal{O}$ carries a natural partial order given by set-theoretic inclusion, so becomes a poset. The corresponding category structure from Example 1.1.9 is canonically isomorphic to $\mathsf{Ouv}(X)$.

(b) The notation Ouv stems from the French word 'ouvert' for 'open'.

21.1.3 Proposition *Let $(X, \mathcal{O})$ be a topological space. Then the category $\mathsf{Ouv}(X)$ has the following properties.*

(i) *The empty set is an initial object in $\mathsf{Ouv}(X)$, the total set X a final object.*

(ii) *Fibered products exist in $\mathsf{Ouv}(X)$. More precisely, if $i_{U,V} : V \hookrightarrow U$ and $i_{W,U} : W \hookrightarrow U$ are two morphisms in $\mathsf{Ouv}(X)$, the fibered product $V \times_U W$ is given by the open set $V \cap W$ together with the canonical embeddings $i_{V,V\cap W} : V \cap W \hookrightarrow V$ and $i_{W,V\cap W} : V \cap W \hookrightarrow W$.*

(iii) *Arbitrary (direct) limits exist in $\mathsf{Ouv}(X)$.*

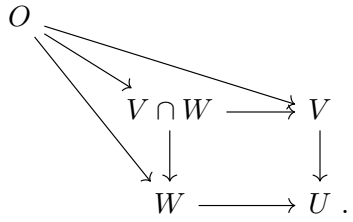
(iv) *Finite colimits exist in $\mathsf{Ouv}(X)$.*

Proof. ad (i). The first claim follows from the fact that $\emptyset$ is contained in every element of $\mathcal{O}$ and that every element of $\mathcal{O}$ is contained in X.

ad (ii). Assume to be given $O \in \mathcal{O}$ such that the following diagram commutes:

$$\begin{array}{ccc} O & \longrightarrow & V \\ \downarrow & & \downarrow \\ W & \longrightarrow & U \ . \end{array}$$

Then $O \subset V \cap W$, and the following diagram commutes with morphisms unique:



ad (*iii*). Assume that □

22. Basic Algebraic Topology

Part IV.

Commutative Algebra

40. The spectrum of a commutative ring

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 $k[x_1, \dots, x_n]$ (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).

40.1. 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 $\Bbbk[x_1, \dots, x_n]$ are of the form $(x_1 - a_1, \dots, x_n - a_n)$ for $(a_1, \dots, a_n) \in \Bbbk[x_1, \dots, x_n]$. 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 $I \subset \Bbbk[x_1, \dots, x_n]$ is an ideal, then the maximal ideals in $\Bbbk[x_1, \dots, x_n]$ correspond to points where everything in I vanishes. See 40.1.6 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.

40.1.1 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

$$V(I) = \{\mathfrak{p} : \mathfrak{p} \supset I\} \subset \operatorname{Spec} R.$$

40.1.2 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. $\emptyset = V((1))$ because no prime contains 1. So $\emptyset$ 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. □

40.1.3 Definition The topology on $\operatorname{Spec} R$ defined above is called the **Zariski topology**. With it, $\operatorname{Spec} R$ is now a topological space.

40.1.4 Remark What is the Spec of the zero ring?

In order to see the geometry of this construction, let us work several examples.

40.1.5 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 $(2), (3), (5), \ldots$, 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).

40.1.6 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,

   40.1.7 Theorem *The $\mathfrak{m}_{s,t}$ are all the maximal ideals in R.*

   This will follow from the *Hilbert Nullstellensatz* to be proved later (43.4.5).
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, $\{(z_1, z_2) : \mathfrak{Re}(z_1) \geq 0\}$ 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.

40.1.8 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 $\operatorname{Spec} R_2$.

40.1.9 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 **??**.

40.1.10 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. 40.1.6 above.

The map from ideals into closed sets is very far from being injective in general, though by definition it is surjective.

40.1.11 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.

40.1.12 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 11.2.5.)

Before proceeding, we must check:

40.1.13 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 $\geq n$, so every term belongs to I. Thus $(x+y)^{2n} \in I$ and, by definition, we see then that $x + y \in \operatorname{Rad}(I)$. □

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:

40.1.14 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:

40.1.15 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.

40.1.16 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 $\mathfrak{p} \cap \left\{1, x, x^2 \ldots, \right\} = \emptyset$. 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 $\{x^n, n \geq 0\}$.

More generally, we will prove:

40.1.17 Lemma *Let S be multiplicatively closed set in any ring R and let I be any ideal with $I \cap S = \emptyset$. 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 11.4.8.

Note that the lemma clearly implies the lemma when applied to $S = \{1, x, \ldots\}$.

Proof of the lemma. Let $P = \{J : J \supset I, J \cap S = \emptyset\}$. Then P is a poset with respect to inclusion. Note that $P \neq \emptyset$ 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 $\mathfrak{p} \cap S = \emptyset$. 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:

$$ab + ar'y + brx + rr'xy \in S.$$ □

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 $\mathfrak{p} \cap S = \emptyset$. □

The upshot of the previous lemmata is:

40.1.18 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 40.1.17) 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.

40.1.19 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 $I \in \mathcal{F}$.

40.1.20 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.

40.1.21 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.

40.1.22 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. □

40.1.23 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 40.1.22, this ideal is necessarily prime (since the family of finitely generated ideals is an Oka family). This contradicts the hypothesis. □

40.1.24 Corollary *If every prime ideal of R is principal, then every ideal of R is principal.*

Proof. This is proved in the same way. □

[2]Later we will say that R is *noetherian.*

40.1.25 Remark Suppose every nonzero prime ideal in R contains a non-zerodivisor. Then R is a domain. (Hint: consider the set S of nonzerodivisors, and argue that any ideal maximal with respect to not intersecting S is prime. Thus, (0) is prime.)

40.1.26 Remark Let R be a ring. Let κ be an infinite cardinal. By applying 40.1.20 and 40.1.22 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

$$R = k[\{x_n\}_{n\geq 1}, \{z_{t,n}\}_{t\in T, n\geq 0}]/(x_n^2, z_{t,n}^2, x_n z_{t,n} - z_{t,n-1})$$

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**.

40.1.27 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.

40.1.28 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 $\mathfrak{p} \in V(f(I))$.

40.1.29 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.

40.1.30 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.

40.1.31 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.

40.1.32 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. $U_0 = \emptyset$ 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 $S^{-1}R$.

40.1.33 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 $\phi^{-1}(\mathfrak{p})$.

40.1.34 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] 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 $x \in \mathfrak{p}$.

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
   $$\psi(U_{y/f^n}) = U_{fy} \subset \operatorname{Spec} R.$$
   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.

40.1.35 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 $S = S_0 \cup S_1 \cup \ldots$, where each S_k is generated by a finite number of elements $f_0, \ldots, f_k$. Then $R_{\mathfrak{p}} = \varinjlim S_k^{-1}R$. So we have
$$S^{-1}R = \varinjlim_k R[f_0^{-1}, f_1^{-1}, \ldots, f_k^{-1}] = \varinjlim R[(f_0 \ldots f_k)^{-1}].$$
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.

40.1.36 Remark Say that R is *semi-local* if it has finitely many maximal ideals. Let $\mathfrak{p}_1$, $\ldots$, $\mathfrak{p}_n \subset R$ be primes. The complement of the union, $S = R \setminus \bigcup \mathfrak{p}_i$, 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.

40.1.37 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 $\mu_{\mathbb{Z}}(\mathbb{Z}) = 1$.

40.1.38 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** □

40.2. 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.

40.2.1 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 $\operatorname{Rad}(R)$ (which is an abuse of notation).

The set of nilpotents is just the radical $\operatorname{Rad}((0))$ 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:

40.2.2 Example For any $n \geq 2$, 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:

40.2.3 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 40.1.16 to the zero ideal. □

We now consider a few examples of nilpotent elements.

40.2.4 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, $\operatorname{Rad}(R[x]) = (\operatorname{Rad}(R))R[x]$.

If $a_0, \dots, a_n$ 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.

40.2.5 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).

40.2.6 Example (Nilpotence in power series rings) Let R be a ring such that $\operatorname{Rad}(R)$ is a finitely generated ideal. (This is satisfied, e.g., if R is *noetherian*, cf. 41.) Let us consider the question of how $\operatorname{Rad}(R)$ and $\operatorname{Rad}(R[[x]])$ are related. The claim is that

$$\operatorname{Rad}(R[[x]]) = (\operatorname{Rad}(R))R[[x]].$$

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 $I = \operatorname{Rad}(R)$. There exists an $N > 0$ such that the ideal power $I^N = 0$ by finite generation. Thus if $f \in IR[[x]]$, then $f^N \in I^N R[[x]] = 0$.

40.2.7 Remark Prove that $x \in R$ is nilpotent if and only if the localization R_x is the zero ring.

40.2.8 Remark Construct an example where $\operatorname{Rad}(R)R[[x]] \neq \operatorname{Rad}(R[[x]])$. (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).

40.2.9 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.

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 $e \in R$. □

[4] Recall that an element $e \in R$ is idempotent if $e^2 = e$.
[5] Not necessarily commutative.

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:

40.2.10 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.

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:

40.2.11 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. □

40.2.12 Lemma *Suppose $V(X) = \operatorname{Spec} R$ for $X \subset R$ an ideal. Then every element of X is nilpotent.*

[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.

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

$$1 = (x+y)^{2k} = \sum_{i=0}^{2k} \binom{2k}{i} x^i y^{2k-i} = \sum\nolimits' + \sum\nolimits''$$

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:

$$I = (x'), \quad , J = (y').$$ □

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:

40.2.13 Proposition *Let A be a ring and I a nilpotent ideal. Then* $\operatorname{Idem}(A) \to \operatorname{Idem}(A/I)$ *is bijective.*

Proof. Indeed, the topological spaces of Spec A and Spec A/I are the same. The result then follows from **??**. □

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.

40.2.14 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.

40.3. Vista: sheaves on $\operatorname{Spec} R$

Presheaves

Let X be a topological space.

40.3.1 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** $\operatorname{res}^V_U : \mathcal{F}(V) \to \mathcal{F}(U)$. The restriction map is required to satisfy:

1. $\operatorname{res}^U_U = \operatorname{id}_{\mathcal{F}(U)}$ for all open sets U.
2. $\operatorname{res}^W_U = \operatorname{res}^V_U \circ \operatorname{res}^W_V$ 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.

40.3.2 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.

40.3.3 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.

40.3.4 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

40.3.5 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.

40.3.6 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.

40.3.7 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.

40.3.8 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:

40.3.9 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 $\mathcal{F} \to \mathcal{F}' \to \mathcal{G}$.

40.3.10 Theorem *We can construct the sheafification of a presheaf $\mathcal{F}$ as follows: $\mathcal{F}'(U) = \{s : U \to \coprod_{x\in U} \mathcal{F}_x | \text{for all } x \in U, s(x) \in \mathcal{F}_x \text{ and there is a neighborhood } V \subset U \text{ and } t \in \mathcal{F}(V) \text{ such that for all } y \in V, s(y) \text{ is the image of } t \text{ in the local ring } \mathcal{F}_y\}$.*

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.

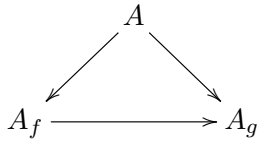
40.3.11 Proposition *Let A be a ring and let $X = \mathrm{Spec}(A)$. Then the assignment of the ring A_f to the basic open set X_f defines a presheaf of rings on X.*

Proof.

Part (i). If $X_g \subset X_f$ are basic open sets, then there exist $n \geq 1$ and $u \in A$ such that $g^n = uf$.

Proof of part (i). Let $S = \{g^n : n \geq 0\}$ and suppose $S \cap (f) = \emptyset$. 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 $\mathfrak{p} \cap S = \emptyset$. Since $(f)^e \in S^{-1}\mathfrak{p}$, we see that $f/1 \in S^{-1}\mathfrak{p}$, so $f \in \mathfrak{p}$. But $S \cap \mathfrak{p} = \emptyset$ implies that $g \notin \mathfrak{p}$. This is a contradiction, since then $\mathfrak{p} \in X_g \setminus X_f$.

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.



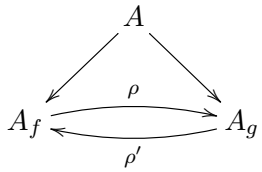
Proof of part (ii). Let $n \geq 1$ and $u \in A$ be such that $g^n = uf$ by part (i). Note that in A_g,

$$(f/1)(u/g^n) = (fu/g^n) = 1/1 = 1$$

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 $\rho : A_f \to A_g$.

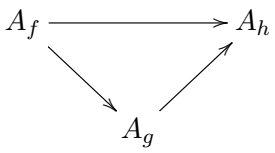
Part (iii). If $X_g = X_f$, then the corresponding restriction $\rho : A_f \to A_g$ is an isomorphism.

Proof of part (iii). The reverse inclusion yields a $\rho' : A_g \to A_f$ such that the diagram



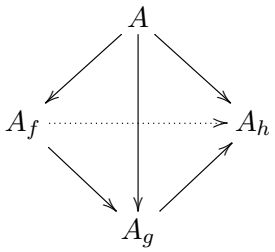
commutes. But since the localization map is epic, this implies that $\rho\rho' = \rho'\rho = \mathbf{1}$.

Part (iv). If $X_h \subset X_g \subset X_f$, then the diagram



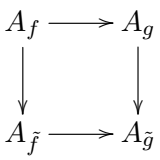
of restriction maps commutes.

Proof of part (iv). Consider the following tetrahedron.



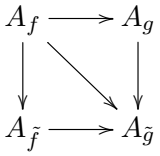
Except for the base, the commutativity of each face of the tetrahedron follows from the universal property of part (ii). But its easy to see that commutativity of the those faces implies commutativity of the base, which is what we want to show.

Part (v). If $X_{\tilde{g}} = X_g \subset X_f = X_{\tilde{f}}$, then the diagram



of restriction maps commutes. (Note that the vertical maps here are isomorphisms.)

Proof of part (v). By part (iv), the two triangles of

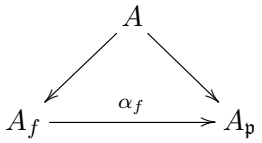


commute. Therefore the square commutes.

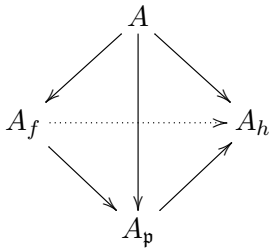
Part (vi). Fix a prime ideal $\mathfrak{p}$ in A. Consider the direct system consisting of rings A_f for every $f \notin \mathfrak{p}$ and restriction maps $\rho_{fg} : A_f \to A_g$ whenever $X_g \subset X_f$. Then $\varinjlim A_f \cong A_{\mathfrak{p}}$.

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 $m \geq 0$. 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.

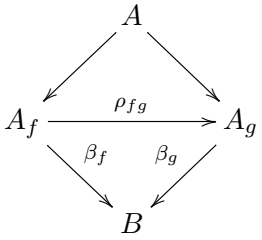


Then consider the following tetrahedron.



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.



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 $\varinjlim A_f \cong A_\mathfrak{p}$. □

40.3.12 Proposition *Let A be a ring and let $X = \mathrm{Spec}(A)$. 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

$$s = s \cdot 1 = s\left(\sum x_i g_i\right) = \sum x_i (s g_i) = 0.$$

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 $(s h_i b_i) h_i^{l_i} = 0$. But in fact, even $l_i = 0$ works:

$$s h_i - b_i = \left(\sum x_j b_j\right) h_i - b_i \left(\sum x_j h_j\right) = \sum x_j \left(h_i b_j - b_i h_j\right) = 0.$$

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. $\square$

41. 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.

41.1. Basics

The noetherian condition

41.1.1 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*.

41.1.2 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 $x_1, \ldots, x_n$. 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. □

Working with noetherian modules over non-noetherian rings can be a little funny, though, so normally this definition is combined with:

41.1.3 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:

41.1.4 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 41.1.7 in the next subsec. So the defining property of noetherian rings is that a submodule of a finitely generated module is finitely generated. (Compare 41.1.10.)

41.1.5 Remark The ring $\mathbb{C}[X_1, X_2, \ldots]$ 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.

41.1.6 Remark Let R be a ring such that every *prime* ideal is finitely generated. Then R is noetherian. See 40.1.23, 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.

41.1.7 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 $n \geq N$ 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.

41.1.8 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:

41.1.9 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.

41.1.10 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 41.1.7 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.

41.1.11 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 $m \in \mathbb{Z}_{\geq 0}$, let $I(n)$ be the collection of elements $a \in R$ consisting of the coefficients of x^n of elements of I of degree $\leq n$. 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 $n \leq N$ and $i \leq m_n$ 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 $\{P_{n,i}, n \leq N, i \leq m_n\}$. 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, $n \geq N$. 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 $P \in J$. □

41.1.12 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.

41.1.13 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 41.1.11 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. □

41.1.14 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.

41.1.15 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. □

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.

41.1.16 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.

41.1.17 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. □

41.2. 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.

41.2.1 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:

41.2.2 Proposition *$M = 0$ if and only if $\operatorname{supp} M = \emptyset$.*

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 $\operatorname{supp} M = \emptyset$. □

41.2.3 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 $\operatorname{Ann}(x)$ **for the annihilator of** $x \in M$**. This has not been defined before. Should we add this in a previous chapter?**

41.2.4 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 $\operatorname{Ass}(M)$.

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.

41.2.5 Remark Show that $\mathfrak{p} \in \operatorname{Ass}(M)$ if and only if there is an injection $R/\mathfrak{p} \hookrightarrow M$.

41.2.6 Remark Let $\mathfrak{p} \in \operatorname{Spec} R$. Then $\operatorname{Ass}(R/\mathfrak{p}) = \{\mathfrak{p}\}$.

41.2.7 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 $\operatorname{Ass}(M) \neq \emptyset$ for nonzero modules.

41.2.8 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 $I = \operatorname{Ann}(x)$ 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 $I \in \operatorname{Ass}(M)$. Indeed, suppose $ab \in I$ where $a, b \in R$. This means that

$$(ab)x = 0.$$

□

Consider the annihilator $\operatorname{Ann}(bx)$ 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, $\operatorname{Ann}(bx)$ contains $(a) + I$, which contains I. By maximality, it must happen that $\operatorname{Ann}(bx) = I$ and $a \in I$.

In either case, we find that one of a, b belongs to I, so that I is prime.

41.2.9 Example (A module with no associated prime) Without the noetherian hypothesis, 41.2.8 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

$$\operatorname{Ass}(R/I) = \emptyset.$$

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.

41.2.10 Proposition *If M is finitely generated, then $\operatorname{Ass}(M)$ 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:

41.2.11 Lemma *Suppose we have an exact sequence of finitely generated R-modules*

$$0 \to M' \to M \to M'' \to 0.$$

Then

$$\operatorname{Ass}(M') \subset \operatorname{Ass}(M) \subset \operatorname{Ass}(M') \cup \operatorname{Ass}(M'')$$

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 $\operatorname{Ass}(M') \subset \operatorname{Ass}(M)$.

The harder direction is the other inclusion. Suppose $\mathfrak{p} \in \operatorname{Ass}(M)$. Then there is $x \in M$ such that $\mathfrak{p} = \operatorname{Ann}(x)$. 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 $\operatorname{Ann}(y) = \mathfrak{p}$ and so $\mathfrak{p} \in \operatorname{Ass}(M')$. 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 $ab \in \mathfrak{p} = \operatorname{Ann}(x)$. But $y = ax \neq 0$, so $a \notin \mathfrak{p}$. As a result, the condition $b \in \operatorname{Ann}(y)$ is the same as $b \in \mathfrak{p}$. In other words,

$$\operatorname{Ann}(y) = \mathfrak{p}$$

which proves the claim.

Suppose now that $Rx \cap M' = 0$. Let $\phi : M \twoheadrightarrow M''$ be the surjection. I claim that $\mathfrak{p} = \operatorname{Ann}(\phi(x))$ and consequently that $\mathfrak{p} \in \operatorname{Ass}(M'')$. The proof is as follows. Clearly $\mathfrak{p}$ annihilates $\phi(x)$ as it annihilates x. Suppose $a \in \operatorname{Ann}(\phi(x))$. 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 $\operatorname{Ann}(\phi(x)) = \mathfrak{p}$. □

The next step in the proof of 41.2.10 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.

41.2.12 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 41.2.10. 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 41.2.11,

$$\operatorname{Ass}(M_{i+1}) \subset \operatorname{Ass}(M_i) \cup \operatorname{Ass}(R/\mathfrak{p}_i) = \operatorname{Ass}(M_i) \cup \{\mathfrak{p}_i\},$$

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 zerodivisors 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 zerodivisor on a finitely generated module M, then $\mathfrak{m} \in \operatorname{Ass}(M)$.

41.2.13 Proposition *If M is a finitely generated module over a noetherian ring R, then the zerodivisors on M are the union $\bigcup_{\mathfrak{p}\in\operatorname{Ass}(M)} \mathfrak{p}$.*

More strongly, if $I \subset R$ is any ideal consisting of zerodivisors 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 zerodivisors. 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 41.2.10.)

For the last statement, we use prime avoidance (theorem 11.4.20): if I consists of zerodivisors, then I is contained in the union $\bigcup_{\mathfrak{p}\in\operatorname{Ass}(M)} \mathfrak{p}$ by the first part of the proof. This is a finite union by **??**, so prime avoidance implies I is contained one of these primes. □

41.2.14 Remark For every module M over any (not necessarily noetherian) ring R, the set of M-zerodivisors$\mathcal{Z}(M)$ 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 $R \smallsetminus Z$ is a *saturated multiplicative set*. This is Kaplansky's Theorem 2.

41.2.15 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 $\operatorname{Ass}(M)$

It turns out to be extremely convenient that the construction $M \to \operatorname{Ass}(M)$ 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 $\operatorname{Ass}(M) \subset \operatorname{Spec} R$ and $\operatorname{Ass}(S^{-1}M) \subset \operatorname{Spec} S^{-1}R$. 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 $\operatorname{Ass}(M)$ and $\operatorname{Ass}(S^{-1}M)$, and one might conjecture (correctly, as it happens) that $\operatorname{Ass}(S^{-1}M) = \operatorname{Ass}(M) \cap \operatorname{Spec} S^{-1}R$.

41.2.16 Proposition *Let R noetherian, M finitely generated and $S \subset R$ multiplicatively closed. Then*

$$\operatorname{Ass}(S^{-1}M) = \left\{ S^{-1}\mathfrak{p} : \mathfrak{p} \in \operatorname{Ass}(M), \mathfrak{p} \cap S = \emptyset \right\}.$$

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 $\mathfrak{p} \cap S = \emptyset$. 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 $\mathrm{Ann}(t_0 tx)$. However, $I \subset \mathrm{Ann}(t_0 x)$ and I is maximal, so $I = \mathrm{Ann}(t_0 tx)$ 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. □

41.2.17 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.

41.2.18 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 $\operatorname{supp}(M)$ contains the set of primes $\mathfrak{p} \in \operatorname{Spec} R$ containing an associated prime; this will imply that $\operatorname{supp}(M) \supset \bigcup_{\mathfrak{q} \in \operatorname{Ass}(M)} \overline{\{\mathfrak{q}\}}$. So let $\mathfrak{q}$ be an associated prime and $\mathfrak{p} \supset \mathfrak{q}$. We need to show that

$$\mathfrak{p} \in \operatorname{supp} M, \text{ i.e. } M_{\mathfrak{p}} \neq 0.$$

But, since $\mathfrak{q} \in \operatorname{Ass}(M)$, 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 $\mathfrak{p} \in \operatorname{supp} M$.

Let us now prove the converse inclusion. Suppose that $\mathfrak{p} \in \operatorname{supp} M$. 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 41.2.16 that $\operatorname{Ass}(M) \cap \operatorname{Spec} R_{\mathfrak{p}}$ 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 $\operatorname{Ass}(M)$. This is precisely what we wanted to prove. □

41.2.19 Corollary *For M finitely generated, $\operatorname{supp} M$ is closed. Further, every minimal element of $\operatorname{supp} M$ lies in $\operatorname{Ass}(M)$.*

Proof. Indeed, the above result says that

$$\operatorname{supp} M = \bigcup_{\mathfrak{q} \in \operatorname{Ass}(M)} \overline{\{\mathfrak{q}\}}.$$

Since $\operatorname{Ass}(M)$ is finite, it follows that $\operatorname{supp} M$ is closed. The above equality also shows that any minimal element of $\operatorname{supp} M$ must be an associated prime. □

41.2.20 Example 41.2.19 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.

41.2.21 Corollary *The ring R has finitely many minimal prime ideals.*

Proof. Clearly, $\operatorname{supp} R = \operatorname{Spec} R$. Thus every prime ideal of R contains an associated prime of R by 41.2.18. □

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 $\operatorname{supp} M$ is a closed subset of $\operatorname{Spec} R$ and is a union of finitely many irreducible subsets. More precisely,

$$\operatorname{supp} M = \bigcup_{\mathfrak{q} \in \operatorname{Ass}(M)} \overline{\{\mathfrak{q}\}}$$

though there might be some redundancy in this expression. Some associated prime might be contained in others.

41.2.22 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

41.2.23 Remark It follows that in a noetherian ring, every minimal prime consists of zerodivisors. 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:

41.2.24 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 $R_x = 0$ (40.2.7 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 zerodivisor 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 zerodivisor.

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.

41.2.25 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.

41.2.26 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.

41.2.27 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 41.2.11. Since $M' \neq 0$, we also know that M' has an associated prime (41.2.8). Thus $\mathrm{Ass}(M') = \{\mathfrak{p}\}$, so M' is $\mathfrak{p}$-primary. □

Let us now return to the proof of the main result, 41.2.26. 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 $\mathrm{supp}\, R/I = \{\mathfrak{q} : \mathfrak{q} \supset I\}$ is the union of the closures of the associated primes. In this case,

$$\mathrm{supp}(R/I) = \{\mathfrak{q} : \mathfrak{q} \supset \mathfrak{p}\}.$$

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.

41.2.28 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 41.2.26. 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 41.2.26, 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.

41.2.29 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 $\operatorname{Rad}(I)$ is prime. This follows from 41.2.28.

41.2.30 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.

41.2.31 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.

41.2.32 Remark (Bourbaki) Let M be a finitely generated R-module. Let $T \subset \operatorname{Ass}(M)$ be a subset of the associated primes. Prove that there is a submodule $N \subset M$ such that

$$\operatorname{Ass}(N) = T, \quad \operatorname{Ass}(M/N) = \operatorname{Ass}(M) - T.$$

41.3. Primary decomposition

This is the structure theorem for modules over a noetherian ring, in some sense. Throughout, we fix a noetherian ring R.

Irreducible and coprimary modules

41.3.1 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.

41.3.2 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.

41.3.3 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.

41.3.4 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.

41.3.5 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 41.3.3 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. □

41.3.6 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.

41.3.7 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.□

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*.

41.3.8 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,

41.3.9 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.

41.3.10 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

$$\begin{array}{ccc} N_i & \longrightarrow & (N_i)_{\mathfrak{p}_i} = 0 \\ \downarrow & & \downarrow \\ M & \longrightarrow & M_{\mathfrak{p}_i}. \end{array}$$

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:

41.3.11 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.

41.3.12 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.

41.4. 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

41.4.1 Definition A commutative ring R is **Artinian** every descending chain of ideals $I_0 \supset I_1 \supset I_2 \supset \dots$ stabilizes.

41.4.2 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.

41.4.3 Remark A module is artinian if and only if every nonempty collection of submodules has a minimal element.

41.4.4 Remark A ring which is a finite-dimensional algebra over a field is artinian.

41.4.5 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.

41.4.6 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:

41.4.7 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 41.4.7:

41.4.8 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 nonzerodivisor, 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:

41.4.9 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

$$\mathfrak{m}_1, \mathfrak{m}_2, \dots$$

of distinct maximal ideals. Then we have the descending chain

$$R \supset \mathfrak{m}_1 \supset \mathfrak{m}_1 \cap \mathfrak{m}_2 \supset \dots.$$

This necessarily stabilizes. So for some n, we have that $\mathfrak{m}_1 \cap \dots \cap \mathfrak{m}_n \subset \mathfrak{m}_{n+1}$. However, this means that $\mathfrak{m}_{n+1}$ contains one of the $\mathfrak{m}_1, \dots, \mathfrak{m}_n$ 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.**

41.4.10 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:

41.4.11 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

$$R \supset J \supset J^2 \supset J^3 \supset \dots.$$

We want to show that this stabilizes at zero. A priori, we know that it stabilizes *somewhere.* For some n, we have

$$J^n = J^{n'}, \quad n' \geq n.$$

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:

41.4.12 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 41.4.11. 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 $\mathrm{Spec}\, R$; 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 $\mathrm{Spec}\, R$ 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:

41.4.13 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. □

41.4.14 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.

41.4.15 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** □

41.4.16 Remark If R is an artinian ring, and S is a finite R-algebra (finite as an R-module), then S is artinian.

41.4.17 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

41.4.18 Definition (von Neumann) An element $a \in R$ is called *von Neumann regular* if there is some $x \in R$ such that $a = axa$.

41.4.19 Definition (McCoy) A element $a \in R$ is *π-regular* if some power of a is von Neumann regular.

41.4.20 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.

41.4.21 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. □

41.4.22 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.

41.4.23 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.

42. 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

$$[z_0, \dots, z_n] = [\lambda z_0, \dots, \lambda z_n], \quad \lambda \in \mathbb{C}^*. \tag{42.0.0.1}$$

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 (42.0.0.1), 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 **??**.

42.1. 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

42.1.1 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 $m, n \in \mathbb{Z}_{\geq 0}$, 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.

42.1.2 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 $m \in \mathbb{Z}_{\geq 0}, n \in \mathbb{Z}$. 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:

42.1.3 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:

42.1.4 Example (An easy example) If R is a graded ring, then R is a graded module over itself.

42.1.5 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.

42.1.6 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.

42.1.7 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.

42.1.8 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.

42.1.9 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.

42.1.10 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 $\{I \cap R_i\}_{i \in \mathbb{Z}_{\geq 0}}$ 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. □

42.1.11 Example If $\mathfrak{a}, \mathfrak{b} \subset R$ are homogeneous ideals, then so is $\mathfrak{ab}$. This is clear from proposition 42.1.10.

42.1.12 Example Let k be a field. The ideal $(x^2 + y)$ in $k[x, y]$ is *not* homogeneous. However, we find from proposition 42.1.10 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.

42.1.13 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.

42.1.14 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 $\emptyset \subset \mathbb{P}^n_{\mathbb{C}}$.

42.1.15 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

$$\begin{aligned} R_n &= RS \cap R_n \text{ by assumption} \\ &= (R_0 \oplus R_1 \oplus \dots \oplus R_{n-1})(S) \cap R_n \text{ because } S \subset R_+ \\ &\subset (R_0[S])(S) \cap R_n \text{ by inductive hypothesis} \\ &\subset R_0(S). \end{aligned}$$

□

42.1.16 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 42.1.15, 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.

42.1.17 Definition Let R be a graded ring and $d \in \mathbb{N}$. We set $R^{(d)} = \bigoplus_{k \in \mathbb{Z}_{\geq 0}} R_{kd}$; 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.

42.1.18 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 $a_1, \ldots, a_p \geq 0$ 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 42.1.15. □

In particular, by proposition 43.1.12 (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 $R^{(d)}$.

42.1.19 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.

42.1.20 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 $r_1, \ldots, r_m$ 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 43.1.12 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:

42.1.21 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 42.1.20 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.

42.1.22 Lemma *Hypotheses as above, there is a pair (d, n_0) such that*

$$R_d M_n = M_{n+d}$$

for $n \geq n_0$.

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 42.1.18; 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 $\geq \frac{d}{d_j}$, we can write each term in the sum as some s_j times something in M_{n-d}.

In particular, $M_n = R_d M_{n-d}$. □

42.1.23 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 42.1.23 (replacing d, which we reserve for something else), and choose n_0 accordingly. So we have $R_{d'} R_m = R_{m+d'}$ whenever $m \geq n_0$. 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 $n \geq d$ 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.

42.1.24 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.

42.1.25 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 42.1.16. proposition 42.1.18 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.

42.1.26 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.

42.1.27 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. □

42.1.28 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 (42.1.25).

More generally, we can define the localization procedure for graded modules.

42.1.29 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 42.1.27.

42.1.30 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 42.1.27. 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:

42.1.31 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**.

42.1.32 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. $V(\sum \mathfrak{b}_i) = \bigcap V(\mathfrak{b}_i)$.
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 $\mathfrak{a}$.

42.1.33 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.

42.1.34 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.

42.1.35 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 42.1.13.

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. □

42.2. 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

42.2.1 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 $m, n \in \mathbb{Z}_{\geq 0}$. 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 $I_m = \bigoplus_{k \geq m} R_k$ (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.

42.2.2 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.

42.2.3 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.

42.2.4 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.

42.2.5 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

$$M \supset IM \supset I^2M \supset \dots,$$

which is also called the **I-adic filtration.**

The associated graded

We shall now describe a construction that produces graded things from filtered ones.

42.2.6 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.

42.2.7 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.

42.2.8 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.

42.2.9 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.

42.2.10 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.

42.2.11 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 $x \in R, n \in \mathbb{Z}_{\geq 0}$. 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!).

42.2.12 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.

42.2.13 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. □

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:

42.2.14 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.

42.2.15 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 $x \in M_n + N$. □

42.3. 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

42.3.1 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 42.1.6). 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:

42.3.2 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 42.1.22; 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 42.3.2 is finitely generated. We will show next that S is noetherian, and consequently submodules of finitely generated modules are finitely generated. Applying lemma 42.3.2 again, we will find that the induced filtration must be I-good.

42.3.3 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 41.1.13), 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.

42.3.4 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 (42.3.1). 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:

42.3.5 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:

42.3.6 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 $a_1, \ldots, a_k \in I$ be generators. For each n, the ideal I^n consists of the values of all homogeneous polynomials in $R[x_1, \ldots, x_k]$ of degree n evaluated on the tuple $(a_1, \ldots, a_k)$, as one may easily see.

It follows that if $b \in \bigcap I^n$, then for each n there is a polynomial $P_n \in R[x_1, \ldots, x_k]$ which is homogeneous of degree n and which satisfies

$$P_n(a_1, \ldots, a_k) = b.$$

The ideal generated by all the P_n in $R[x_1, \ldots, x_k]$ 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 $Q_i \in R[x_1, \ldots, x_k]$. 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 $(a_1, \ldots, a_k)$ we find

$$b = b(Q_1(a_1, \ldots, a_k) + \cdots + Q_{N-1}(a_1, \ldots, a_k)).$$

But the $Q_i(a_1, \ldots, a_k)$ 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. □

43. 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 **??**) 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.

43.1. Integrality

Fundamentals

As stated in the introduction to the chapter, integrality is a condition on rings parallel to that of algebraicity for field extensions.

43.1.1 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'.

43.1.2 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 43.1.5 below).

43.1.3 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.

43.1.4 Example Let R be a graded ring. Then the subring $R^{(d)} \subset R$ was defined in definition 42.1.17; 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 $x \in R'$.

43.1.5 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 $\{1\}, \{1, x\}, \left\{1, x, x^2\right\}, \ldots$ 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. □

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.

43.1.6 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.

43.1.7 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. □

43.1.8 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 $x \in R$.

43.1.9 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.

43.1.10 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 13.3.12).

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 43.1.1). 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.

43.1.11 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 $R[s_1, \ldots, s_n]$. As a result, the module $R[s_1, \ldots, s_n, t]$ is finitely generated over the ring $R' = R[s_1, \ldots, s_n]$. By the following proposition 43.1.12, R' is a finitely generated R-module. In particular, $R[s_1, \ldots, s_n, t]$ is a finitely generated R-module (not just a finitely generated R'-module).

[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.

Thus the R-module $R[s_1, \ldots, s_n, t]$ is a faithful R' module, finitely generated over R, which is preserved under multiplication by t. □

We now prove a result that can equivalently be phrased as "finite type plus integral implies finite" for a map of rings.

43.1.12 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 43.1.5. In general, we will have $R' = R[\alpha_1, \ldots, \alpha_n]$ for some $\alpha_i \in R'$. By the inductive hypothesis, $R[\alpha_1, \ldots, \alpha_{n-1}]$ is a finite R-module; by the case of one generator, R' is a finite $R[\alpha_1, \ldots, \alpha_{n-1}]$-module. This establishes the result by the next exercise. □

43.1.13 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.

43.1.14 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.

43.1.15 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.

43.1.16 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.

43.1.17 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

$$v_q(x + y) \geq \min(v_q(x), v_q(y)). \tag{43.1.17.1}$$

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

$$v_q(xy) = v_q(x) + v_q(y). \tag{43.1.17.2}$$

Now return to the proof that $\mathbb{Z}$ is normal. We would like to show that $v_q(a/b) \geq 0$. 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:

$$n v_q(a/b) \geq \min_{i>0} v_q(c_i (a/b)^{n-i}).$$

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 $v_q(x) \geq 0$ 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 (43.1.17.1) and (43.1.17.2) above such that R is the set of elements such that $v(x) \geq 0, v \in \mathcal{V}$ (along with 0), then R is integrally closed in K. We will talk more about this, and about valuation rings, below.

43.1.18 Example We saw earlier (example 43.1.2) 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.

43.1.19 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.

43.1.20 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:

43.1.21 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.

43.2. 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.

43.2.1 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

[3] It is, however, true that if R is *noetherian* (see Chapter 41) 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**.

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

$$\begin{array}{ccc} \operatorname{Spec}(R-\mathfrak{p})^{-1}R' & \longrightarrow & \operatorname{Spec} R_{\mathfrak{p}} \\ \downarrow & & \downarrow \\ \operatorname{Spec} R' & \longrightarrow & \operatorname{Spec} R \end{array}$$

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

$$\mathfrak{p}R' \neq R', \tag{43.2.1.1}$$

or that $\mathfrak{p}$ does not generate the unit ideal in R'. If we prove (43.2.1.1), 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 13.1.22). 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 (43.2.1.1) 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 (43.2.1.1) 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 43.1.12. 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. □

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.

43.2.2 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. □

43.2.3 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

$$x^n + r_1 x^{n-1} + \cdots + r_n = 0, \quad r_i \in R, \qquad \square$$

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.

43.2.4 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 43.2.1: 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 43.2.2, 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. □

43.2.5 Remark (exercise) Let k be a field. Show that $k[\mathbb{Q}_{\geq 0}]$ is integral over the polynomial ring $k[T]$. Although this is a *huge* extension, the prime ideal (T) lifts in only one way to $\operatorname{Spec} k[\mathbb{Q}_{\geq 0}]$.

43.2.6 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 43.2.1) 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'.

43.2.7 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 43.2.1), 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 43.2.1 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 $\mathfrak{q}_2$. □

43.3. 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

43.3.1 Definition A **valuation ring** is a domain R such that for every pair of elements $a, b \in R$, either $a \mid b$ or $b \mid a$.

43.3.2 Example $\mathbb{Z}$ is not a valuation ring. It is neither true that 2 divides 3 nor that 3 divides 2.

43.3.3 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.

43.3.4 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. □

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.

43.3.5 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 $x, y \in K$, $v(x + y) \geq \min(v(x), v(y))$.

Suppose that K is a field and $v : K \to A \cup \{\infty\}$ is a valuation (i.e. $v(0) = \infty$). Define $R = \{x \in K : v(x) \geq 0\}$.

43.3.6 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

$$v(x+y) \geq \min v(x), v(y),$$

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 $v(x) \geq 0$ or $v(x^{-1}) \geq 0$ since A is totally ordered.[4] So either $x, x^{-1} \in R$. □

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.

43.3.7 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 $x \leq y$ if $y/x \in R$—this doesn't depend on the representatives in K^* chosen. Note that either $x \leq y$ or $y \leq x$ must hold, since R is a valuation ring. The combination of $x \leq y$ and $y \leq x$ 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 $v(x), v(y)$. □

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.

43.3.8 Definition A valuation ring R is **discrete** if we can choose A to be $\mathbb{Z}$.

43.3.9 Example $\mathbb{Z}_{(p)}$ is a discrete valuation ring.

The notion of a valuation ring is a useful one.

[4] Otherwise $0 = v(x) + v(x^{-1}) < 0$, contradiction.

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.

43.3.10 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.

43.3.11 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:

43.3.12 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 $v(1) \geq \min v(x), v(1 - x)$ 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:

43.3.13 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.

43.3.14 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.

43.3.15 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.

43.3.16 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.

43.3.17 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:

43.3.18 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.

43.3.19 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.

43.3.20 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 $v(x) \geq 0$ 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. □

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.

43.4. 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.

43.4.1 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 43.4.1 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 $a_i \in \mathbb{C}$ $(1 \leq i \leq n)$ 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:

43.4.2 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:

43.4.3 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:

43.4.4 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 $1 \leq i \leq m$,

$$x_i' = x_i + x_n^{N^i}.$$

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

$$f(\left\{x_i' - x_n^{N_i}\right\}, x_n)$$

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

$$x_n^{a_m N^m + a_{m-1} N^{m-1} + \dots + b}.$$

We can't, however, get any exponents of x_n in the expression $f(\left\{x_i' - x_n^{N_i}\right\}, x_n)$ 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 $k[x'_1, \ldots, x'_m]$.

We have thus proved the normalization lemma in the codimension one case. What about the general case? We repeat this. Say we have

$$k[x_1, \ldots, x_m] \subset R.$$

Let R' be the subring of R generated by $x_1, \ldots, x_m, x_{m+1}$. The argument we just gave implies that we can choose $x'_1, \ldots, x'_m$ such that R' is integral over $k[x'_1, \ldots, x'_m]$, and the x'_i are algebraically independent. We know in fact that $R' = k[x'_1, \ldots, x'_m, x_{m+1}]$.

Let us try repeating the argument while thinking about x_{m+2}. Let $R'' = k[x'_1, \ldots, x'_m, x_{m+2}]$ 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 $x''_1, \ldots, x''_m$ are algebraically independent and R'' is integral over $k[x''_1, \ldots, x''_m]$. We have furthermore that $k[x''_1, \ldots, x''_m, x_{m+2}] = R''$.

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 $k[x''_1, \ldots, x''_m]$.

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 $k[x''_1, \ldots, x''_m]$ in R thus contains

$$k[x''_1, \ldots, x''_m, x_{m+2}] = R''. \qquad \square$$

However, R'' contains the elements $x'_1, \ldots, x'_m$. But by construction, x_{m+1} is integral over the $x'_1, \ldots, x'_m$. The integral closure of $k[x''_1, \ldots, x''_m]$ 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 $k[x_1, \ldots, x_m]$ for some m. If $m > 0$, then this polynomial ring has more than one prime. For instance, (0) and $(x_1, \ldots, x_m)$. 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

$$\operatorname{Spec} R \to \operatorname{Spec} k[x_1, \ldots, x_m]$$

is surjective. If R is a field, this means $\operatorname{Spec} k[x_1, \ldots, x_m]$ 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:

43.4.5 Theorem *Let $I \subset \mathbb{C}[x_1, \ldots, x_n]$. 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 $\mathrm{Rad}(I)$ *is precisely the collection of f such that $f|_V = 0$. In particular,*

$$\mathrm{Rad}(I) = \bigcap_{\mathfrak{m} \supset I, \mathfrak{m} \text{ maximal}} \mathfrak{m}.$$

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

$$\mathrm{Rad}(I) = \bigcap_{\mathfrak{p} \supset I, \mathfrak{p} \text{ prime}} \mathfrak{p},$$

true in any commutative ring. However, this general result is not necessarily true.

43.4.6 Example The intersection of all primes in a DVR is zero, but the intersection of all maximals is nonzero.

Proof of theorem 43.4.5. It now suffices to show that for every $\mathfrak{p} \subset \mathbb{C}[x_1, \ldots, x_n]$ prime, we have

$$\mathfrak{p} = \bigcap_{\mathfrak{m} \supset I \text{ maximal}} \mathfrak{m}$$

since every radical ideal is an intersection of primes.

Let $R = \mathbb{C}[x_1, \ldots, x_n]/\mathfrak{p}$. 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. □

43.4.7 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 $\operatorname{Spec}_{\mathrm{m}} A$ denotes the set of maximal ideals in A. Consider the natural k-algebra structure on $\operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} A, k)$. We have a map

$$A \to \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} A, k)$$

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 $\to a = 0$.

43.4.8 Definition A function $f \in \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} A, k)$ 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 $\Phi : \operatorname{Spec}_{\mathrm{m}} B \to \operatorname{Spec}_{\mathrm{m}} A$ given by taking pre-images.

43.4.9 Definition A map $\Phi : \operatorname{Spec}_{\mathrm{m}} B \to \operatorname{Spec}_{\mathrm{m}} A$ 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.

43.4.10 Proposition *A map $\Phi : \operatorname{Spec}_{\mathrm{m}} B \to \operatorname{Spec}_{\mathrm{m}} A$ is algebraic if and only if for any algebraic function $f \in \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} A, k)$, the pullback $f \circ \Phi \in \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} B, k)$ is algebraic.*

Proof. Suppose that Φ is algebraic. It suffices to check that the following diagram is commutative:

$$\begin{array}{ccc} \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} A, k) & \xrightarrow{-\circ\Phi} & \operatorname{Funct}(\operatorname{Spec}_{\mathrm{m}} B, k) \\ \uparrow & & \uparrow \\ A & \xrightarrow[\phi]{} & B \end{array}$$

where $\phi : A \to B$ is the map that gives rise to Φ.

[⇐] Suppose that for all algebraic functions $f \in \mathrm{Funct}(\mathrm{Spec}_{\mathrm{m}} A, k)$, the pull-back $f \circ \Phi$ is algebraic. Then we have an induced map, obtained by chasing the diagram counterclockwise:

$$\begin{array}{ccc} \mathrm{Funct}(\mathrm{Spec}_{\mathrm{m}} A, k) & \xrightarrow{-\circ\Phi} & \mathrm{Funct}(\mathrm{Spec}_{\mathrm{m}} B, k) \\ \uparrow & & \uparrow \\ A & \xrightarrow[\phi]{} & B \end{array}$$

From ϕ, we can construct the map $\Phi' : \mathrm{Spec}_{\mathrm{m}} B \to \mathrm{Spec}_{\mathrm{m}} A$ given by $\Phi'(\mathfrak{m}) = \phi^{-1}(\mathfrak{m})$. I claim that $\Phi = \Phi'$. If not, then for some $\mathfrak{m} \in \mathrm{Spec}_{\mathrm{m}} B$ we have $\Phi(\mathfrak{m}) \neq \Phi'(\mathfrak{m})$. By definition, for all algebraic functions $f \in \mathrm{Funct}(\mathrm{Spec}_{\mathrm{m}} A, k)$, $f \circ \Phi = f \circ \Phi'$ so to arrive at a contradiction we show the following lemma:
Given any two distinct points in $\mathrm{Spec}_{\mathrm{m}} A = V(I) \subset k^n$, 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. □

43.5. 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:

43.5.1 Definition A ring R is **reduced** if it has no nonzero nilpotents.

43.5.2 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

$$\begin{array}{ccc} R & \longrightarrow & \prod M_i \\ \downarrow & & \downarrow \\ \prod R_{\mathfrak{p}_i} & \longrightarrow & \prod (M_i)_{\mathfrak{p}_i} \end{array}$$

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:

43.5.3 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.*

43.5.4 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:

43.5.5 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.

43.5.6 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 nonzerodivisors 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:

43.5.7 Proposition *Let R be noetherian. The set of nonzerodivisors 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 zerodivisor.

Conversely, suppose x is a zerodivisor, 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:

43.5.8 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 zerodivisor 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

$$\begin{array}{ccc} M & \longrightarrow & \prod M_i \\ \downarrow & & \downarrow \\ \prod M_{\mathfrak{p}_i} & \longrightarrow & \prod (M_i)_{\mathfrak{p}_i} \end{array} .$$

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

$$S^{-1}M \to \prod_{\mathfrak{p} \in \mathrm{Ass}(M)} M_{\mathfrak{p}}.$$

43.5.9 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

43.5.10 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 nonzerodivisors 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:

43.5.11 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.

43.5.12 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. □

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:

43.5.13 Theorem *If R is a noetherian domain*

$$R = \bigcap_{\mathfrak{p} \in \mathrm{Ass}(R/s), s \in R-0} R_{\mathfrak{p}}$$

[5] In general, this would be equivalent to $ty \in tsM$ for some $t \in S$; but S consists of nonzerodivisors on M.

Serre's criterion

We can now state a result.

43.5.14 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.

43.5.15 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 43.5.13 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.

43.5.16 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. □

144. 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 13.4.7) 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!

144.1. Faithful flatness

Faithfully flat modules

Let R be a commutative ring.

144.1.1 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.

144.1.2 Example The direct sum of faithfully flat modules is faithfully flat.

144.1.3 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.

144.1.4 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. □

144.1.5 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.

144.1.6 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.

144.1.7 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 144.1.4, 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 $N \in \mathbb{Z}_{\geq 0}$.)

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."

144.1.8 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

$$\mathrm{im}(f) \otimes_R M = \mathrm{im}(f \otimes 1_M)$$

so that if $f \otimes 1_M = 0$, then $\mathrm{im}(f) \otimes M = 0$. Thus by faithful flatness, $\mathrm{im}(f) = 0$ by Proposition 144.1.4.

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 144.1.4, we are done again. □

144.1.9 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:

144.1.10 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 144.1.4. 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 144.1.4 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 144.1.4, it follows that M is faithfully flat □

144.1.11 Corollary *A nonzero finitely generated flat module over a* local *ring is faithfully flat.*

Proof. This follows from proposition 144.1.10 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.□

144.1.12 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.

144.1.13 Definition S is a **faithfully flat R-algebra** if it is faithfully flat as an R-module.

144.1.14 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:

144.1.15 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.

144.1.16 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. □

144.1.17 Example The converse of proposition 144.1.16 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).

144.1.18 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:

144.1.19 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 144.1.10 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.

144.1.20 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 144.1.19. 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 144.1.15, $S_\mathfrak{q}$ 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.

144.1.21 Remark (exercise) Let $R \to S$ be a faithfully flat morphism of rings. If S is noetherian, so is R. The converse is false!

144.1.22 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.

144.1.23 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:

144.1.24 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:

144.1.25 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

$$\phi(R - \mathfrak{p}) \cap \phi(\mathfrak{p})S = \emptyset.$$

Now $T = \phi(R - \mathfrak{p})$ is a multiplicatively closed subset. There is a morphism

$$R_{\mathfrak{p}} \to T^{-1}S \tag{144.1.25.1}$$

which sends elements of $\mathfrak{p}$ into non-units, by (144.1.25.1) 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 $\operatorname{Spec} S$. □

144.1.26 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 144.1.25 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 144.1.10.

144.1.27 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

$$\begin{array}{ccc} \operatorname{Spec} S/\mathfrak{p}S & \longrightarrow & \operatorname{Spec} R/\mathfrak{p} \\ \downarrow & & \downarrow \\ \operatorname{Spec} S & \longrightarrow & \operatorname{Spec} R \end{array}$$

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.

144.1.28 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** □

144.2. 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

$$d_i : B^{\otimes n} \to B^{\otimes n+1}, \quad 0 \leq i \leq n+1$$

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.

144.2.1 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:

144.2.2 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:

144.2.3 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

$$b_1 \otimes \cdots \otimes b_{n+1} \mapsto s(b_{n+1}) \left(b_1 \otimes \cdots \otimes b_n\right).$$

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. □

144.2.4 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] 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 $b \otimes m \otimes b'$.

144.2.5 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:

(144.2.5.1)
$$\begin{array}{ccc} B \otimes_A B \otimes_A N & \xrightarrow{\phi_{23}} & B \otimes_A N \otimes_A B \\ & {\scriptstyle \phi_{13}} \searrow \quad \swarrow {\scriptstyle \phi_{12}} & \\ & N \otimes_A B \otimes_A B & \end{array}$$

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

(144.2.5.2)
$$\begin{array}{ccc} N \otimes_A B & \xrightarrow{\phi} & B \otimes_A N \\ \downarrow {\scriptstyle f \otimes 1} & & \downarrow {\scriptstyle 1 \otimes f} \\ N' \otimes_A B & \xrightarrow{\psi} & B \otimes_A N' \end{array}$$

As we have seen, there is a functor F from A-modules to descent data. Strictly speaking, we should check the commutativity of (144.2.5.1), but this is clear: for $N = M \otimes_A B$, (144.2.5.1) looks like

$$\begin{array}{ccc} B \otimes_A B \otimes_A M \otimes_A B & \xrightarrow{\phi_{23}} & B \otimes_A M \otimes_A B \otimes_A B \\ & {\scriptstyle \phi_{13}} \searrow \quad \swarrow {\scriptstyle \phi_{12}} & \\ & M \otimes_A B \otimes_A B \otimes_A B & \end{array}$$

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:

144.2.6 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.

[2] This is the cocycle condition.

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 (144.2.5.2) commutes. The diagram looks like

$$\begin{array}{ccc} M_B \otimes_A B & \longrightarrow & B \otimes_A M_B \\ \downarrow{\scriptstyle f \otimes 1} & & \downarrow{\scriptstyle 1 \otimes f} \\ N \otimes_A B & \xrightarrow{\phi} & B \otimes_A N \end{array}.$$

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:

$$\begin{array}{ccccccc} 0 & \longrightarrow & M \otimes_A B & \longrightarrow & N \otimes_A B & \longrightarrow & N \otimes_A B \otimes_A B\ . \\ & & \downarrow & & \downarrow \phi & & \downarrow \phi_{13}^{-1} \\ 0 & \longrightarrow & N & \longrightarrow & B \otimes_A N & \longrightarrow & B \otimes_A B \otimes_A N \end{array}$$

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

144.3. 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:

144.3.1 Definition The derived functors of the tensor product functor $N \mapsto N \otimes_R M$ are denoted by $\mathrm{Tor}_R^i(N, M), i \geq 0$. We shall sometimes denote omit the subscript R.

So in particular, $\mathrm{Tor}_R^0(M,N) = M \otimes N$. 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, $\mathrm{Tor}_R^i(M,N) \simeq \mathrm{Tor}_R^i(N,M)$ 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 $\mathrm{Tor}_R^i(M,N)$ is the ith homology of the complex $M \otimes P_*$. We can use this to compute Tor in the case of abelian groups.

144.3.2 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

$$\begin{aligned}
\mathrm{Tor}_0^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong A/dA \\
\mathrm{Tor}_n^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong {}_dA/(m/d)A \quad n \text{ odd}, n \geq 1 \\
\mathrm{Tor}_n^{\mathbb{Z}/m\mathbb{Z}}(A, \mathbb{Z}/d\mathbb{Z}) &\cong {}_{m/d}A/dA \quad n \text{ even}, n \geq 2,
\end{aligned}$$

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.

144.3.3 Proposition *Let M be an R-module, $\{N_i\}$ a filtered system of R-modules. Then the natural morphism*

$$\varinjlim_i \operatorname{Tor}_R^i(M, N_i) \to \operatorname{Tor}_R^i(M, \varinjlim_i N_i)$$

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 $i \geq 1$ (whatever be N). Indeed, recall that $\operatorname{Tor}(M, N)$ 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:

144.3.4 Proposition *M is flat iff $\operatorname{Tor}^1(M, R/I) = 0$ for all finitely generated ideals $I \subset R$.*

Proof. We have just seen one direction. Conversely, suppose $\operatorname{Tor}^i(M, R/I) = 0$ for all finitely generated ideals I and $i > 0$. Then the result holds, first of all, for all ideals I, because of proposition 144.3.3 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 $\operatorname{Tor}^i(M, N) = 0$ 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 $\operatorname{Tor}^i(M, N) = 0$ for all $i \geq 1$.

Thus we see that $\operatorname{Tor}^i(M, N) = 0$ whenever N is finitely generated. Since any module is a filtered colimit of finitely generated ones, we are done by proposition 144.3.3. □

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:

144.3.5 Corollary *If the map*

$$I \otimes M \to M$$

is injective for all ideals I, then M is flat.

144.4. 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:

144.4.1 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.

144.4.2 Theorem *The finitely generated S-module M is flat over R iff*

$$\mathrm{Tor}_R^1(k,M) = 0.$$

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

$$\operatorname{Tor}_R^1(N, M) = 0. \tag{144.4.2.1}$$

This is because N has by dévissage (proposition 41.2.12) 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 $\operatorname{Tor}^1(N_i, M) = 0$ for each i.

Indeed, the claim (144.4.2.1) is true $N_0 = 0 \subset N$ trivially. We climb up the filtration piece by piece inductively; if $\operatorname{Tor}_R^1(N_i, M) = 0$, then the exact sequence

$$0 \to N_i \to N_{i+1} \to k \to 0$$

yields an exact sequence

$$\operatorname{Tor}_R^1(N_i, M) \to \operatorname{Tor}_R^1(N_{i+1}, M) \to 0$$

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 $\operatorname{Tor}_R^1(M, R/(I + \mathfrak{m}^t)) = 0$ by (144.4.2.1), as $R/(I + \mathfrak{m}^t)$ is of finite length.

Let us draw the following commutative diagram:

(144.4.2.2)

$$\begin{array}{ccccc} & & 0 & & \\ & & \downarrow & & \\ \mathfrak{m}^t \cap I \otimes M & \longrightarrow & I \otimes M & \longrightarrow & I/I \cap \mathfrak{m}^t \otimes M \\ & & & & \downarrow \\ & & & & M/\mathfrak{m}^t M \end{array}$$

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

144.4.3 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 (144.4.2.2)

$$\begin{array}{ccccc} & & 0 & & \\ & & \downarrow & & \\ \mathfrak{m}^t \cap I \otimes M & \longrightarrow & I \otimes M & \longrightarrow & I/I \cap \mathfrak{m}^t \otimes M \\ & & & & \downarrow \\ & & & & M/\mathfrak{m}^t M \end{array} .$$

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.

144.4.4 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:

144.4.5 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 144.4.2. 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 144.4.2 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 144.4.5, 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.

144.4.6 Remark One situation where the strong form of the local criterion, theorem 144.4.5, 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 $\operatorname{Spec} B$).

144.4.7 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 $\operatorname{Tor}_1^R(M, S) = 0$.*
2. *$\operatorname{Tor}_1^R(M, N) = 0$ 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

$$\operatorname{Tor}_p^S(\operatorname{Tor}_q^R(Q, S), P) \implies \operatorname{Tor}_{p+q}^R(Q, P).$$

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 $\operatorname{Tor}_1^R(M, S) = 0$. We want to show that 2 holds, so let N be any S-module. Consider the E_2 page of the above spectral sequence $\operatorname{Tor}_p^S(\operatorname{Tor}_q^R(M, S), N) \implies \operatorname{Tor}_{p+q}^R(M, N)$. In the terms such that $p + q = 1$, we have the two terms $\operatorname{Tor}_0^S(\operatorname{Tor}_1^R(M, S), N), \operatorname{Tor}_1^S(\operatorname{Tor}_0^R(M, S), N)$. But by hypotheses these are both zero. It follows that $\operatorname{Tor}_1^R(M, N) = 0$.
2. Suppose $\operatorname{Tor}_1^R(M, N) = 0$ 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 $\operatorname{Tor}_1^S(M \otimes_R S, N) = 0$ for each S-module N. It follows that $M \otimes_R S$ is S-flat. Hence the higher terms $\operatorname{Tor}_p^S(M \otimes_R S, N) = 0$ 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 $\operatorname{Tor}_1^R(M, S) \otimes_S N = 0$ for every S-module N, which clearly implies $\operatorname{Tor}_1^R(M, S) = 0$. □

As a result, we shall be able to deduce the result alluded to in the motivation following the statement of theorem 144.4.5.

144.4.8 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 144.4.7, once one notes that any N as in the statement admits a finite filtration whose successive quotients are annihilated by J. □

Proof of theorem 144.4.5. 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.

144.4.9 Lemma *Let the hypotheses be as in theorem 144.4.5 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 144.4.9 and the argument in theorem 144.4.3. (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 144.4.10. 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 **?**:

144.4.10 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 $\operatorname{Tor}_1^A(M, A/I) = 0$.*
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

$$M/IM \otimes_{A/I} I^n/I^{n+1} \to I^nM/I^{n+1}M \tag{144.4.10.1}$$

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 $\operatorname{Tor}_1^A(M, A/I) = 0$ (by the local criterion). Then we are, finally, reduced to showing:

144.4.11 Lemma *Let A be a ring, $I \subset A$ be a nilpotent ideal, and M any A-module. If (144.4.10.1) is an isomorphism for each n, then $\operatorname{Tor}_1^A(M, A/I) = 0$.*

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.

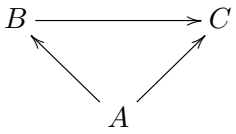
Suppose $I^{n+1} \otimes_A M \to I^{n+1}M$ is an isomorphism. Consider the diagram

$$\begin{array}{ccccccc} & & I^{n+1} \otimes_A M & \longrightarrow & I^n \otimes_A M & \longrightarrow & I^n/I^{n+1} \otimes_A M \to 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & I^{n+1}M & \longrightarrow & I^nM & \longrightarrow & I^nM/I^{n+1}M \longrightarrow 0. \end{array}$$

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:

144.4.12 Proposition *Let $(A, \mathfrak{m}), (B, \mathfrak{n}), (C, \mathfrak{n}')$ be noetherian local rings. Suppose given a commutative diagram of local homomorphisms*



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

$$I^n/I^{n+1} \otimes_{B/I} C/IC \to I^n C/I^{n+1}C \tag{144.4.12.1}$$

is an isomorphism. By theorem 144.4.10, this will do it. Now we have:

$$\begin{aligned} I^n/I^{n+1} \otimes_{B/I} C/IC &\simeq \mathfrak{m}^n B/\mathfrak{m}^{n+1}B \otimes_{B/\mathfrak{m}B} C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A B/\mathfrak{m}B \otimes_B C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A B \otimes_B C/\mathfrak{m}C \\ &\simeq (\mathfrak{m}^n/\mathfrak{m}^{n+1}) \otimes_A C/\mathfrak{m}C \\ &\simeq \mathfrak{m}^n C/\mathfrak{m}^{n+1}C \simeq I^n C/I^{n+1}C. \end{aligned}$$

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 (144.4.12.1) 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.

144.4.13 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* $\leq$ 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

$$\dim B/(x) \leq \dim A/(x) + \dim B/\mathfrak{m}B$$

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, $\dim B/(x) \leq \dim B - 1$. 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 nonzerodivisor in B.

In other words, we have found $x \in A$ which is both A-regular and B-regular (i.e. nonzerodivisors 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. □

144.4.14 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 144.4.10). 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.

144.4.15 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 144.4.16. □

We shall need the following lemma to see that we preserve noetherianness when we pass to the limit.

144.4.16 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 **??**.

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 **??** 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.

144.4.17 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 144.1.24. 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!

144.4.18 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 41.2.12), 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 $t_1, \ldots, t_d$, which are algebraically independent over $K(R)$ and such that $T \otimes_R K(R)$ is integral over $K(R)[t_1, \ldots, t_d]$. (Here d is the transcendence degree of $K(T)/K(R)$.) If we localize at some highly divisible element, we can assume that $t_1, \ldots, t_d$ 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 $R[t_1, \ldots, t_d] \subset T$ is a polynomial ring. Moreover, each of the finitely many generators of T/R satisfies a monic polynomial equation over $K(R)[t_1, \ldots, t_d]$ (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 $R[t_1, \ldots, t_d]$. We have thus reduced to the following case. T is a finitely generated domain over R, *integral* over the polynomial ring $R[t_1, \ldots, t_d]$. In particular, it is a finitely generated module over the polynomial ring $R[t_1, \ldots, t_d]$. Thus we have some r and an exact sequence

$$0 \to R[t_1, \ldots, t_d]^r \to T \to Q \to 0,$$

where Q is a torsion $R[t_1, \ldots, t_d]^r$-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. □

Part V.

Homological Algebra

50. Homological algebra à la Cartan–Eilenberg

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).

50.1. (Co)Chain complexes and their (co) homology

Chain complexes

The chain complex is the most fundamental construction in homological algebra.

50.1.1 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 $k \leq n$ (respectively $C_k = 0$ for all $k \geq n$). 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.

50.1.2 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.

50.1.3 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.

50.1.4 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*.

50.1.5 Remark In general, a chain complex need not be exact, and this failure of exactness is measured by its homology.

50.1.6 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.

50.1.7 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

$$\begin{array}{ccccccccc}
\dashrightarrow & C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \dashrightarrow \\
 & \downarrow{\scriptstyle f_{k+1}} & & \downarrow{\scriptstyle f_k} & & \downarrow{\scriptstyle f_{k-1}} & \\
\dashrightarrow & D_{k+1} & \xrightarrow[\delta_{k+1}]{} & D_k & \xrightarrow[\delta_k]{} & D_{k-1} & \dashrightarrow
\end{array}$$

commutes. We will denote such a morphism of chain complexes by $f : (C_\bullet, \partial) \to (D_\bullet, \delta)$.

50.1.8 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.

50.1.9 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 $\mathsf{Ch}^+_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^-_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^{\mathrm{b}}_\bullet(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^{\geq 0}_\bullet(R\text{-}\mathsf{Mod})$, and $\mathsf{Ch}^{\leq 0}_\bullet(R\text{-}\mathsf{Mod})$, 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. □

50.1.10 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:

$$\begin{array}{ccccccccc}
\dashrightarrow & C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \dashrightarrow \\
 & \downarrow{\scriptstyle f_{k+1}} & & \downarrow{\scriptstyle f_k} & & \downarrow{\scriptstyle f_{k-1}} & \\
\dashrightarrow & D_{k+1} & \xrightarrow[\delta_{k+1}]{} & D_k & \xrightarrow[\delta_k]{} & D_{k-1} & \dashrightarrow
\end{array} \quad .$$

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. □

Long exact sequences

add: OMG! We have all this and not the most basic theorem of them all.

50.1.11 Definition If M is a complex then for any integer k, we define a new complex $M[k]$ by shifting indices, i.e. $(M[k])^i := M^{i+k}$.

50.1.12 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.

50.1.13 Proposition *A map $f : M \to N$ is a quasi-isomorphism if and only if* $\mathrm{Cone}(f)$ *is acyclic.*

50.1.14 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

$$\cdots \to H^{i-1}(\mathrm{Cone}(f)) \to H^i(M) \to H^i(N) \to H^i(\mathrm{Cone}(f)) \to \dots$$

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.

50.1.15 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 $k \leq n$ (respectively $C^k = 0$ for all $k \geq n$). 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:

$$\begin{array}{ccccccc} \dashrightarrow & C^{k-1} & \xrightarrow{d^{k-1}} & C^k & \xrightarrow{d^k} & C^{k+1} & \dashrightarrow \\ & \downarrow{\scriptstyle g^{k-1}} & & \downarrow{\scriptstyle g^k} & & \downarrow{\scriptstyle g^{k+1}} & \\ \dashrightarrow & D^{k-1} & \xrightarrow[\delta^{k-1}]{} & D^k & \xrightarrow[\delta^k]{} & D^{k-1} & \dashrightarrow \end{array} .$$

We will denote such a morphism of cochain complexes usually by $g : (C^\bullet, d) \to (D^\bullet, \delta)$.

50.1.16 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* $\mathsf{Ch}^\bullet_+(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_-(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_{\mathrm{b}}(R\text{-}\mathsf{Mod})$, $\mathsf{Ch}^\bullet_{\geq 0}(R\text{-}\mathsf{Mod})$, *and* $\mathsf{Ch}^\bullet_{\leq 0}(R\text{-}\mathsf{Mod})$, *respectively.*

Proof. The proof is completely dual to the proof of Proposition 50.1.16. □

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.

50.1.17 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).

50.2. Chain Homotopies

50.2.1 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.

50.2.2 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 $f, g : C^bullet \to D^bullet$ 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 $f - g = dh + hd$.

50.2.3 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 $H^i(D)$. □

50.2.4 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. □

50.2.5 Example Not every quasi-isomorphism is a homotopy equivalence. Consider the complex

$$\cdots \to 0 \to \mathbb{Z}/\cdot 2 \to \mathbb{Z} \to 0 \to 0 \to \dots$$

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

$$\cdots \to 0 \to 0 \to \mathbb{Z}/2\mathbb{Z} \to 0 \to 0 \to \dots$$

but no inverse can be defined (no map from $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$).

50.2.6 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."

50.2.7 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.

50.3. 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 $\bigoplus_k H_k(C_\bullet)$.

50.3.1 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.

50.3.2 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

$$\begin{array}{ccc} M & \xrightarrow{d} & M \\ \downarrow & & \downarrow \\ N & \xrightarrow{d'} & N \end{array}$$

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.

50.3.3 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.

50.4. Derived functors

Projective resolutions

Fix a ring R. Let us recall (13.2.7) 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.

50.4.1 Definition Let M be an arbitrary module, a projective resolution of M is an exact sequence

$$\cdots \to P_i \to P_{i-1} \to P_{i-2} \cdots \to P_1 \to P_0 \to M \tag{50.4.1.1}$$

where the P_i are projective modules.

50.4.2 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 $n \geq 0$, 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 11.6.6), 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. □

Here is a useful observation:

50.4.3 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. □

50.4.4 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.

50.4.5 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:

$$\begin{array}{ccccccc} \ldots & \longrightarrow & P_1 & \longrightarrow & P_0 & \longrightarrow & M \\ & & \downarrow & & \downarrow & & \downarrow f \\ \ldots & \longrightarrow & P'_1 & \longrightarrow & P'_0 & \longrightarrow & N \end{array}$$

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

$$\begin{array}{ccccccccccc} \ldots & \longrightarrow & P_n & \xrightarrow{d} & P_{n-1} & \xrightarrow{d} & \ldots & \xrightarrow{d} & P_0 & \longrightarrow & M & \longrightarrow & 0 \\ & & & & & & & & & & \downarrow f & & \\ \ldots & \longrightarrow & P'_n & \xrightarrow{d} & P'_{n-1} & \longrightarrow & \ldots & \xrightarrow{d} & P'_0 & \longrightarrow & N & \longrightarrow & 0 \end{array}$$

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

$$\begin{array}{ccc} P_0 & \longrightarrow & M \\ \downarrow & & \downarrow f \\ P_0' & \longrightarrow & N \end{array}$$

commute. Suppose we have defined maps $P_i \to P_i'$ for $i \leq n$ such that the following diagram commutes:

$$\begin{array}{ccccccccc} P_n & \xrightarrow{d} & P_{n-1} & \xrightarrow{d} & \dots & \xrightarrow{d} & P_0 & \longrightarrow & M & \longrightarrow & 0 \\ \downarrow & & \downarrow & & & & \downarrow & & \downarrow f & & \\ P_n' & \xrightarrow{d} & P_{n-1}' & \longrightarrow & \dots & \xrightarrow{d} & P_0' & \longrightarrow & N & \longrightarrow & 0 \end{array}$$

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

$$\begin{array}{ccc} P_{n+1} & \longrightarrow & P_n \\ & & \downarrow \\ P_{n+1}' & \longrightarrow & P_n' \end{array},$$

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

$$\begin{array}{ccccc} & & P_{n+1} & & \\ & & \downarrow & & \\ P_{n+1}' & \longrightarrow & I & \longrightarrow & 0 \end{array}$$

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

$$\begin{array}{ccc} P_0 & \longrightarrow & M \\ \downarrow & & \downarrow \\ P_0' & \longrightarrow & N \end{array}$$

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 $i \leq n$ such that $dD_i + D_{i-1}d = f - g$ for $i \leq n$. We will define D_{n+1}. There is a diagram

$$\begin{array}{ccccccc} & & P_{n+1} & \longrightarrow & P_n & \longrightarrow & P_{n-1} \\ & & \downarrow & \swarrow^{D_n} & \downarrow & \swarrow^{D_{n-1}} & \downarrow \\ P'_{n+2} & \longrightarrow & P'_{n+1} & \longrightarrow & P'_n & \longrightarrow & P'_{n-1} \end{array}$$

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.

50.4.6 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. $\square$

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.

50.4.7 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 $L_i F$ 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 **??**.

50.4.8 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:*

(50.4.8.1)
$$\cdots L_iF(A) \to L_iF(B) \to L_iF(C) \to L_{i-1}F(A) \cdots \to L_1(C) \to L_0F(A) \to L_0F(B) \to L_0F(C) \to 0$$

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**

50.4.9 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.

50.4.10 Definition The homology of the complex $\{\hom_R(P_i, N)\}$ is denoted $\operatorname{Ext}^i_R(M, N)$. 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)

50.4.11 Proposition $\operatorname{Ext}^0_R(M, N) = \hom_R(M, N)$.

Proof. This is obvious from the left-exactness of $\hom(-, N)$. (We discussed this.) □

50.4.12 Proposition $\operatorname{Ext}^i(M, N)$ *is a functor of* N.

Proof. Obvious from the definition. □

Here is a harder statement.

50.4.13 Proposition $\operatorname{Ext}^i(M, N)$ *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. □

50.4.14 Proposition *If* M *is annihilated by some ideal* $I \subset R$*, then so is* $\operatorname{Ext}^i(M, N)$ *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 $\operatorname{Ext}^i(M, N)$.

50.4.15 Proposition $\operatorname{Ext}^i(M, N) = 0$ *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. □

50.4.16 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*

$$0 \to \hom(M, N') \to \hom(M, N) \to \hom(M, N'') \to \operatorname{Ext}^1(M, N') \to \operatorname{Ext}^1(M, N) \to \dots$$

[1] I.e. a map $M \to M'$ induces $\operatorname{Ext}^i(M', N) \to \operatorname{Ext}^i(M, N)$.

Proof. This proof will assume a little homological algebra. Choose a projective resolution $P_* \to M$. (The notation P_* means the chain complex $\cdots \to P_2 \to P_1 \to P_0$.) 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:

50.4.17 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

$$0 \to \hom_R(M'', N) \to \hom_R(M, N) \to \hom_R(M', N) \to \operatorname{Ext}^1(M'', N) \to \operatorname{Ext}^1(M, N) \to \dots.$$

Proof. Omitted. □

We now can characterize projectivity:

50.4.18 Corollary *TFAE:*

1. *M is projective.*
2. *$\operatorname{Ext}^i(M, N) = 0$ for all R-modules N and $i > 0$.*
3. *$\operatorname{Ext}^1(M, N) = 0$ 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

$$0 \to \hom(M, K) \to \hom(P, K) \to \hom(K, K) \to \operatorname{Ext}^1(M, K) = 0$$

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.

50.4.19 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.

50.4.20 Definition We call these cohomologies

$$\operatorname{Ext}^i_R(M, N)' = \ker(\hom(M, Q^i) \to \hom(M, Q^{i+1}))/\operatorname{im}(\hom(M, Q^{i-1}) \to \hom(M, Q^i)).$$

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:

50.4.21 Theorem *There are canonical isomorphisms:*

$$\operatorname{Ext}^i(M, N)' \simeq \operatorname{Ext}^i(M, N).$$

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,

$$\operatorname{Ext}^i(M, N) \to H^i(C'_*), \quad \operatorname{Ext}^i(M, N)' \to H^i(C'_*).$$

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. □

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 $\operatorname{Ext}^i(M, N)$ for each i, and proved various properties. We forgot to mention one.

50.4.22 Proposition *If R noetherian, and M, N are finitely generated, $\operatorname{Ext}^i(M, N)$ 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. □

51. Homological algebra à la Grothendieck

51.1. Additive categories

51.1.1 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 $\mathrm{Mor}(A,B)$ carries an abelian group structure

$$+_{(A,B)} : \mathrm{Mor}(A,B) \times \mathrm{Mor}(A,B) \to \mathrm{Mor}(A,B), \quad (f,g) \mapsto f+g$$

such that composition of morphisms in A is bilinear in the following sense:

(BL) If A, B, C are objects of A, $f, f' \in \mathrm{Mor}(A,B)$ and $g, g' \in \mathrm{Mor}(B,C)$, then

$$g \circ (f+f') = (g \circ f) + (g \circ f') \quad \text{and} \quad (g+g') \circ f = (g \circ f) + (g' \circ f)\,.$$

51.1.2 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.

51.1.3 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

$$\sum_{l=1}^n i_l \circ p_l = \mathrm{id}_{\prod_{l=1}^n A_l} \tag{51.1.3.1}$$

holds true.

(2) *If* $\coprod_{l=1}^n A_l$ *is a coproduct with canonical injections* $i_k : \coprod_{l=1}^n A_l \to A_k$, $k = 1, \dots, n$, *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

$$\sum_{l=1}^n i_l \circ p_l = \mathrm{id}_{\coprod_{l=1}^n A_l} \tag{51.1.3.2}$$

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 $k = 1, \dots, n$,

$$p_k \circ \Big(\sum_{l=1}^n i_l \circ p_l \Big) = \sum_{l=1}^n p_k \circ i_l \circ p_l = p_k \, .$$

By the universal property of the product, Equation (51.1.3.1) follows. Now let $f_k : A_k \to X$, $k = 1, \dots, n$, 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

$$\begin{aligned} f - \tilde{f} \; &= \big(f - \tilde{f}\big) \circ \Big(\sum_{l=1}^n i_l \circ p_l \Big) = \sum_{l=1}^n \big(f - \tilde{f}\big) \circ i_l \circ p_l = \\ &= \sum_{l=1}^n \big(f - \tilde{f}\big) \circ i_l \circ p_l = \sum_{l=1}^n \big(f_l - f_l\big) \circ p_l = 0 \, . \end{aligned}$$

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, $k = 1, \dots, n$ 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.

51.1.4 Definition A pre-additive category A is called *additive*, if it has the following properties:

($\mathsf{A}0$) A has a zero object.

($\mathsf{A}1$) Every finite family of objects has a product.

($\mathsf{A}1$)° Every finite family of objects has a coproduct.

51.1.5 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.

51.2. Abelian categories

51.2.1 Definition By an *abelian category* one understands an additive category A which fulfills the following axioms by Grothendieck:

($\mathsf{AB}1$) Every morphism has a kernel and a cokernel.

($\mathsf{AB}2$) For every morphism f the induced canonical morphism $\operatorname{coim} f \to \operatorname{im} f$ is an isomorphism.

51.2.2 Proposition *Assume that* A *is an abelian category, and let*

$$\begin{array}{ccc} X & \xrightarrow{f} & A \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle r} \\ B & \xrightarrow[s]{} & Y \end{array} \tag{51.2.2.1}$$

be a commutative diagram in A.

(1) *The diagram is cartesian if and only if the sequence*

$$0 \longrightarrow X \xrightarrow{i_1 f + i_2 g} A \oplus B \xrightarrow{r p_1 - s p_2} Y \tag{51.2.2.2}$$

is exact.

(2) *The diagram is cocartesian, if and only if*

$$X \xrightarrow{i_1 f - i_2 g} A \oplus B \xrightarrow{r p_1 + s p_2} Y \longrightarrow 0 \tag{51.2.2.3}$$

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*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker s & \longrightarrow & X & \xrightarrow{f} & A & \longrightarrow & 0 \\ & & \| & & {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle r} & & \\ 0 & \longrightarrow & \ker s & \longrightarrow & B & \xrightarrow[s]{} & Y & \longrightarrow & 0\,. \end{array} \tag{51.2.2.4}$$

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*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & X & \xrightarrow{f} & A & \longrightarrow & \operatorname{coker} f & \longrightarrow & 0 \\ & & {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle r} & & \| & & \\ 0 & \longrightarrow & B & \xrightarrow[s]{} & Y & \longrightarrow & \operatorname{coker} f & \longrightarrow & 0. \end{array} \tag{51.2.2.5}$$

In particular this means that the cokernel of f factors through r then.

Proof. To prove (1), consider the sequence

$$0 \longrightarrow K \xrightarrow{k} A \oplus B \xrightarrow{rp_1 - sp_2} Y, \tag{51.2.2.6}$$

where k is the kernel of $rp_1 - sp_2$. Given a commutative diagram

$$\begin{array}{ccc} P & \xrightarrow{l} & A \\ {\scriptstyle m}\downarrow & & \downarrow{\scriptstyle r} \\ B & \xrightarrow[s]{} & Y, \end{array}$$

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

$$\begin{array}{ccc} K & \xrightarrow{p_1 k} & A \\ {\scriptstyle p_2 k}\downarrow & & \downarrow{\scriptstyle r} \\ B & \xrightarrow[s]{} & Y \end{array}$$

commutes as well, this implies that (51.2.2.1) is cartesian if and only if the sequence (51.2.2.2) is exact.

Next let us show (3). So assume that the diagram (51.2.2.1) 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 (51.2.2.2) and (51.2.2.3) 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 (51.2.2.1) 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. □

51.3. 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

51.3.1 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.

51.3.2 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.*

52. Homotopical algebra

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.

52.1. Model categories

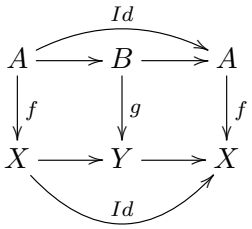
Definition

We need to begin with the notion of a *retract* of a map.

52.1.1 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

$$\begin{array}{ccc} A & \longrightarrow & C \\ \downarrow & & \downarrow \\ B & \longrightarrow & D \end{array} .$$

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:



For instance, one can prove:

52.1.2 Proposition *In any category, isomorphisms are closed under retracts.*

We leave the proof as an exercise.

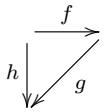
52.1.3 Definition A **model category** is a category $\mathcal{C}$ equipped with three classes of maps called *cofibrations*, *fibrations*, and *weak equivalences*. They have to satisfy five axioms $M1 - M5$.

Denote cofibrations as $\hookrightarrow$, fibrations as $\twoheadrightarrow$, and weak equivalences as $\to \sim$.

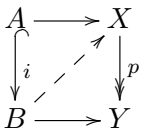
(M1) $\mathcal{C}$ is closed under all limits and colimits.[1]

(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.



(M4) (*Lifts*) Suppose we have a diagram

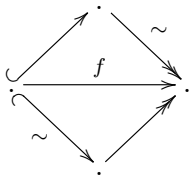


Here $i : A \to B$ is a cofibration and $p : X \to Y$ is a fibration. Then a lift exists if i or p is a weak equivalence.

[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:



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.)

52.1.4 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:

52.1.5 Example

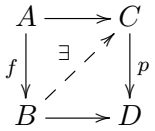
1. Given a complete and cocomplete category $\mathcal{C}$, then we can give a model structure to $\mathcal{C}$ by taking the weak equivalences to be the isomorphisms and the cofibrations and fibrations to be all maps.
2. If R is a *Frobenius ring*, or the classes of projective and injective R-modules 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.

52.1.6 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.

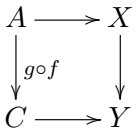
52.1.7 Definition Let $\mathcal{C}$ be any category. Suppose that P is a class of maps of $\mathcal{C}$. A map $f : A \to B$ has the **left lifting property** with respect to P iff: for all $p : C \to D$ in P and all diagrams



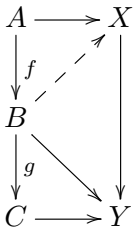
a lift represented by the dotted arrow exists, making the diagram commute. We abbreviate this property to **LLP**. There is also a notion of a **right lifting property**, abbreviated **RLP**, where f is on the right.

52.1.8 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.*

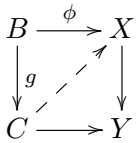
Proof. This will be a diagram chase. Suppose $f : A \to B$ and $g : B \to C$ have the LLP with respect to maps in P. Suppose given a diagram



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:



The lifting property provides a map $\phi : B \to X$ as in the dotted line in the diagram. This gives a diagram



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. □

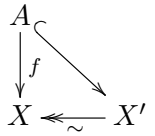
The axioms of a model category imply that cofibrations have the LLP with respect to trivial fibrations, and acyclic cofibrations have the LLP with respect to fibrations. There are dual statements for fibrations. It turns out that these properties *characterize* cofibrations and fibrations (and acyclic ones).

52.1.9 Theorem *Suppose C is a model category. Then:*

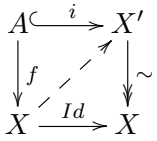
(1) A map f is a cofibration iff it has the left lifting property with respect to the class of acyclic fibrations.

(2) A map is a fibration iff it has the right lifting property w.r.t. the class of acyclic cofibrations.

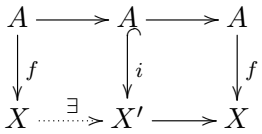
Proof. Suppose you have a map f, that has LLP w.r.t. all acyclic fibrations and you want it to be a cofibration. (The other direction is an axiom.) Somehow we're going to have to get it to be a retract of a cofibration. Somehow you have to use factorization. Factor f:



We had assumed that f has LLP. There is a lift:



This implies that f is a retract of i.

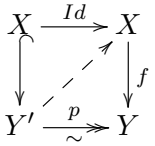


□

52.1.10 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.

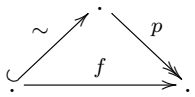
Suppose we know the cofibrations. Then we don't know the weak equivalences, or the fibrations, but we know the maps that are both. If we know the fibrations, we know the maps that are both weak equivalences and cofibrations. This is basically the same argument. One direction is easy: if a map is an acyclic 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.



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.

52.1.11 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



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.

52.1.12 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.

52.1.13 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:

$$A = \{D^n \times \{0\} \to D^n \times [0,1] \mid n \geq 0\}$$

$$B = A \cup \{S^{n-1} \to D^n \mid n \geq 0, S^{-1} = \emptyset\}$$

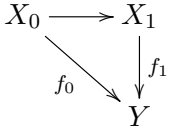
These are compact, in a category-theory sense. By definition of Serre fibrations, a map is a fibration iff it has the right lifting property with respect to A. A map is an acyclic fibration iff it has the RLP w.r.t. B. (This was on the homework.) I need another general fact:

52.1.14 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*

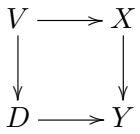
$$A_0 \hookrightarrow A_1 \to A_2 \to \cdots$$

we mean the map $A \to colim_n ßA_n$.

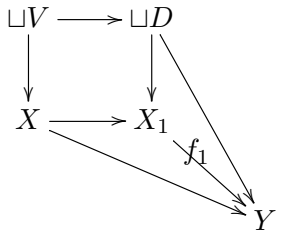
Suppose I have a map $f_0 : X_0 \to Y_0$. We want to produce a diagram:



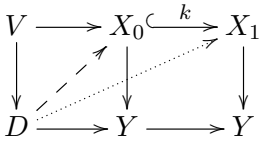
We have $\sqcup V \to \sqcup D$ where the disjoint union is taken over commutative diagrams



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:

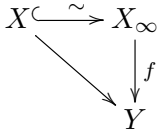


where we have subsequently made the pushout to Y. By construction, every lifting problem in X_0 can be solved in X_1.

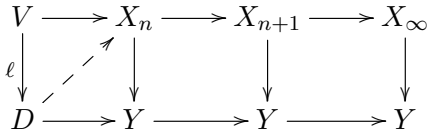


We know that every map in A is a cofibration. Also, $\sqcup V \to \sqcup D$ is a homotopy equivalence. k is an acylic cofibration because it is a weak equivalence (recall that it is a homotopy equivalence) and a cofibration.

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:



by which we mean



□

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 VI.

Algebraic Topology

Licenses

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.

Creative Commons Attribution 4.0 International

==========

Creative Commons Corporation ("Creative Commons") is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an "as-is" basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and conditions that creators and other rights holders may use to share original works of authorship and other material subject to copyright and certain other rights specified in the public license below. The following considerations are for informational purposes only, are not exhaustive, and do not form part of our licenses.

```
 Considerations for licensors: Our public licenses are
 intended for use by those authorized to give the public
 permission to use material in ways otherwise restricted by
 copyright and certain other rights. Our licenses are
 irrevocable. Licensors should read and understand the terms
 and conditions of the license they choose before applying it.
 Licensors should also secure all rights necessary before
 applying our licenses so that the public can reuse the
 material as expected. Licensors should clearly mark any
 material not subject to the license. This includes other CC-
 licensed material, or material used under an exception or
 limitation to copyright. More considerations for licensors:
wiki.creativecommons.org/Considerations_for_licensors

 Considerations for the public: By using one of our public
 licenses, a licensor grants the public permission to use the
 licensed material under specified terms and conditions. If
 the licensor's permission is not necessary for any reason--for
 example, because of any applicable exception or limitation to
 copyright--then that use is not regulated by the license. Our
 licenses grant only permissions under copyright and certain
 other rights that a licensor has authority to grant. Use of
 the licensed material may still be restricted for other
 reasons, including because others have copyright or other
 rights in the material. A licensor may make special requests,
 such as asking that all changes be marked or described.
 Although not required by our licenses, you are encouraged to
```

```
 respect those requests where reasonable. More_considerations
 for the public:
wiki.creativecommons.org/Considerations_for_licensees
```

==========

Creative Commons Attribution 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. Adapted Material means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. Adapter's License means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. Copyright and Similar Rights means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

d. Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

e. Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

f. Licensed Material means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

g. Licensed Rights means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

h. Licensor means the individual(s) or entity(ies) granting rights under this Public License.

i. Share means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

j. Sui Generis Database Rights means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

k. You means the individual or entity exercising the Licensed Rights under this Public License. Your has a corresponding meaning.

Section 2 – Scope.

a. License grant.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

```
a. reproduce and Share the Licensed Material, in whole or
   in part; and

b. produce, reproduce, and Share Adapted Material.
```

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)

   (4) never produces Adapted Material.

5. Downstream recipients.

   a. Offer from the Licensor -- Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

   b. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. Other rights.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
2. Patent and trademark rights are not licensed under this Public License.
3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. Attribution.

1. If You Share the Licensed Material (including in modified form), You must:

   a. retain the following if it is supplied by the Licensor with the Licensed Material:

      i. identification of the creator(s) of the Licensed Material and any others designated to receive

```
            attribution, in any reasonable manner requested by
            the Licensor (including by pseudonym if
            designated);

         ii. a copyright notice;

        iii. a notice that refers to this Public License;

         iv. a notice that refers to the disclaimer of
            warranties;

          v. a URI or hyperlink to the Licensed Material to the
            extent reasonably practicable;

      b. indicate if You modified the Licensed Material and
         retain an indication of any previous modifications; and

      c. indicate the Licensed Material is licensed under this
         Public License, and include the text of, or the URI or
         hyperlink to, this Public License.
```

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.
3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.
4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

a. UNLESS OTHERWISE SEPARATELY UNDERTAKEN BY THE LICENSOR, TO THE EXTENT POSSIBLE, THE LICENSOR OFFERS THE LICENSED MATERIAL AS-IS AND AS-AVAILABLE, AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE LICENSED MATERIAL, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHER. THIS INCLUDES, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OR ABSENCE OF ERRORS, WHETHER OR NOT KNOWN OR DISCOVERABLE. WHERE DISCLAIMERS OF WARRANTIES ARE NOT ALLOWED IN FULL OR IN PART, THIS DISCLAIMER MAY NOT APPLY TO YOU.

b. TO THE EXTENT POSSIBLE, IN NO EVENT WILL THE LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY (INCLUDING, WITHOUT LIMITATION, NEGLIGENCE) OR OTHERWISE FOR ANY DIRECT, SPECIAL, INDIRECT, INCIDENTAL, CONSEQUENTIAL, PUNITIVE, EXEMPLARY, OR OTHER LOSSES, COSTS, EXPENSES, OR DAMAGES ARISING OUT OF THIS PUBLIC LICENSE OR USE OF THE LICENSED MATERIAL, EVEN IF THE LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH LOSSES, COSTS, EXPENSES, OR DAMAGES. WHERE A LIMITATION OF LIABILITY IS NOT ALLOWED IN FULL OR IN PART, THIS LIMITATION MAY NOT APPLY TO YOU.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

    1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or

    2. upon express reinstatement by the Licensor.

    For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

==========

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the "Licensor." Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark "Creative Commons" or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at `https://creativecommons.org`.

Bibliography

Alexandroff, P. & Hopf, H. (1965). *Topologie*. Erster Band. Grundbegriffe der mengentheoretischen Topologie, Topologie der Komplexe, topologische Invarianzsätze und anschliessende Begriffsbildunge n, Verschlingungen im n-dimensionalen euklidischen Raum, stetige Abbildungen vo n 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. (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. (1957). Sur quelques points d'algèbre homologique. *Tôhoku Math. J. (2)*, *9*, 119–221.

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.

Lang, S. (2002). *Algebra* (third ed.)., volume 211 of *Graduate Texts in Mathematics*. New York: 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. Springer, New York.

Steen, L. A. & Seebach, Jr., J. A. (1995). *Counterexamples in topology*. Dover Publications, Inc., Mineola, NY. Reprint of the second (1978) edition.