Commutative Algebra

Math 225 Winter 2021

and

Homological Algebra

Math 224 Spring 2021

Eloísa Grifo
University of California, Riverside

May 20, 2021

Warning!

Proceed with caution. These notes are under construction and are 100% guaranteed to contain typos. If you find any typos or errors, I will be most grateful to you for letting me know. If you are looking for a place where to learn commutative algebra or homological algebra, I strongly recommend the following excellent resources:

- Mel Hochster's Lecture notes
- Jack Jeffries' Lecture notes (either his UMich 614 notes or his CIMAT notes)
- Atiyah and MacDonald's *Commutative Algebra* [AM69]
- Matsumura's *Commutative Ring Theory* [Mat89], or his other less known book *Commutative Algebra* [Mat80]
- Eisenbud's *Commutative Algebra with a view towards algebraic geometry* [Eis95]
- Rotman's *An introduction to homological algebra*, second edition. [Rot09]

Acknowledgements

Part I of these notes is heavily based on Jack Jeffries and Alexandra Seceleanu's notes, and I thank them for sharing their notes with me. Thank you also to all the students in my commutative algebra class at UCR in Winter 2021 for their comments and questions that lead to multiple improvements, especially Brandon Massaro, Adam Richardson, Khoa Ta, Ryan Watson, and Noble Williamson, who found typos and errors.

Contents

Chapter 0

Setting the stage

In this chapter we set the stage for what's to come in the rest of the class. The definitions and facts we collect here should be somewhat familiar to you already, and so we present them in rapid fire succession. You can learn more about the basic theory of (commutative) rings and R-modules in any introductory algebra book, such as [DF04].

0.1 Basic definitions: rings and ideals

Roughly speaking, Commutative Algebra is the branch of algebra that studies commutative rings and modules over such rings. For a commutative algebraist, every ring is commutative and has a $1 \neq 0$.

Definition 0.1 (Ring)**.** A **ring** is a set R equipped with two binary operations $+$ and $\cdot$ satisfying the following properties:

1) R is an abelian group under the addition operation $+$, with additive identity 0.[1] Explicitly, this means that
   - $a + (b + c) = (a + b) + c$ for all $a, b, c \in R$,
   - $a + b = b + a$ for all $a, b \in R$,
   - there is an element $0 \in R$ such that $0 + a = a$ for all $a \in R$, and
   - for each $a \in R$ there exists an element $-a \in R$ such that $a + (-a) = 0$.
2) R is a commutative monoid under the multiplication operation $\cdot$, with multiplicative identity 1.[2] Explicitly, this means that
   - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$,
   - $a \cdot b = b \cdot a$ for all $a, b \in R$, and
   - there exists an element $1 \in R$ such that $1 \cdot a = a \cdot 1$ for all $a \in R$.

[1] Or 0_R if we need to specify which ring we are talking about.
[2] If we need to specify the corresponding ring, we may write 1_R.

3) multiplication is distributive with respect to addition, meaning that

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

for all $a, b, c \in R$.

4) $1 \neq 0$.

We typically write ab for $a \cdot b$.

While in some branches of algebra rings might fail to be commutative, we will explicitly say we have a *noncommutative ring* if that is the case, and otherwise all rings are assumed to be commutative. There also branches of algebra where rings might be assumed to not necessarily have a multiplicative identity; we recommend [Poo19] for an excellent read on the topic of *Why rings should have a 1.*

Example 0.2. Here are some examples of the kinds of rings we will be talking about.

a) The integers $\mathbb{Z}$.

b) Any quotient of $\mathbb{Z}$, which we write compactly as $\mathbb{Z}/n$.

c) A polynomial ring. When we say polynomial ring, we typically mean $R = k[x_1, \ldots, x_n]$, a polynomial ring in finitely many variables over a field k.

d) A quotient of a polynomial ring by an ideal I, say $R = k[x_1, \ldots, x_n]/I$.

e) Rings of polynomials in infinitely many variables, $R = k[x_1, x_2, \ldots]$.

f) Power series rings $R = k[\![x_1, \ldots, x_n]\!]$. The elements are (formal) power series $\displaystyle\sum_{a_i \geqslant 0} c_{a_1,\ldots,a_n} x_1^{a_1} \cdots x_n^{a_n}$.

g) While any field k is a ring, we will see that fields on their own are not very exciting from the perspective of the kinds of things we will be discussing in this class.

Definition 0.3 (ring homomorphism)**.** A map $R \xrightarrow{f} S$ between rings is a **ring homomorphism** if f preserves the operations and the multiplicative identity, meaning

- $f(a + b) = f(a) + f(b)$ for all $a, b \in R$,
- $f(ab) = f(a)f(b)$ for all $a, b \in R$, and
- $f(1) = 1$.

A bijective ring homomorphism is an **isomorphism**. We should think about a ring isomorphism as a relabelling of the elements in our ring.

Definition 0.4. A subset $R \subseteq S$ of a ring S is a **subring** if R is also a ring with the structure induced by S, meaning that the each operation on R is the restrictions of the corresponding operation on S to R, and the 0 and 1 in R are the 0 and 1 in S, respectively.

Often, we care about the ideals in a ring more than we care about individual elements.

Definition 0.5 (ideal)**.** A nonempty subset I of a ring R is an **ideal** if it is closed for the addition and for multiplication by any element in R: for any $a, b \in I$ and $r \in R$, we must have $a + b \in I$ and $ra \in I$.

The **ideal generated by** $f_1, \ldots, f_n$, denoted $(f_1, \ldots, f_n)$, is the smallest ideal containing $f_1, \ldots, f_n$, or equivalently,

$$(f_1, \ldots, f_n) = \{r_1 f_1 + \cdots r_n f_n \,|\, r_i \in R\}\,.$$

Example 0.6. Every ring has always at least 2 ideals, the zero ideal $(0) = \{0\}$ and the unit ideal $(1) = R$.

We will follow the convention that when we say *ideal* we actually mean every ideal $I \neq R$.

Exercise 1. The ideals in $\mathbb{Z}$ are the sets of multiples of a fixed integer, meaning every ideal has the form (n). In particular, every ideal in $\mathbb{Z}$ can be generated by one element.

This makes $\mathbb{Z}$ the canonical example of a **principal ideal domain**.

A **domain** is a ring with no zerodivisors, meaning that $rs = 0$ implies that $r = 0$ or $s = 0$. A **principal ideal** is an ideal generated by one element. A **principal ideal domain** or **PID** is a domain where every ideal is principal.

Exercise 2. Given a field k, $R = k[x]$ is a principal ideal domain, so every ideal in R is of the form $(f) = \{fg \,|\, g \in R\}$.

Exercise 3. While $R = k[x, y]$ is a domain, it is **not** a PID. We will see later that every ideal in R is finitely generated, and yet we can construct ideals in R with arbitrarily many generators!

Example 0.7. While $\mathbb{Z}[x]$ is a domain, it is also **not** a PID. For example, $(2, x)$ is not a principal ideal.

Finally, here is an elementary fact we will need, known as the Chinese Remainder Theorem:

Theorem 0.8. *Let R be a ring and $I_1, \ldots, I_n$ be pairwise coprime ideals in R, meaning $I_i + I_j = R$ for all $i \neq j$. Then $I := I_1 \cap \cdots \cap I_n = I_1 \cdots I_n$, and there is an isomorphism of rings*

$$\begin{aligned} R/I &\xrightarrow{\cong} R/I_1 \times \cdots \times R/I_n\ . \\ r + I &\longmapsto (r + I_1, \ldots, r + I_n) \end{aligned}$$

0.2 Basic definitions: modules

Similarly to how linear algebra is the study of vector spaces over fields, commutative algebra often focuses on the structure of modules over a given commutative ring R. While in other branches of algebra modules might be left- or right-modules, all our modules are two sided, and we refer to them simply as modules.

Definition 0.9 (Module)**.** Given a ring R, an R**-module** $(M, +)$ is an abelian group equipped with an R-action that is compatible with the group structure. More precisely, there is an operation $\cdot : R \times M \longrightarrow M$ such that

- $r \cdot (a + b) = r \cdot a + r \cdot b$ for all $r \in R$ and $a, b \in M$,
- $(r + s) \cdot a = r \cdot a + s \cdot a$ for all $r, s \in R$ and $a \in M$,
- $(rs) \cdot a = r \cdot (s \cdot a)$ for all $r, s \in R$ and $a \in M$, and
- $1 \cdot a = a$ for all $a \in M$.

We typically write ra for $r \cdot a$. We denote the additive identity in M by 0, or 0_M if we need to distinguish it from 0_R.

The definitions of submodule, quotient of modules, and homomorphism of modules are very natural and easy to guess, but here they are.

Definition 0.10. If $N \subseteq M$ are R-modules with compatible structures, we say that M is a **submodule** of M.

A map $M \xrightarrow{f} N$ between R-modules is a **homomorphism of** R**-modules** if it is a homomorphism of abelian groups that preserves the R-action, meaning $f(ra) = rf(a)$ for all $r \in R$ and all $a \in M$. We sometimes refer to R-module homomorphisms as R**-module maps**, or **maps of** R**-modules**. An isomorphism of R-modules is a bijective homomorphism, which we really should think about as a relabeling of the elements in our module. If two modules M and N are isomorphic, we write $M \cong N$.

Given an R-module M and a submodule $N \subseteq M$, the **quotient** M/N is an R-module whose elements are the equivalence classes determined by the relation on M given by $a \sim b \Leftrightarrow a - b \in N$. One can check that this set naturally inherits an R-module structure from the R-module structure on M, and it comes equipped with a natural **canonical map** $M \longrightarrow M/N$ induced by sending 1 to its equivalence class.

Example 0.11. The modules over a field k are precisely all the k-vector spaces. Linear transformations are precisely all the k-module maps.

While vector spaces make for a great first example, be warned that many of the basic facts we are used to from linear algebra are often a little more subtle in commutative algebra. These differences are features, not bugs.

Example 0.12. The $\mathbb{Z}$-modules are precisely all the abelian groups.

Example 0.13. When we think of the ring R as a module over itself, the submodules of R are precisely the ideals of R.

Exercise 4. The kernel $\ker f$ and image $\operatorname{im} f$ of an R-module homomorphism $M \xrightarrow{f} N$ are submodules of M and N, respectively.

Theorem 0.14 (First Isomorphism Theorem)**.** *For any homomorphism of R-modules $M \xrightarrow{f} N$, $M/\ker f \cong \operatorname{im} f$.*

The first big noticeable difference between vector spaces and more general R-modules is that while every vector space has a basis, most R-modules do not.

Definition 0.15. A subset $\Gamma \subseteq M$ of an R-module M is a **generating set**, or a **set of generators**, if every element in M can be written as a finite linear combination of elements in M with coefficients in R. A **basis** for an R-module M is a generating set Γ for M such that $\sum_i a_i\gamma_i = 0$ implies $a_i = 0$ for all i. An R-module is **free** if it has a basis.

Remark 0.16. Every vector space is a free module.

Remark 0.17. Every free R-module is isomorphic to a direct sum of copies of R. Indeed, let's construct such an isomorphism for a given free R-module M. Given a basis $\Gamma = \{\gamma_i\}_{i\in I}$ for M, let

$$\begin{aligned} \textstyle\bigoplus_{i\in I} R &\xrightarrow{\ \pi\ } M \qquad . \\ (r_i)_{i\in I} &\longrightarrow \sum_i r_i\gamma_i \end{aligned}$$

The condition that Γ is a basis for M can be restated into the statement that π is an isomorphism of R-modules.

One of the key things that makes commutative algebra so rich and beautiful is that most modules are in fact *not* free. In general, every R-module has a generating set — for example, M itself. Given some generating set Γ for M, we can always repeat the idea above and write a **presentation** $\oplus_{i\in I} R \xrightarrow{\pi} M$ for M, but in general the resulting map π will have a nontrivial kernel. A nonzero kernel element $(r_i)_{i\in I} \in \ker \pi$ corresponds to a **relation** between the generators of M.

Remark 0.18. Given a set of generators for an R-module M, any homomorphism of R-modules $M \longrightarrow N$ is determined by the images of the generators.

We say that a module is **finitely generated** if we can find a finite generating set for M. The simplest finitely generated modules are the cyclic modules.

Example 0.19. An R-module is **cyclic** if it can be generated by one element. Equivalently, we can write M as a quotient of R by some ideal I. Indeed, given a generator m for M, the kernel of the map $R \xrightarrow{\pi} M$ induced by $1 \mapsto m$ is some ideal I. Since we assumed that m generates M, π is automatically surjective, and thus induces an isomorphism $R/I \cong M$.

Similarly, if an R-module has n generators, we can naturally think about it as a quotient of R^n by the submodule of relations among those n generators.

0.3 Why study commutative algebra?

There are many reasons why one would want to study commutative algebra. For starters, it's fun! Also, modern commutative algebra has connections with many fields of mathematics, including:

- Algebra Geometry
- Algebraic Topology
- Homological Algebra
- Category Theory
- Number Theory
- Arithmetic Geometry
- Combinatorics
- Invariant Theory
- Representation Theory
- Differential Algebra
- Lie Algebras
- Cluster Algebras

Part I

Commutative Algebra

Chapter 1

Finiteness conditions

1.1 Noetherian rings and modules

The most common assumption in commutative algebra is to require that our rings be Noetherian. Noetherian rings are named after Emmy Noether, who is in many ways the mother of modern commutative algebra. Many rings that one would naturally want to study are noetherian.

Definition 1.1 (Noetherian ring)**.** A ring R is *Noetherian* if every ascending chain of ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

eventually stabilizes: there is some N for which $I_n = I_{n+1}$ for all $n \geqslant N$.

This condition can be restated in various equivalent forms.

Proposition 1.2. *Let R be a ring. The following are equivalent:*

1) R is a Noetherian ring.

2) Every nonempty family of ideals has a maximal element (under $\subseteq$).

3) Every ascending chain of finitely generated ideals of R stabilizes.

4) Given any generating set S for an ideal I, I is generated by a finite subset of S.

5) Every ideal of R is finitely generated.

Proof.
(1)⇒(2): We prove the contrapositive. Suppose there is a nonempty family of ideals with no maximal element. This means that we can keep inductively choosing larger ideals from this family to obtain an infinite properly ascending chain.

(2)⇒(1): An ascending chain of ideals is a family of ideals, and the maximal ideal in the family indicates where our chain stabilizes.

(1)⇒(3): Clear.

(3)⇒(4): Let's prove the contrapositive. Suppose that there is an ideal I and a generating set S for I such that no finite subset of S generates I. So for any finite $S' \subseteq S$ we have

$(S') \subsetneq (S) = I$, so there is some $s \in S \smallsetminus (S')$. Thus, $(S') \subsetneq (S' \cup \{s\})$. Inductively, we can continue this process to obtain an infinite proper chain of finitely generated ideals, contradicting (3).

(4)$\Rightarrow$(5): Clear.

(5)$\Rightarrow$(1): Given an ascending chain of ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

let $I = \bigcup_{n \in \mathbb{N}} I_n$. In general, the union of two ideals might fail to be an ideal, but the union of a chain of ideals is an ideal (exercise). By assumption, the ideal I is finitely generated, say $I = (a_1, \ldots, a_t)$, and since each a_i is in some I_{n_i}, there is an N such that every a_i is in I_N. But then $I_N = I$, and thus $I_n = I_{n+1}$ for all $n \geqslant N$. □

Remark 1.3. When we say that every non-empty family of ideals has a maximal element, that maximal element does not have to be unique in any way. An ideal I is maximal in the family $\mathcal{F}$ if $I \subseteq J$ for some $J \in \mathcal{F}$ implies $I = J$; we might have many incomparable maximal elements in $\mathcal{F}$. For example, every element in the family of ideals in $\mathbb{Z}$ given by

$$\mathcal{F} = \{(p) \mid p \text{ is a prime integer}\}$$

is maximal.

Remark 1.4. If R is a Noetherian ring and S is a non-empty set of ideals in R, not only does S have a maximal element, but every element in S must be contained in a maximal element of S. Given an element $I \in S$, the subset T of S of ideals in S that contain I is nonempty, and must then contain a maximal element J by Proposition 1.2. If $J \subseteq L$ for some $L \in S$, then $I \subseteq L$, so $L \in T$, and thus by maximality of J in T, we must $J = L$. This proves that J is in fact a maximal element in S, and by construction it contains I.

Example 1.5.

1) If $R = k$ is a field, the only ideals in k are (0) and $(1) = k$, so k is a Noetherian ring.

2) $\mathbb{Z}$ is a Noetherian ring. More generally, if R is a PID, then R is Noetherian. Indeed, every ideal is finitely generated!

3) As a special case of the previous example, consider the ring of germs of complex analytic functions near 0,

$$\mathbb{C}\{z\} := \{f(z) \in \mathbb{C}[\![z]\!] \mid f \text{ is analytic on a neighborhood of } z = 0\}.$$

   This ring is a PID: every ideal is of the form (z^n), since any $f \in \mathbb{C}\{z\}$ can be written as $z^n g(z)$ for some $g(z) \neq 0$, and any such $g(z)$ is a unit in $\mathbb{C}\{z\}$.

4) A ring that is *not* Noetherian is a polynomial ring in infinitely many variables over a field k, $R = k[x_1, x_2, \ldots]$: the ascending chain of ideals

$$(x_1) \subseteq (x_1, x_2) \subseteq (x_1, x_2, x_3) \subseteq \cdots$$

   does *not* stabilize.

5) The ring $R = K[x, x^{1/2}, x^{1/3}, x^{1/4}, x^{1/5}, \dots]$ is also *not* Noetherian . A nice ascending chain of ideals is
$$(x) \subsetneq (x^{1/2}) \subsetneq (x^{1/3}) \subsetneq (x^{1/4}) \subsetneq \cdots .$$

6) The ring of continuous real-valued functions $\mathcal{C}(\mathbb{R}, \mathbb{R})$ is *not* Noetherian: the chain of ideals
$$I_n = \{f(x) \mid f|_{[-1/n,1/n]} \equiv 0\}$$
is increasing and proper. The same construction shows that the ring of infinitely differentiable real functions $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ is not Noetherian: properness of the chain follows from, e.g., Urysohn's lemma (though it's not too hard to find functions distinguishing the ideals in the chain). Note that if we asked for analytic functions instead of infinitely-differentiable functions, every element of the chain would be the zero ideal!

Remark 1.6. If R is Noetherian, and I is an ideal of R, then R/I is Noetherian as well, since there is an order-preserving bijection
$$\{\text{ideals of } R \text{ that contain } I\} \longleftrightarrow \{\text{ideals of } R/I\}.$$

This gives us many more examples, by simply taking quotients of the examples above. We will also see huge classes of easy examples once we learn about localization.

Similarly, we can define noetherian modules.

Definition 1.7 (Noetherian module)**.** An R-module M is *Noetherian* if every ascending chain of submodules of M eventually stabilizes.

There are analogous equivalent definitions for modules as we had above for rings, so we leave the proof as an exercise.

Proposition 1.8 (Equivalence definitions for Noetherian module)**.** *Let M be an R-module. The following are equivalent:*

1) M is a Noetherian module.

2) Every nonempty family of submodules has a maximal element.

3) Every ascending chain of finitely generated submodules of M eventually stabilizes.

4) Given any generating set S for a submodule N, the submodule N is generated by a finite subset of S.

5) Every submodule of M is finitely generated.

In particular, a Noetherian module must be finitely generated.

Remark 1.9. A ring R is a Noetherian ring if and only if R is Noetherian as a module over itself. However, a Noetherian ring need not be a Noetherian module over a subring. For example, consider $\mathbb{Z} \subseteq \mathbb{Q}$. These are both Noetherian *rings*, but $\mathbb{Q}$ is not a noetherian $\mathbb{Z}$-module; for example, the following is an ascending chain of submodules which does not stabilize:
$$0 \subsetneq \frac{1}{2}\mathbb{Z} \subsetneq \frac{1}{2}\mathbb{Z} + \frac{1}{3}\mathbb{Z} \subsetneq \frac{1}{2}\mathbb{Z} + \frac{1}{3}\mathbb{Z} + \frac{1}{5}\mathbb{Z} \subsetneq \cdots .$$

Definition 1.10. An **exact sequence** of R-modules is a sequence

$$\cdots \xrightarrow{f_{n-1}} M_n \xrightarrow{f_n} M_{n+1} \xrightarrow{f_{n+1}} \cdots$$

of R-modules and R-module homomorphisms such that $\operatorname{im} f_n = \ker f_{n+1}$ for all n. An exact sequence of the form

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

is a **short exact sequence**.

Remark 1.11. The sequence

$$0 \longrightarrow M \xrightarrow{f} N$$

is exact if and only if f is injective. Similarly,

$$M \xrightarrow{f} N \longrightarrow 0$$

is exact if and only if f is surjective. So

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is a short exact sequence if and only if

- f is injective
- g is surjective
- $\operatorname{im} f = \ker g$.

So when this is indeed a short exact sequence, we can identify A with its image $f(A)$, and $A = \ker g$. Moreover, since g is surjective, by the First Isomorphism Theorem we conclude that $C \cong B/A$, so we might abuse notation and identify C with B/A.

Lemma 1.12 (Noetherianity in exact sequences)**.** *In an exact sequence of modules*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

B is Noetherian if and only if A and C are Noetherian.

Proof. Assume B is Noetherian. Since A is a submodule of B, and its submodules are also submodules of B, A is Noetherian. Moreover, any submodule of B/A is of the form D/A for some submodule $D \supseteq A$ of B. Since every submodule of B is finitely generated, every submodule of C is also finitely generated. Therefore, C is Noetherian.

Conversely, assume that A and C are Noetherian, and let

$$M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$$

be a chain of submodules of B. First, note that

$$M_1 \cap A \subseteq M_2 \cap A \subseteq \cdots$$

is an ascending chain of submodules of A, and thus it stabilizes. Moreover,

$$g(M_1) \subseteq g(M_2) \subseteq g(M_3) \subseteq \cdots$$

is a chain of submodules of C, and thus it also stabilizes. Pick a large enough index n such that both of these chains stabilize. We claim that $M_n = M_{n+1}$, so that the original chain stabilizes as well. To show that, take $x \in M_{n+1}$. Then

$$g(x) \in g(M_{n+1}) = g(M_n)$$

so we can choose some $y \in M_n$ such that $g(x) = g(y)$. Then $x - y \in \ker g = \operatorname{im} f = A$. Now note that $x - y \in M_{n+1}$, so

$$x - y \in M_{n+1} \cap A = M_n \cap A.$$

Then $x - y \in M_n$, and since $y \in M_n$, we must have $x \in M_n$ as well. $\square$

Corollary 1.13. *If A and B are Notherian R-modules, then $A \oplus B$ is a Noetherian R-module.*

Proof. Apply the previous lemma to the short exact sequence

$$0 \longrightarrow A \longrightarrow A \oplus B \longrightarrow B \longrightarrow 0\ . \qquad \square$$

Corollary 1.14. *A module M is Noetherian if and only if M^n is Noetherian for some n. In particular, if R is a Noetherian ring then R^n is a Noetherian module.*

Proof. We will do induction on n. The case $n = 1$ is a tautology. For $n > 1$, consider the short exact sequence

$$0 \longrightarrow M^{n-1} \longrightarrow M^n \longrightarrow M \longrightarrow 0$$

Lemma 1.12 and the inductive hypothesis give the desired conclusion. $\square$

Proposition 1.15. *Let R be a Noetherian ring. Given an R-module M, M is a Noetherian R-module if and only if M is finitely generated. Consequently, any submodule of a finitely generated R-module is also finitely generated.*

Proof. If M is Noetherian, M is finitely generated by the equivalent definitions above, and so are all of its submodules.

Now let R be Noetherian and M be a finitely generated R-module. Then M is isomorphic to a quotient of R^n for some n, which is Noetherian. $\square$

Remark 1.16. The Notherianity hypothesis is important: if M is a finitely generated R-module over a non-Noetherian ring, M might not be Noetherian. For a dramatic example, note that R itself is a finitely generated R-module, but not Noetherian.

David Hilbert had a big influence in the early years of commutative algebra, in many different ways. Emmy Noether's early work in algebra was in part inspired by some of his work, and he later invited Emmy Noether to join the Göttingen Math Department — many of her amazing contributions to algebra happened during her time in Göttingen. Unfortunately, some of the faculty was opposed to having a woman joining the department, and for her first two years in Göttingen Noether did not have an official position nor was she paid. Hilbert's contributions also include three of the most fundamental results in commutative algebra — Hilbert's Basis Theorem, the Hilbert Syzygy Theorem, and Hilbert's Nullstellensatz. We can now prove the first.

Theorem 1.17 (Hilbert's Basis Theorem)**.** *Let R be a Noetherian ring. Then the rings $R[x_1, \ldots, x_d]$ and $R[\![x_1, \ldots, x_d]\!]$ are Noetherian.*

Remark 1.18. We can rephrase this theorem in a way that can be understood by anyone with a basic high school algebra (as opposed to abstract algebra) knowledge:

Proof. We give the proof for polynomial rings, and indicate the difference in the power series argument. By induction on d, we can reduce to the case $d = 1$. Given $I \subseteq R[x]$, let

$$J = \{a \in R \mid \text{there is some } ax^n + \text{lower order terms (wrt } x) \in I\}.$$

So $J \subseteq R$ consists of all the leading coefficients of polynomials in I. We can check (exercise) that this is an ideal of R. By our hypothesis, J is finitely generated, so let $J = (a_1, \ldots, a_t)$. Pick $f_1, \ldots, f_t \in R[x]$ such that the leading coefficient of f_i is a_i, and set $N = \max_i \{\deg f_i\}$.

Given any $f \in I$ of degree greater than N, we can cancel off the leading term of f by subtracting a suitable combination of the f_i, so any $f \in I$ can be written as $f = g + h$ where $h \in (f_1, \ldots, f_t)$ and $g \in I$ has degree at most N, so $g \in I \cap (R + Rx + \cdots + Rx^N)$. Note that since $I \cap (R + Rx + \cdots + Rx^N)$ is a submodule of the finitely generated free R-module $R + Rx + \cdots + Rx^N$, it is also finitely generated as an R-module. Given such a generating set, say $I \cap (R + Rx + \cdots + Rx^N) = (f_{t+1}, \ldots, f_s)$, we can write any such $f \in I$ as an $R[x]$-linear combination of these generators and the f_i's. Therefore, $I = (f_1, \ldots, f_t, f_{t+1}, \ldots, f_s)$ is finitely generated, and $R[x]$ is a Noetherian ring.

In the power series case, take J to be the coefficients of *lowest degree* terms. $\square$

1.2 Algebra finite-extensions

If R is a subring of S, then S is an **algebra** over R, meaning that S is a ring with a (natural) structure of an R-module that also satisfies

$$r(s_1 s_2) = (r s_1) s_2 \text{ for all } r \in R \text{ and } s_1, s_2 \in S.$$

More generally, given any ring homomorphism $\varphi : R \to S$, we can view S as an algebra over R via φ by setting $r \cdot s = \varphi(r)s$. We may abuse notation and write $r \in S$ for its image $\varphi(r) \in S$. We will see that in a lot of situations we want to study, it is enough to consider the case when φ is injective, so this abuse of notation makes sense. Giving a ring homomorphism $R \to S$ is the same as giving an R-algebra structure to S. In particular, a ring S can have different R-algebra structures given by different homomorphisms $R \to S$.

A set of elements $\Lambda \subseteq S$ **generates** S as an R-algebra if the following equivalent conditions hold:

- The only subring of S containing $\varphi(R)$ and Λ is S itself.
- Every element of S admits a polynomial expression in Λ with coefficients in $\varphi(R)$.

- Given a polynomial ring $R[X]$ on $|\Lambda|$ indeterminates, the ring homomorphism

$$\begin{aligned} R[X] &\xrightarrow{\psi} S \\ x_i &\longmapsto \lambda_i \end{aligned}$$

  is surjective.

Let S be an R-algebra and $\Lambda \subseteq S$ be a set of algebra generators for S over R. The ideal of **relations** on the elements Λ over R is the kernel of the map $\psi : R[X] \longrightarrow S$ above. This ideal consists of the polynomial functions with R-coefficients that the elements of Λ satisfy. Given an R-algebra S with generators Λ and ideal of relations I, we have a ring isomorphism $S \cong R[X]/I$ by the First Isomorphism Theorem. If we understand the ring R and generators and relations for S over R, we can get a pretty concrete understanding of S. If a sequence of elements has no nonzero relations, we say they are *algebraically independent* over R.

Remark 1.19. If $s_1, \dots, s_n \in S$ are algebraically independent over R, then $R[s_1, \dots, s_n]$ is isomorphic to the polynomial ring in n variables obver R.

We say that $\varphi : R \to S$ is **algebra-finite**, or S is a **finitely generated R-algebra**, or S is of **finite type** over R, if there exists a *finite* set of elements $f_1, \dots, f_t \in S$ that generates S as an R-algebra. A better name might be *finitely generatable*, since to say that an algebra is finitely generated does not require knowing any actual finite set of generators. From the discussion above, we conclude that S is a finitely generated R-algebra if and only if S is a quotient of some polynomial ring $R[x_1, \dots, x_d]$ over R in finitely many variables. If S is generated over R by $f_1, \dots, f_d$, we will use the notation $R[f_1, \dots, f_d]$ to denote S. Of course, for this notation to properly specify a ring, we need to understand how these generators behave under the operations. This is no problem if A and $\underline{f}$ are understood to be contained in some larger ring.

Remark 1.20. Any surjective ring homomorphism $\varphi : R \to S$ is algebra-finite, since S must then be generated over R by 1. Moreover, we can always factor φ as the surjection $R \twoheadrightarrow R/\ker(\varphi)$ followed by the inclusion $R/\ker(\varphi) \hookrightarrow S$, so to understand algebra-finiteness it suffices to restrict our attention to injective homomorphisms.

Example 1.21. Every ring is a $\mathbb{Z}$-algebra, but generally not a finitely generated one.

Remark 1.22. Let $A \subseteq B \subseteq C$ be rings. It follows from the definitions that

- $\begin{array}{c} A \subseteq B \text{ algebra-finite} \\ \text{and} \\ B \subseteq C \text{ algebra-finite} \end{array} \quad \Longrightarrow \quad A \subseteq C \text{ algebra-finite}$

- $A \subseteq C$ algebra-finite $\Longrightarrow B \subseteq C$ algebra-finite.

However, $A \subseteq C$ algebra-finite $\not\Longrightarrow$ $A \subseteq B$ algebra-finite.

Example 1.23. Let k be a field and

$$B = k[x, xy, xy^2, xy^3, \cdots] \subseteq C = k[x, y],$$

where x and y are indeterminates. While B and C are both k-algebras, C is a finitely generated k-algebra, while B is not. Indeed, any finitely generated subalgebra of B is contained in $k[x, xy, \dots, xy^m]$ for some m, since we can write the elements in any finite generating set as polynomial expressions in finitely many of the specified generators of B. However, note that every element of $k[x, xy, \dots, xy^m]$ is a k-linear combination of monomials with the property that the y exponent is no more than m times the x exponent, so this ring does not contain xy^{m+1}. Thus, B is not a finitely generated A-algebra.

There are many basic questions about algebra generators that are surprisingly difficult. Let $R = \mathbb{C}[x_1, \dots, x_n]$ and $f_1, \dots, f_n \in R$. When do $f_1, \dots, f_n$ generate R over $\mathbb{C}$? It is not too hard to show that the Jacobian determinant

$$\det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

must be a nonzero constant. It is a big open question whether this is in fact a sufficient condition!

Finally, note that an easy corollary of the Hilbert Basis Theorem is that finitely generated algebras over noetherian rings are also noetherian.

Corollary 1.24. *If R is a Noetherian ring, then any finitely generated R-algebra is Noetherian. In particular, any finitely generated algebra over a field is Noetherian.*

Proof. By our discussion above, a finitely generated R-algebra is isomorphic to a quotient of a polynomial ring over R in finitely many variables; polynomial rings over noetherian rings are Noetherian, by Hilbert's Basis Theorem, and quotients of Noetherian rings are Noetherian. □

The converse to this statement is false: there are lots of Noetherian rings that are not finitely generated algebras over a field. For example, $\mathbb{C}\{z\}$ is not algebra-finite over $\mathbb{C}$. We will see more examples of these when we talk about local rings.

1.3 Module-finite extensions

Given a ring homomorphism $\varphi : R \to S$, saying that S acquires an R-module structure via φ by $a \cdot r = \varphi(a)r$ is a particular case of *restriction of scalars*. By restriction of scalars, we mean that any S-module M also gains a new R-module structure given by $r \cdot m = \varphi(r)m$.[1] We may write ${}_{\varphi}M$ for this R-module if we need to emphasize which map we are talking about.

[1]This gives a functor from the category of S-modules to the category of R-modules.

Given an R-algebra S, we can consider the *algebra* structure of S over R, or its *module* structure over R. So instead of asking about how S is generated as an *algebra* over R, we can ask how it is generated as a *module* over R. Recall that an A-module M is generated by a set of elements $\Gamma \subseteq M$ if the following equivalent conditions hold:

- The smallest submodule of M that contains Γ is M itself.
- Every element of M can be written as an A-linear combination of elements in Γ.
- Given a free R-module on $|\Gamma|$ basis elements $R^{\oplus Y}$, the homomorphism

$$\begin{aligned} R^{\oplus Y} &\xrightarrow{\theta} M \\ y_i &\longrightarrow \gamma_i \end{aligned}$$

  is surjective.

We use the notation $M = \sum_{\gamma \in \Gamma} A\gamma$ to indicate that M is generated by Γ as a module. We say that $\varphi : A \to R$ is *module-finite* if R is a finitely-generated A-module. This is also called simply *finite* in the literature, but we'll stick with the unambiguous "module-finite."

As with algebra-finiteness, surjective maps are always module-finite in a trivial way, and it suffices to understand this notion for ring inclusions.

The notion of module-finite is much stronger than algebra-finite, since a linear combination is a very special type of polynomial expression.

Example 1.25.

a) If $K \subseteq L$ are fields, saying L is module-finite over K just means that L is a finite field extension of K.

b) The Gaussian integers $\mathbb{Z}[i]$ satisfy the well-known property (or definition, depending on your source) that any element $z \in \mathbb{Z}[i]$ admits a unique expression $z = a + bi$ with $a, b \in \mathbb{Z}$. That is, $\mathbb{Z}[i]$ is generated as a $\mathbb{Z}$-module by $\{1, i\}$; moreover, they form a free module basis!

c) If R is a ring and x an indeterminate, $R \subseteq R[x]$ is not module-finite. Indeed, $R[x]$ is a free R-module on the basis $\{1, x, x^2, x^3, \dots\}$.

d) Another map that is *not* module-finite is the inclusion $k[x] \subseteq k[x, 1/x]$. First, note that any element of $k[x, 1/x]$ can be written in the form $f(x)/x^n$ for some $f \in k[x]$ and some $n \geqslant 0$. Since $k[x]$ is a Noetherian ring, $k[x, 1/x]$ is a finitely-generated $k[x]$-module if and only if it is a Noetherian $k[x]$-module. But here is an infinite chain of submodules of $k[x, \frac{1}{x}]$:

$$k[x] \cdot \frac{1}{x} \subseteq k[x] \cdot \frac{1}{x^2} \subseteq k[x] \cdot \frac{1}{x^3} \subseteq \cdots$$

Remark 1.26. If R is an A-algebra,

- $A \subseteq R$ is algebra-finite if $R = A[f_1, \dots, f_n]$ for some $f_1, \dots, f_n \in R$.
- $A \subseteq R$ is module-finite if $R = Af_1 + \cdots + f_n$ for some $f_1, \dots, f_n \in R$.

Lemma 1.27. *If $R \subseteq S$ is module-finite and N is a finitely generated S-module, then N is a finitely generated R-module by restriction of scalars. In particular, the composition of two module-finite ring maps is module-finite.*

Proof. Let $S = Ra_1 + \cdots + Ra_r$ and $N = Sb_1 + \cdots + Sb_s$. Then we claim that

$$N = \sum_{i=1}^{r} \sum_{j=1}^{s} Ra_i b_j.$$

Indeed, given $n = \sum_{j=1}^{s} s_j b_j$, rewrite each $s_j = \sum_{i=1}^{r} r_{ij} a_i$ and substitute to get

$$n = \sum_{i=1}^{r} \sum_{j=1}^{s} r_{ij} a_i b_j$$

as an R-linear combination of the $a_i b_j$. □

Remark 1.28. Let $A \subseteq B \subseteq C$ be rings. It follows from the definitions that

- $$\begin{array}{c} A \subseteq B \text{ module-finite} \\ \text{and} \\ B \subseteq C \text{ module-finite} \end{array} \implies A \subseteq C \text{ module-finite}$$
- $A \subseteq C$ module-finite $\implies B \subseteq C$ module-finite.

However, $A \subseteq C$ module-finite $\not\Longrightarrow$ $A \subseteq B$ module-finite. Note that if A is Noetherian, then $A \subseteq C$ module-finite *does* in fact imply $A \subseteq B$ module-finite, so to find an example of this bad behavior we need A to be non-Noetherian. You will construct an example in the next problem set.

1.4 Integral extensions

In field theory, there is a close relationship between (vector space-)finite field extensions and algebraic equations. The situation for rings is similar.

Definition 1.29 (Integral element/extension)**.** Let R be an A-algebra. The element $r \in R$ is **integral** over A if there are elements $a_0, \ldots, a_{n-1} \in A$ such that

$$r^n + a_{n-1} r^{n-1} + \cdots + a_1 r + a_0 = 0;$$

i.e., r satisfies an **equation of integral dependence** over A. We say that R is **integral over** A if every $r \in R$ is integral over A.

Integral automatically implies algebraic, but the condition that there exists an equation of algebraic dependence that is *monic* is stronger in the setting of rings.

Again, we can restrict our focus to inclusion maps $A \subseteq R$.

Remark 1.30. An element $r \in R$ is integral over A if and only if r is integral over the subring $\varphi(A) \subseteq R$, so we might as well assume that φ is injective.

Definition 1.31. Given an inclusion of rings $A \subseteq R$, the **integral closure** of A in R is the set of elements in R that are integral over A. The integral closure of a domain R in its field of fractions is usually denoted by $\overline{R}$. We say A is **integrally closed** in R if A is its own integral closure in R; a **normal domain** is a domain R that is integrally closed in its field of fractions, meaning $R = \overline{R}$.

Example 1.32. The ring of integers $\mathbb{Z}$ is a normal domain, meaning its integral closure in its fraction field $\mathbb{Q}$ is $\mathbb{Z}$ itself.

Example 1.33. The ring $\mathbb{Z}[\sqrt{d}]$, where $d \in \mathbb{Z}$ is not a perfect square, is integral over $\mathbb{Z}$. Indeed, $\sqrt{d}$ satisfies the monic polynomial $r^2 - d$, and since the integral closure of $\mathbb{Z}$ is a ring containing $\mathbb{Z}$ and $\sqrt{d}$, every element in $\mathbb{Z}[\sqrt{d}]$ is integral over $\mathbb{Z}$.

Proposition 1.34. *Let $A \subseteq R$ be rings.*

1) *If $r \in R$ is integral over A then $A[r]$ is module-finite over A.*
2) *If $r_1, \dots, r_t \in R$ are integral over A then $A[r_1, \dots, r_t]$ is module-finite over A.*

Proof.

1) Suppose r is integral over A, and $r^n + a_{n-1}r^{n-1} + \cdots + a_1 r + a_0 = 0$. Then we claim that $A[r] = A + Ar + \cdots + Ar^{n-1}$. First, note that to show that any polynomial $p(r) \in A[r]$ is in $A + Ar + \cdots + Ar^{n-1}$, it is enough to show that $r^m \in A + Ar + \cdots + Ar^{n-1}$ for all m. Using induction on m, the base cases $1, r, \dots, r^{n-1} \in A + Ar + \cdots + Ar^{n-1}$ are obvious. On the other hand, we can use induction to conclude that $r^m \in A + Ar + \cdots + Ar^{n-1}$ for all $m \geqslant n-1$, since we can use the equation above to rewrite r^m as

$$r^m = r^{m-n}(a_{n-1}r^{n-1} + \cdots + a_1 r + a_0),$$

which has degree $m-1$ in r.

2) Write

$$A_0 := A \subseteq A_1 := A[r_1] \subseteq A_2 := A[r_1, r_2] \subseteq \cdots \subseteq A_t := A[r_1, \dots, r_t].$$

Note that r_i is integral over A_{i-1}, via the same monic equation of r_i over A. Then, the inclusion $A \subseteq A[r_1, \dots, r_t]$ is a composition of module-finite maps, and thus it is also module-finite. $\square$

The name “ring” is roughly based on this idea: in an extension as above, the powers wrap around (like a ring).

We will need a linear algebra fact. The **classical adjoint** of an $n \times n$ matrix $B = [b_{ij}]$ is the matrix $\operatorname{adj}(B)$ with entries $\operatorname{adj}(B)_{ij} = (-1)^{i+j} \det(\widehat{B_{ji}})$, where $\widehat{B_{ji}}$ is the matrix obtained from B by deleting its jth row and ith column. You may remember this matrix from linear algebra.

Lemma 1.35 (Determinantal trick). *Let R be a ring, $B \in M_{n\times n}(R)$, $v \in R^{\oplus n}$, and $r \in R$.*

1) $\operatorname{adj}(B)B = \det(B)I_{n\times n}$.

2) If $Bv = rv$, then $\det(rI_{n\times n} - B)v = 0$.

Proof.

1) When R is a field, this is a basic linear algebra fact. We deduce the case of a general ring from the field case.

   The ring R is a $\mathbb{Z}$-algebra, so we can write R as a quotient of some polynomial ring $\mathbb{Z}[X]$. Let $\psi : \mathbb{Z}[X] \twoheadrightarrow R$ be a surjection, $a_{ij} \in \mathbb{Z}[X]$ be such that $\psi(a_{ij}) = b_{ij}$, and let $A = [a_{ij}]$. Note that

$$\psi(\operatorname{adj}(A)_{ij}) = \operatorname{adj}(B)_{ij} \quad \text{and} \quad \psi((\operatorname{adj}(A)A)_{ij}) = (\operatorname{adj}(B)B)_{ij},$$

   since ψ is a homomorphism, and the entries are the same polynomial functions of the entries of the matrices A and B, respectively. Thus, it suffices to establish

$$\operatorname{adj}(B)B = \det(B)I_{n\times n}$$

   in the case when $R = \mathbb{Z}[X]$, and we can do this entry by entry. Now, $R = \mathbb{Z}[X]$ is an integral domain, hence a subring of a field (its fraction field). Since both sides of the equation

$$(\operatorname{adj}(B)B)_{ij} = (\det(B)I_{n\times n})_{ij}$$

   live in R and are equal in the fraction field (by linear algebra) they are equal in R. This holds for all i, j, and thus 1) holds.

2) We have $(rI_{n\times n} - B)v = 0$, so by part 1)

$$\det(rI_{n\times n} - B)v = \operatorname{adj}(rI_{n\times n} - B)(rI_{n\times n} - B)v = 0. \qquad \square$$

Theorem 1.36 (Module finite implies integral). *Let $A \subseteq R$ be module-finite. Then R is integral over A.*

Proof. Given $r \in R$, we want to show that r is integral over A. The idea is to show that multiplication by r, realized as a linear transformation over A, satisfies the characteristic polynomial of that linear transformation.

Write $R = Ar_1 + \cdots Ar_t$. We may assume that $r_1 = 1$, perhaps by adding module generators. By assumption, we can find $a_{ij} \in A$ such that

$$rr_i = \sum_{j=1}^{t} a_{ij} r_j$$

for each i. Let $C = [a_{ij}]$, and v be the column vector $(r_1, \dots, r_t)$. We have $rv = Cv$, so by the determinant trick, $\det(rI_{n\times n} - C)v = 0$. Since we chose one of the entries of v to be 1, we have in particular that $\det(rI_{n\times n} - C) = 0$. Expanding this determinant as a polynomial in r, this is a monic equation with coefficients in A. $\square$

Collecting the previous results, we now have a useful characterization of module-finite extensions:

Corollary 1.37 (Characterization of module-finite extensions)**.** *Let $A \subseteq R$ be rings. R is module-finite over A if and only if R is integral and algebra-finite over A.*

Proof. ($\Rightarrow$): A generating set for R as an A-module serves as a generating set as an A-algebra. The remainder of this direction comes from the previous theorem. ($\Leftarrow$): If $R = A[r_1, \dots, r_t]$ is integral over A, so that each r_i is integral over A, then R is module-finite over A by Proposition 1.34. ☐

Corollary 1.38. *If R is generated over A by integral elements, then R is integral. Thus, if $A \subseteq S$, the set of elements of S that are integral over A form a subring of S.*

Proof. Let $R = A[\Lambda]$, with λ integral over A for all $\lambda \in \Lambda$. Given $r \in R$, there is a finite subset $L \subseteq \Lambda$ such that $r \in A[L]$. By the theorem, $A[L]$ is module-finite over A, and $r \in A[L]$ is integral over A.

For the latter statement, the first statement implies that

$$\{\text{integral elements}\} \subseteq A[\{\text{integral elements}\}] \subseteq \{\text{integral elements}\},$$

so equality holds throughout, and $\{$integral elements$\}$ is a ring. ☐

Definition 1.39. If $A \subseteq R$, the **integral closure of A in R** is the set of elements of R that are integral over A.

So the previous result says that the integral closure of A in R is a subring of R (containing A).

Example 1.40.

1) Let $R = \mathbb{C}[x, y] \subseteq S = \mathbb{C}[x, y, z]/(x^2 + y^2 + z^2)$. Then S is module-finite over R: indeed, S is generated over R as an algebra by one element, z, and z satisfies the monic equation $r^2 + x^2 + y^2 = 0$, so it is integral over R.
2) Not all integral extensions are module-finite. Consider
$$A = k[x] \subseteq R = k[x, x^{1/2}, x^{1/3}, x^{1/4}, x^{1/5}, \dots].$$
R is generated by integral elements over $k[x]$, but it is not algebra-finite over $k[x]$.

Finally, we can prove a technical sounding result that puts together all our finiteness conditions in a useful way. We will then be able to answer a classical question using this result.

Theorem 1.41 (Artin-Tate Lemma)**.** *Let $A \subseteq B \subseteq C$ be rings. Assume that*

- *A is Noetherian,*
- *C is module-finite over B, and*
- *C is algebra-finite over A.*

Then, B is algebra-finite over A.

Proof. Let $C = A[f_1, \dots, f_r]$ and $C = Bg_1 + \cdots + Bg_s$. Then,

$$f_i = \sum_j b_{ij} g_j \quad \text{and} \quad g_i g_j = \sum_k b_{ijk} g_k$$

for some $b_{ij}, b_{ijk} \in B$. Let $B_0 = A[\{b_{ij}, b_{ijk}\}] \subseteq B$. Since A is Noetherian, so is B_0.

We claim that $C = B_0\, g_1 + \cdots + B_0 g_s$. Given an element $c \in C$, write c as a polynomial expression in $f_1, \dots, f_r$, and since the f_i are linearly combinations of the g_i, we can rewrite $c \in A[\{b_{ij}\}][g_1, \dots, g_s]$. Then using the equations for $g_i g_j$ we can write c in the form required.

Now, since B_0 is Noetherian, C is a finitely generated B_0-module, and $B \subseteq C$, then B is a finitely generated B_0-module, too. In particular, $B_0 \subseteq B$ is algebra-finite. We conclude that $A \subseteq B$ is algebra-finite, as required. $\square$

1.5 An application to invariant rings

Historically, commutative algebra has roots in classical questions of algebraic and geometric flavors, including the following natural question:

Question 1.42. Given a (finite) set of symmetries, consider the collection of polynomial functions that are fixed by all of those symmetries. Can we describe all the fixed polynomials in terms of finitely many of them?

To make this precise, let G be a group acting on a ring R, or just as well, a group of automorphisms of R. The main case we have in mind is when $R = k[x_1, \dots, x_d]$ and k is a field. We are interested in the set of elements that are *invariant* under the action,

$$R^G := \{r \in R \mid g(r) = r \text{ for all } g \in G\}.$$

Note that R^G is a subring of R. Indeed, given $r, s \in R^G$, then

$$r + s = g(r) + g(s) = g(r + s) \quad \text{and} \quad rs = g(r)g(s) = g(rs) \qquad \text{for all } g \in G,$$

since each g is a homomorphism. Note also that if $G = \langle g_1, \dots, g_t \rangle$, then $r \in R^G$ if and only if $g_i(r) = r$ for $i = 1, \dots, t$. The question above can now be rephrased as follows:

Question 1.43. Given a finite group G acting on $R = k[x_1, \dots, x_d]$, is R^G a finitely generated k-algebra?

Proposition 1.44. *Let k be a field, R be a finitely-generated k-algebra, and G a finite group of automorphisms of R that fix k. Then $R^G \subseteq R$ is module-finite.*

Proof. Since integral implies module-finite, we will show that R is algebra-finite and integral over R^G.

First, since R is generated by a finite set as a k-algebra, and $k \subseteq R^G$, it is generated by the same finite set as an R^G-algebra as well. Extend the action of G on R to $R[t]$ with G

fixing t. Now, for $r \in R$, consider the polynomial $F_r(t) = \prod_{g \in G}(t - g(r)) \in R[t]$. Then G fixes $F_r(t)$, since for each $h \in G$,

$$h(F_r(t)) = h \prod_{g \in G}(t - g(r)) = \prod_{g \in G}(h \cdot t - hg(r)) = F_r(t)$$

Thus, $F_r(t) \in (R[t])^G$. Notice that $(R[t])^G = R^G[t]$, since

$$g(a_n t^n + \cdots + a_0) = a_n t^n + \cdots + a_0 \implies (g \cdot a_n)t^n + \cdots + (g \cdot a_0) = a_n t^n + \cdots + a_0.$$

Therefore, $F_r(t) \in R^G[t]$. The leading term (with respect to t) of $F_r(t)$ is $t^{|G|}$, so $F_r(t)$ is monic, and r is integral over R^G. Therefore, R is integral over R^G. $\square$

Theorem 1.45 (Noether's finiteness theorem for invariants of finite groups)**.** *Let k be a field, R be a polynomial ring over k, and G be a finite group acting k-linearly on R. Then R^G is a finitely generated k-algebra.*

Proof. Observe that $k \subseteq R^G \subseteq R$, that k is Noetherian, $k \subseteq R$ is algebra-finite, and $R^G \subseteq R$ is module-finite. Thus, by the Artin-Tate Lemma, we are done! $\square$

Chapter summary

- R is a Noetherian ring $\iff$ every ideal I in R is Noetherian
- M is a Noetherian R-module $\underset{R \text{ Noeth}}{\overset{\text{general}}{\rightleftarrows}}$ M is a finitely generated R-module

$A \subseteq R$ extension of rings:

- $A \subseteq R$ module-finite $\iff$ $R = Af_1 + \cdots + Af_n$ for some $f_i \in R$ $\iff$ $R \cong A^n/N$, $N \subseteq A^n$ submod
- $A \subseteq R$ algebra-finite $\iff$ $R = A[f_1, \ldots, f_n]$ for some $f_i \in R$ $\iff$ $R \cong A[x_1, \ldots, x_i]/I$, x_i indeterminates
    - $A \subseteq R$ algebra-finite $\iff$ $R = A[f_1, \ldots, f_n]$, $f_i \in R$
    - $A \subseteq R$ algebra-finite, A Noetherian $\implies$ R Noetherian ring
    - $A \subseteq R$ module-finite $\iff$ $\left\{\begin{array}{l} \text{algebra-finite} \not\Longrightarrow \\ \text{and integral} \not\Longrightarrow \end{array}\right.$ module-finite

Artin-Tate Lemma: $\underbrace{\underset{\text{Noeth}}{A} \subseteq \underbrace{B \subseteq C}_{\text{mod-fin}}}_{\text{alg-fin}}$

Chapter 2

Graded rings

2.1 Graded rings

When we think of a polynomial ring R, we often think of R with its graded structure, even if we have never formalized what that means. Other rings we have seen also have a graded structure, and this structure is actually very powerful.

Definition 2.1. A ring R is $\mathbb{N}$**-graded** if we can write a direct sum decomposition of R as an abelian group indexed by $\mathbb{N}$[1]

$$R = \bigoplus_{a \geqslant 0} R_a,$$

where $R_a R_b \subseteq R_{a+b}$ for every $a, b \in \mathbb{N}$, meaning that for any $r \in R_a$ and $s \in R_b$, we have $rs \in R_{a+b}$. More generally, given a monoid T. The ring R is T**-graded** if there exists a direct sum decomposition of R as an abelian group indexed by T:

$$R = \bigoplus_{a \in T} R_a$$

satisfying $R_a R_b \subseteq R_{a+b}$.

An element that lies in one of the summands R_a is said to be **homogeneous** of **degree** a; we write $|r|$ or $\deg(r)$ to denote the degree of a homogeneous element r.

By definition, an element in a graded ring is a *unique* sum of homogeneous elements, which we call its **homogeneous components** or **graded components**. One nice thing about graded rings is that many properties can usually be sufficiently checked on homogeneous elements, and these are often easier to deal with.

Remark 2.2. Note that whenever R is a graded ring, the multiplicative identity 1 must be a homogeneous element whose degree is the identity in T. In particular, if R is $\mathbb{N}$ or $\mathbb{Z}$-graded, then $1 \in R_0$ and R_0 is a subring of R.

Example 2.3.

a) Any ring R is trivially an $\mathbb{N}$-graded ring, by setting $R_0 = R$ and $R_n = 0$ for $n \neq 0$.

[1]We follow the convention that 0 is a natural number.

b) If k is a field and $R = k[x_1, \ldots, x_n]$ is a polynomial ring, there is an $\mathbb{N}$-grading on R called the *standard grading* where R_d is the k-vector space with basis given by the monomials of total degree d, meaning those of the form $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ with $\sum_i \alpha_i = d$. Of course, this is the notion of degree familiar from middle school. So $x_1^2 + x_2x_3$ is homogeneous in the standard grading, while $x_1^2 + x_2$ is not.

c) If k is a field, and $R = k[x_1, \ldots, x_n]$ is a polynomial ring, we can give different $\mathbb{N}$-gradings on R by fixing some tuple $(\beta_1, \ldots, \beta_n) \in \mathbb{N}^n$ and letting x_i be a homogeneous element of degree β_i; we call this a grading with *weights* $(\beta_1, \ldots, \beta_n)$.

   For example, in $k[x_1, x_2]$, $x_1^2 + x_2^3$ is not homogeneous in the standard grading, but it is homogeneous of degree 6 under the $\mathbb{N}$-grading with weights $(3, 2)$.

d) A polynomial ring $R = k[x_1, \ldots, x_n]$ also admits a natural $\mathbb{N}^n$-grading, with $R_{(d_1,\ldots,d_n)} = k \cdot x_1^{d_1} \cdots x_n^{d_n}$. This is called the *fine grading*.

e) Let $\Gamma \subseteq \mathbb{N}^n$ be a subsemigroup of $\mathbb{N}^n$. Then

$$\bigoplus_{\gamma \in \Gamma} k \cdot \underline{x}^{\gamma} \subseteq k[\underline{x}] = k[x_1, \ldots, x_n]$$

   is an $\mathbb{N}^n$-graded subring of $k[x_1, \ldots, x_n]$. Conversely, every $\mathbb{N}^n$-graded subring of $k[x_1, \ldots, x_n]$ is of this form.

f) Polynomial rings in Macaulay2 are graded with the standard grading by default. To define a different grading, we give Macaulay2 a list with the grading of each of the variables:

```
i1 : R = ZZ/101[a,b,c,Degrees=>{{1,2},{2,1},{1,0}}];
```

   We can check whether an element of R `isHomogeneous`, and the function `degree` applied to an element of R returns the least upper bound of the degrees of its monomials:

```
i2 : degree (a+b)
o2 = {2, 2}
o2 : List

i3 : isHomogeneous(a+b)
o3 = false
```

Remark 2.4. You may have seen the term *homogeneous polynomial* used to refer to a polynomial $f(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]$ that satisfies

$$f(\lambda x_1, \ldots, \lambda x_n) = \lambda^d f(x_1, \ldots, x_n)$$

for some d. This is equivalent to saying that all the terms in f have the same total degree, or that f is homogeneous with respect to the standard grading.

Similarly, a polynomial is *quasi-homogeneous*, or *weighted homogeneous*, if there exist integers $w_1, \ldots, w_n$ such that he sum $w = a_1w_1 + \cdots + a_nw_n$ is the same for all monomials $x_1^{a_1} \cdots x_n^{a_n}$ appearing in f. So f satisfies

$$f(\lambda^{w_1} x_1, \ldots, \lambda^{w_n} x_n) = \lambda^w f(x_1, \ldots, x_n),$$

and $f(x_1^{w_1}, \ldots, x_n^{w_n})$ is homogeneous (in the previous sense, so with respect to the standard grading). This condition is equivalent to asking that f be homogeneous with respect to some weighted grading on $k[x_1, \ldots, x_n]$.

Definition 2.5. An ideal I in a graded ring R is called *homogeneous* if it can be generated by homogeneous elements.

Remark 2.6. Observe that an ideal is homogeneous if and only if I has the following property: for any element $f \in R$ we have $f \in I$ if and only if every homogeneous component of f lies in I. We can repackage this by saying that I is homogeneous if

$$I = \bigoplus_{a \in T} I_a,$$

where $I_a = I \cap R_a$.

Indeed, if I has this property, take a generating set $\{f_\lambda\}_\Lambda$ for I; by assumption, all of the homogeneous components of each f_λ lie in I, and since each f_λ lies in the ideal generated by these components, the set of all the components generates I, and I is homogeneous. On the other hand, if all the components of f lie in I then so does f, whether or not I is homogeneous. If I is homogeneous and $f \in I$, write f as a combination of the (homogeneous) generators of I, say $f_1, \ldots, f_n$:

$$f = r_1 f_1 + \cdots + r_n f_n.$$

Now by writing each r_i as a sum of its components, say $r_i = r_{i,1} + \cdots + r_{i,n_i}$, each $r_{i,j} f_i \in I$, and these contain all the components of f (and potentially some redundant terms).

Example 2.7. Given an $\mathbb{N}$-graded ring R, then $R_+ = \bigoplus_{d>0} R_d$ is a homogeneous ideal.

We now observe the following:

Lemma 2.8. *Let R be an T-graded ring, and I be a homogeneous ideal. Then R/I has a natural T-graded structure induced by the T-graded structure on R.*

Proof. The ideal I decomposes as the direct sum of its graded components, so we can write

$$R/I = \frac{\oplus R_a}{\oplus I_a} \cong \oplus \frac{R_a}{I_a}. \qquad \square$$

Example 2.9.

a) The ideal $I = (w^2x + wyz + z^3, x^2 + 3xy + 5xz + 7yz + 11z^2)$ in $R = k[w, x, y, z]$ is homogeneous with respect to the standard grading on R, and thus the ring R/I admits an $\mathbb{N}$-grading with $|w| = |x| = |y| = |z| = 1$.

b) In contrast, the ring $R = k[x, y, z]/(x^2 + y^3 + z^5)$ does not admit a grading with $|x| = |y| = |z| = 1$, but does admit a grading with $|x| = 15, |y| = 10, |z| = 6$.

Definition 2.10. Let R be a T-graded ring, and M an R-module. The module R is **T-graded** if there exists a direct sum decomposition of M as an abelian group indexed by T:

$$M = \bigoplus_{a \in T} M_a \text{ such that } R_a M_b \subseteq M_{a+b}$$

for all $a, b \in T$.

The notions of homogeneous element of a module and degree of a homogeneous element of a module take the obvious meanings. A notable abuse of notation: we will often talk about $\mathbb{Z}$-graded modules over $\mathbb{N}$-graded rings, and likewise.

We can also talk about graded homomorphisms.

Definition 2.11. Let R and S be T-graded rings with the same grading monoid T. A ring homomorphism $\varphi : R \to S$ is **graded** or **degree-preserving** if $\varphi(R_a) \subseteq S_a$ for all $a \in T$.

Note that our definition of ring homomorphism requires $1_R \mapsto 1_S$, and thus it does not make sense to talk about graded ring homomorphisms of degree $d \neq 0$. But we can have graded module homomorphisms of any degree.

Definition 2.12. Let M and N be $\mathbb{Z}$-graded modules over the $\mathbb{N}$-graded ring R. A homomorphism of R-modules $\varphi : R \to S$ is **graded** if $\varphi(M_a) \subseteq N_{a+d}$ for all $a \in \mathbb{Z}$ and some fixed $d \in \mathbb{Z}$, called the **degree** of φ. A graded homomorphism of degree 0 is also called **degree-preserving**.

Example 2.13.

a) Consider the ring map $k[x, y, z] \to k[s, t]$ given by $x \mapsto s^2, y \mapsto st, z \mapsto t^2$. If $k[s, t]$ has the fine grading, meaning $|s| = (1, 0)$ and $|t| = (0, 1)$, then the given map is degree preserving if and only if $k[x, y, z]$ is graded by

$$|x| = (2, 0), |y| = (1, 1), |z| = (0, 2).$$

b) Let k be a field, and let $R = k[x_1, \dots, x_n]$ be a polynomial ring with the standard grading. Given $c \in k = R_0$, the homomorphism of R-modules $R \to R$ given by $f \mapsto cf$ is degree preserving. However, if instead we take $g \in k = R_d$ for some $d > 0$, then the map

$$\begin{aligned} R &\longrightarrow R \\ f &\longmapsto gf \end{aligned}$$

is not degree preserving, although it is a graded map of degree d. We can make this a degree-preserving map if we shift the grading on R by defining $R(-d)$ to be the R-module R but with the $\mathbb{Z}$-grading given by $R(-d)_t = R_{t-d}$. With this grading, the component of degree d of $R(-d)$ is $R(-d)_d = R_0 = k$. Now the map

$$\begin{aligned} R(-d) &\longrightarrow R \\ f &\longmapsto gf \end{aligned}$$

is degree preserving.

We observed earlier an important relationship between algebra-finiteness and Noetherianity that followed from the Hilbert basis theorem: if R is Noetherian, then any algebra-finite extension of R is also Noetherian. There isn't a converse to this in general: there are lots of algebras over fields K that are Noetherian but not algebra-finite over K. However, for graded rings, this converse relation holds.

Proposition 2.14. *Let R be an $\mathbb{N}$-graded ring, and consider homogeneous elements $f_1, \ldots, f_n \in R$ of positive degree. Then $f_1, \ldots, f_n$ generate the ideal $R_+ := \bigoplus_{d>0} R_d$ if and only if $f_1, \ldots, f_n$ generate R as an R_0-algebra.*

Therefore, an $\mathbb{N}$-graded ring R is Noetherian if and only if R_0 is Noetherian and R is algebra-finite over R_0.

Proof. If $R = R_0[f_1, \ldots, f_n]$, then any element $r \in R_+$ can be written as a polynomial expression $r = P(f_1, \ldots, f_n)$ for some $P \in R_0[\underline{x}]$ with no constant term. Each monomial of P is a multiple of some x_i, and thus $r \in (f_1, \ldots, f_n)$.

To show that $R_+ = (f_1, \ldots, f_n)$ implies $R = R_0[f_1, \ldots, f_n]$, it suffices to show that any homogeneous element $r \in R$ can be written as a polynomial expression in the f's with coefficients in R_0. We induce on the degree of r, with degree 0 as a trivial base case. For r homogeneous of positive degree, we must have $r \in R_+$, so by assumption we can write $r = a_1 f_1 + \cdots + a_n f_n$; moreover, since r and $f_1, \ldots, f_n$ are all homogeneous, we can choose each coefficient a_i to be homogeneous of degree $|r| - |f_i|$. By the induction hypothesis, each a_i is a polynomial expression in the f's, so we are done.

For the final statement, if R_0 is Noetherian and R algebra-finite over R_0, then R is Noetherian by the Hilbert Basis Theorem. If R is Noetherian, then $R_0 \cong R/R_+$ is Noetherian. Moreover, R is algebra-finite over R_0 since R_+ is generated as an ideal by finitely many homogeneous elements by Noetherianity, so by the first statement, we get a finite algebra generating set for R over R_0. $\square$

There are many interesting examples of $\mathbb{N}$-graded algebras with $R_0 = k$; in that case, R_+ is the largest homogeneous ideal in R. In fact, R_0 is the only maximal ideal of R that is also homogeneous, so we can call it *the* **homogeneous maximal ideal**; it is sometimes also called the **irrelevant maximal ideal** of R. This ideal plays a very important role — in many ways, R and R_+ behave similarly to a local ring R and its unique maximal ideal. We will discuss this further when we learn about local rings.

2.2 Another application to invariant rings

If R is a graded ring, and G is a group acting on R by degree-preserving automorphisms, then R^G is a graded subring of R, meaning R^G is graded with respect to the same grading monoid. In particular, if G acts k-linearly on a polynomial ring over k, the invariant ring is $\mathbb{N}$-graded.

Using this perspective, we can now give a different proof of the finite generation of invariant rings that works under different hypotheses. The proof we will discuss now is essentially Hilbert's proof. To do that, we need another notion that is very useful in commutative algebra.

Definition 2.15. Let S be an R-algebra corresponding to the ring homomorphism $\varphi : R \to S$. We say that R is a **direct summand** of S if the map φ **splits** as a map of R-modules, meaning there is an R-module homomorphism

$$\begin{tikzcd} R \arrow[r, "\varphi"'] & S \arrow[l, bend right, dashed, "\pi"'] \end{tikzcd}$$

such that $\pi\varphi$ is the identity on R.

First, observe that the condition on π implies that φ must be injective, so we can assume that $R \subseteq S$, perhaps after renaming elements. Then the condition on π is that $\pi(rs) = r\pi(s)$ for all $r \in R$ and $s \in S$ and that $\pi|_R$ is the identity. We call the map π the *splitting* of the inclusion. Note that given any R-linear map $\pi : S \to R$, if $\pi(1) = 1$ then π is a splitting: indeed, $\pi(R) = \pi(r \cdot 1) = r\pi(1) = r$ for all $r \in R$.

Being a direct summand is really nice, since many good properties of S pass onto its direct summands.

Notation 2.16. Let $R \subseteq S$ be an extension of rings. Given an ideal I in S, we write $I \cap R$ for the **contraction** of R back into R, meaning the preimage of I via the inclusion map $R \subseteq S$. More generally, we may use the notation $I \cap R$ to denote the preimage of I via a given ring map $R \to S$, even if the map is not injective.

Given a ring map $R \to S$, and an ideal I in R, the **expansion** of I in S is the ideal of S generated by the image of I via the given ring map; we naturally denote this by IS.

Lemma 2.17. *Let R be a direct summand of S. Then, for any ideal $I \subseteq R$, we have $IS \cap R = I$.*

Proof. Let π be the corresponding splitting. Clearly, $I \subseteq IS \cap R$. Conversely, if $r \in IS \cap R$, we can write $r = s_1 f_1 + \cdots + s_t f_t$ for some $f_i \in I$, $s_i \in S$. Applying π, we have

$$r = \pi(r) = \pi\left(\sum_{i=1}^{t} s_i f_i\right) = \sum_{i=1}^{t} \pi\left(s_i f_i\right) = \sum_{i=1}^{t} \pi\left(s_i\right) f_i \in I.$$

□

Proposition 2.18. *Let R be a direct summand of S. If S is Noetherian, then so is R.*

Proof. Let

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

be a chain of ideals in R. The chain of ideals in S

$$I_1 S \subseteq I_2 S \subseteq I_3 S \subseteq \cdots$$

stabilizes, so there exist J, N such that $I_n R = J$ for $n \geqslant N$. Contracting to R, we get that $I_n = I_n S \cap R = J \cap R$ for $n \geqslant N$, so the original chain also stabilizes. □

Proposition 2.19. *Let k be a field, and R be a polynomial ring over k. Let G be a finite group acting k-linearly on R. Assume that the characteristic of k does not divide $|G$. Then R^G is a direct summand of R.*

Remark 2.20. The condition that the characteristic of k does not divide the order of G is trivially satisfied if k has characteristic zero.

Proof. We consider the map $\rho : R \to R^G$ given by

$$\rho(r) = \frac{1}{|G|} \sum_{g \in G} g \cdot r.$$

First, note that the image of this map lies in R^G, since acting by g just permutes the elements in the sum, so the sum itself remains the same. We claim that this map ρ is a splitting for the inclusion $R^G \subseteq R$. To see that, let $s \in R^G$ and $r \in R$. We have

$$\rho(sr) = \frac{1}{|G|} \sum_{g \in G} g \cdot (sr) = \frac{1}{|G|} \sum_{g \in G} (g \cdot s)(g \cdot r) = \frac{1}{|G|} \sum_{g \in G} s(g \cdot r) = s \frac{1}{|G|} \sum_{g \in G} (g \cdot r) = s\rho(r),$$

so ρ is R^G-linear, and for $s \in R^G$,

$$\rho(s) = \frac{1}{|G|} \sum_{g \in G} g \cdot s = s.$$

□

Theorem 2.21 (Hilbert's finiteness theorem for invariants)**.** *Let k be a field, and R be a polynomial ring over k. Let G be a group acting k-linearly on R. Assume that G is finite and $|G|$ does not divide the characteristic of k, or more generally, that R^G is a direct summand of R. Then R^G is a finitely generated k-algebra.*

Proof. Since G acts linearly, R^G is an $\mathbb{N}$-graded subring of R with $R_0 = k$. Since R^G is a direct summand of R, R^G is Noetherian by Proposition 2.18. By our characterization of Noetherian graded rings, R^G is finitely generated over $R_0 = k$. □

One important thing about this proof is that it applies to many infinite groups. In particular, for any *linearly reductive group*, including $\mathrm{GL}_n(\mathbb{C})$, $\mathrm{SL}_n(\mathbb{C})$, and $(\mathbb{C}^\times)^n$, we can construct a splitting map ρ.

Chapter 3

Algebraic Geometry

Colloquially, we often identify systems of equations with their solution sets. We will make this correspondence more precise for systems of polynomial equations, and develop the beginning of a rich dictionary between algebraic and geometric objects.

Question 3.1. Let k be a field. To what extent is a system of polynomial equations

$$\begin{cases} f_1 = 0 \\ \quad\vdots \\ f_t = 0 \end{cases}$$

where polynomials $f_1, \dots, f_t \in k[x_1, \dots, x_d]$, determined by its solution set?

Let's consider one polynomial equation in one variable. Over $\mathbb{R}, \mathbb{Q}$, or other fields that are not algebraically closed, there are many polynomials with an empty solution set; for example, $z^2 + 1$ has an empty solution set over $\mathbb{R}$. On the other hand, over $\mathbb{C}$, or any algebraically closed field, if $a_1, \dots, a_d$ are the solutions to $f(z) = 0$, we know that we can write f in the form $f(z) = \alpha(z - a_1)^{n_1} \cdots (z - a_d)^{n_d}$, so f is completely determined up to scalar multiple and repeated factors. If we insist that f have no repeated factors, then (f) is uniquely determined.

More generally, given any system of polynomial equations

$$\begin{cases} f_1 = 0 \\ \quad\vdots \\ f_t = 0 \end{cases}$$

where $f_i \in k[z]$ for some field k, notice that that $z = a$ is a solution to the system if and only if it is a solution for any polynomial $g \in (f_1, \dots, f_t)$. But since $k[z]$ is a PID, we have $(f_1, \dots, f_t) = (f)$, where f is a greatest common divisor of $f_1, \dots, f_t$. Therefore, $z = a$ is a solution to the system if and only if $f(a) = 0$.

3.1 Varieties

Definition 3.2. Given a field k, the *affine d-space over k*, denoted $\mathbb{A}_k^d$, is the set

$$\mathbb{A}_k^d = \{(a_1, \dots, a_d) \mid a_i \in k\}.$$

Definition 3.3. For a subset T of $k[x_1, \ldots, x_d]$, we define $\mathcal{Z}(T) \subseteq \mathbb{A}^d_k$ to be the set of common zeros or the *zero set* of the polynomials (equations) in T:

$$\mathcal{Z}(T) = \{(a_1, \ldots, a_d) \in \mathbb{A}^d_k \mid f(a_1, \ldots, a_d) = 0 \text{ for all } f \in T\}.$$

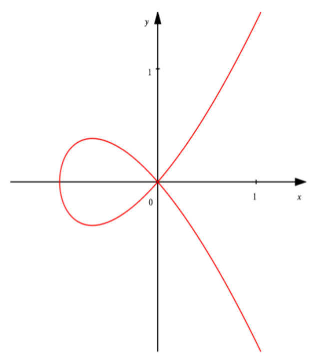
Sometimes, in order to emphasize the role of k, we will write this as $\mathcal{Z}_k(T)$.

A subset of $\mathbb{A}^d_k$ of the form $\mathcal{Z}(T)$ for some subset T is called an **algebraic subset** of $\mathbb{A}^d_k$, or an **affine algebraic variety**. So a variety $\mathbb{A}^d_k$ is the set of common solutions of some (possibly infinite) collection of polynomial equations. A variety is **irreducible** if it cannot be written as the union of two proper varieties.

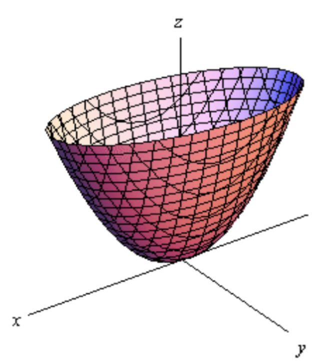
Note that some authors use the word *variety* to refer only to irreducible algebraic sets. Note also that the definitions given here are only completely standard when k is algebraically closed.

Example 3.4. Here are some simple examples of algebraic varieties:

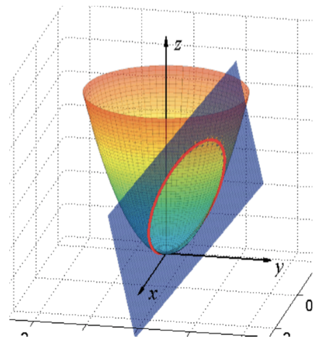
a) For $k = \mathbb{R}$ and $n = 2$, $\mathcal{Z}(y^2 + x^2(x-1))$ is a "nodal curve" in $\mathbb{A}^2_{\mathbb{R}}$, the real plane. Note that we've written x for x_1 and y for x_2 here.



b) For $k = \mathbb{R}$ and $n = 3$, $\mathcal{Z}(z - x^2 - y^2)$ is a paraboloid in $\mathbb{A}^3_{\mathbb{R}}$, real three space.



c) For $k = \mathbb{R}$ and $n = 3$, $\mathcal{Z}(z - x^2 - y^2, 3x - 2y + 7z - 7)$ is circle in $\mathbb{A}^3_{\mathbb{R}}$.



d) For $k = \mathbb{R}$, $\mathcal{Z}_{\mathbb{R}}(x^2 + y^2 + 1) = \emptyset$. Note that $\mathcal{Z}_{\mathbb{C}}(x^2 + y^2 + 1) \neq \emptyset$.

e) The subset $\mathbb{A}^2_k \setminus \{(0,0)\}$ is not an algebraic subset of $\mathbb{A}^2_k$ if k is infinite. Why?

f) The graph of the sine function is not an algebraic subset of $\mathbb{A}^2_{\mathbb{R}}$. Why not?

g) For $k = \mathbb{R}$, $\mathcal{Z}(y - x^2, xz - y^2, z - xy)$ is the so-called **twisted cubic (affine) curve**. It is the curve parametrized by (t, t^2, t^3), meaning it is the image of the map

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{R}^3 \\ t &\longmapsto (t, t^2, t^3) \end{aligned} \quad .$$

We can check this with Macaulay2:

```
i1 : k = RR;

i2 : R = k[x,y,z];

i3 : f = map(k[t],R,{t,t^2,t^3});

i4 : ker f

               2                    2
o4 = ideal (y  - x*z, x*y - z, x  - y)

o4 : Ideal of R
```

So in our computation above, f sets $x = t$, $y = t^2$, and $z = t^3$, and its kernel consists precisely of the polynomials that vanish at every point of this form. Note that computations over the reals in Macaulay2 are experimental, and yet we obtain the correct answer; we can also run the same computation over $k = \mathbb{Q}$.

h) For any field k and elements $a_1, \ldots, a_d \in k$, we have

$$\mathcal{Z}(x_1 - a_1, \ldots, x_d - a_d) = \{(a_1, \ldots, a_d)\}.$$

So, all one element subsets of $\mathbb{A}^d_k$ are algebraic subsets.

We can consider the equations that a subset of affine space satisfies.

Definition 3.5. Given any subset X of $\mathbb{A}_k^d$ for a field k, define

$$\mathcal{I}(X) = \{g(x_1, \dots, x_d) \in k[x_1, \dots, x_d] \mid g(a_1, \dots, a_d) = 0 \text{ for all } (a_1, \dots, a_d) \in X\}.$$

Exercise 5. $\mathcal{I}(X)$ is an ideal in $k[x_1, \dots, x_d]$ for any $X \subseteq \mathbb{A}_k^d$.

Example 3.6.

a) $\mathcal{I}(\{(a_1, \dots, a_d)\}) = (x_1 - a_1, \dots, x_d - a_d)$, for any field k.

b) $\mathcal{I}(\text{graph of the sine function in } \mathbb{A}_{\mathbb{R}}^2) = (0)$.

Exercise 6. Here are some properties of the functions $\mathcal{Z}$ and $\mathcal{I}$:

a) For any field, we have $\mathcal{Z}(0) = \mathbb{A}_k^n$ and $\mathcal{Z}(1) = \emptyset$.

b) $\mathcal{I}(\emptyset) = (1) = k[x_1, \dots, x_d]$ (the improper ideal).

c) $\mathcal{I}(\mathbb{A}_k^d) = (0)$ if and only if k is infinite.

d) If $I \subseteq J \subseteq k[x_1, \dots, x_d]$ then $\mathcal{Z}(I) \supseteq \mathcal{Z}(J)$.

e) If $S \subseteq T$ are subsets of $\mathbb{A}_k^n$ then $\mathcal{I}(S) \supseteq \mathcal{I}(T)$.

f) If $I = (T)$ is the ideal generated by the elements of $T \subseteq k[x_1, \dots, x_d]$, then $\mathcal{Z}(T) = \mathcal{Z}(I)$.

So we will talk about the solution set of an ideal, rather than of an arbitrary set. Hilbert's Basis Theorem implies that every ideal in $k[x_1, \dots, x_d]$ is finitely generated, so any system of equations in $k[x_1, \dots, x_d]$ can be replaced with a system of *finitely many* equations.

Example 3.7. Let

$$X = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}$$

be a 2×3 matrix of variables — we usually call these *generic* matrices — and let

$$R = k[X] = k\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}.$$

Let $\Delta_1, \Delta_2, \Delta_3$ the 2×2-minors of X. Consider the ideal $I = (\Delta_1, \Delta_2, \Delta_3)$. Thinking of these generators as equations, a solution to the system corresponds to a choice of 2×3 matrix whose 2×2 minors all vanish — that is, a matrix of rank at most one. So $\mathcal{Z}(I)$ is the set of rank at most one matrices. Note that $I \subseteq (x_1, x_2, x_3) =: J$, and $\mathcal{Z}(J)$ is the set of 2×3 matrices with top row zero. The containment $\mathcal{Z}(J) \subseteq \mathcal{Z}(I)$ we obtain from $I \subseteq J$ translates to the fact that a 2×3 matrix with a zero row has rank at most 1.

Finally, the union and intersection of varieties is also a variety.

Exercise 7. Suppose that I and J are ideals in $k[x_1, \dots, x_d]$.

a) $\mathcal{Z}(I) \cap \mathcal{Z}(J) = \mathcal{Z}(I + J)$.

b) $\mathcal{Z}(I) \cup \mathcal{Z}(J) = \mathcal{Z}(I \cap J) = \mathcal{Z}(IJ)$.

However, note that in general $IJ \neq I \cap J$.

3.2 Prime and maximal ideals

Before we talk more about geometry, let's recall some basic facts about prime and maximal ideals. As we will discover through the rest of the course, prime ideals play a very prominent role in commutative algebra.

Definition 3.8. An ideal $P \neq R$ is **prime** if $ab \in P$ implies $a \in P$ or $b \in P$.

Exercise 8. An ideal P in a ring R is prime if and only if R/P is a domain.

Example 3.9. The prime ideals in $\mathbb{Z}$ are those of the form (p) for p a prime integer, and (0).

Example 3.10. When k is a field, in $k[x]$ are easy to describe: $k[x]$ is a principal ideal domain, and $(f) \neq 0$ is prime if and only if f is an irreducible polynomial. Moreover, (0) is also a prime ideal, since $k[x]$ is a domain.

The prime ideals in $k[x_1, \ldots, x_d]$ are, however, not so easy to describe. We will see many examples throughout the course; here are some.

Example 3.11. The ideal $P = (x^3 - y^2)$ in $R = k[x, y]$ is prime; one can show that $R/P \cong k[t^2, t^3] \subseteq k[t]$, which is a domain.

Example 3.12. The k-algebra $R = k[s^3, s^2t, st^2, t^2] \subseteq k[s, t]$ is a domain, so its defining ideal I in $k[x_1, x_2, x_3, x_4]$ is prime. This is the kernel of the presentation of R sending x_1, x_2, x_3, x_4 to each of our 4 algebra generators, which we can compute with Macaulay2:

```
1 : k = QQ;

i2 : f = map(k[s,t],k[x_1 .. x_4],{s^3,s^2*t,s*t^2,t^3})
                                3   2      2   3
o2 = map(QQ[s..t],QQ[x ..x ],{s , s t, s*t , t })
                      1   4

o2 : RingMap QQ[s..t] <--- QQ[x ..x ]
                                1   4
i3 : I = ker f
               2                        2
o3 = ideal (x  - x x , x x  - x x , x  - x x )
             3    2 4   2 3    1 4   2    1 3

o3 :  Ideal of QQ[x ..x]
                   1   4
```

Later we will show that prime ideals correspond to irreducible varieties; more precisely, that X is irreducible if and only if $\mathcal{I}(X)$ is prime.

Definition 3.13 (maximal ideal)**.** An ideal $\mathfrak{m}$ in R is **maximal** if for any ideal I

$$I \supseteq \mathfrak{m} \implies I = \mathfrak{m} \text{ or } I = R.$$

Exercise 9. An ideal $\mathfrak{m}$ in R is maximal if and only if $R/\mathfrak{m}$ is a field.

Given a maximal ideal $\mathfrak{m}$ in R, the **residue field** of $\mathfrak{m}$ is the field $R/\mathfrak{m}$. A field k is a residue field of R if $k \cong R/\mathfrak{m}$ for some maximal ideal $\mathfrak{m}$.

Remark 3.14. A ring may have many different residue fields. For example, the residue fields of $\mathbb{Z}$ are all the finite fields with prime many elements, $\mathbb{F}_p \cong \mathbb{Z}/p$.

Exercise 10. Every maximal ideal is prime.

However, not every prime ideal is maximal. For example, in $\mathbb{Z}$, (0) is a prime ideal that is not maximal.

Theorem 3.15. *Given a ring R, every proper ideal $I \neq R$ is contained in some maximal ideal.*

Fun fact: this is actually *equivalent* to the Axiom of Choice. We will prove it (but not its equivalence to the Axiom of Choice!) using Zorn's Lemma, another equivalent version of the Axiom of Choice. Zorn's Lemma says that

> Every non-empty partially ordered set in which every chain (i.e., totally ordered subset) has an upper bound contains at least one maximal element.

So let's prove that every ideal is contained in some maximal ideal.

Proof. First, we will show that Zorn's Lemma applies to proper ideals in any ring R. The statement will then follow by applying Zorn's Lemma to the non-empty set of ideals $J \supseteq I$, which is partially ordered by inclusion.

So consider a chain of proper ideals in R, say $\{I_i\}_i$. Now $I = \bigcup_i I_i$ is an ideal as well, and $I \neq R$ since $1 \notin I_i$ for all i. Note that unions of ideals are not ideals in general, but a union of totally ordered ideals *is* an ideal. Then I is an upper bound for our chain $\{I_i\}_i$, and Zorn's Lemma applies to the set of proper ideals in R with inclusion $\subseteq$. $\square$

3.3 Nullstellensatz

Lemma 3.16. *Let k be a field, and $R = k[x_1, \ldots, x_d]$ be a polynomial ring. There is a bijection*

$$\begin{array}{ccc} \mathbb{A}^d_k & \longrightarrow & \left\{ \begin{array}{c} \textit{maximal ideals } \mathfrak{m} \textit{ of } R \\ \textit{with } R/\mathfrak{m} \cong k \end{array} \right\} \\ & & \\ (a_1, \ldots, a_d) & \longmapsto & (x_1 - a_1, \ldots, x_d - a_d) \end{array}$$

Proof. Each $\mathfrak{m} = (x_1 - a_1, \ldots, x_d - a_d)$ is a maximal ideal satisfying $R/\mathfrak{m} \cong k$. Moreover, these ideals are distinct: if $x_i - a_i, x_i - a_i'$ are in the same ideal for $a_i \neq a_i'$, then the unit $a_i - a_i'$ is in the ideal, so it is not proper. Therefore, our map is injective. To see that it is surjective, let $\mathfrak{m}$ be a maximal ideal with $R/\mathfrak{m} \cong k$. Each class in $R/\mathfrak{m}$ corresponds to a unique $a \in k$, so in particular each x_i is in the class of a unique $a_i \in k$. This means that $x_i - a_i \in \mathfrak{m}$, and thus $(x_1 - a_1, \ldots, x_d - a_d) \subseteq \mathfrak{m}$. Since $(x_1 - a_i, \ldots, x_d - a_i)$ is a maximal ideal, we must have $(x_1 - a_1, \ldots, x_d - a_d) = \mathfrak{m}$. $\square$

Example 3.17. Not all maximal ideals in $k[x_1, \ldots, x_d]$ are necessarily of this form. For example, if $k = \mathbb{R}$ and $d = 1$, the ideal $(x^2 + 1)$ is maximal, but

$$k[x]/(x^2 + 1) \cong \mathbb{C} \ncong k.$$

But this won't happen if k is algebraically closed.

Theorem 3.18 (Zariski's Lemma)**.** *Consider an extension of fields $k \subseteq L$. If L is a finitely generated k-algebra, then L is a finite dimensional k-vector space. In particular, if k is algebraically closed then $L = k$.*

This is a nice application of the Artin-Tate Lemma, together with some facts about transcendent elements. We will skip the proof, but you can find it in Jeffries' notes.

Corollary 3.19 (Nullstellensatz)**.** *Let $S = k[x_1, \ldots, x_d]$ be a polynomial ring over an algebraically closed field k. There is a bijection*

$$\begin{array}{ccc} \mathbb{A}^d_k & \longrightarrow & \{\text{maximal ideals } \mathfrak{m} \text{ of } S\} \\ (a_1, \ldots, a_d) & \longmapsto & (x_1 - a_1, \ldots, x_d - a_d) \end{array}$$

If R is a finitely generated k-algebra, we can write $R = S/I$ for a polynomial ring S, and there is an induced bijection

$$\mathcal{Z}_k(I) \subseteq \mathbb{A}^d_k \longleftrightarrow \{\text{maximal ideals } \mathfrak{m} \text{ of } R\}.$$

Proof. The first part follows immediately from Lemma 3.16 and Lemma 3.18.

To show the second statement, fix an ideal I in S, and $R = S/I$. The maximal ideal ideals in R are in bijection with the maximal ideals $\mathfrak{m}$ in S that contain I; those are the ideals of the form $(x_1 - a_1, \ldots, x_d - a_d)$ with $I \subseteq (x_1 - a_1, \ldots, x_d - a_d)$. These are in bijection with the points $(a_1, \ldots, a_d) \in \mathbb{A}^d_k$ satisfying $(a_1, \ldots, a_d) \in \mathcal{Z}_k(I)$. $\square$

Theorem 3.20 (Weak Nullstellensatz)**.** *Let k be an algebraically closed field. If I is a proper ideal in $R = k[x_1, \ldots, x_d]$, then $\mathcal{Z}_k(I) \neq \varnothing$.*

Proof. If $I \subseteq R$ is a proper ideal, there is a maximal ideal $\mathfrak{m} \supseteq I$, so $\mathcal{Z}(\mathfrak{m}) \subseteq \mathcal{Z}(I)$. Since $\mathfrak{m} = (x_1 - a_1, \ldots, x_d - a_d)$ for some $a_i \in k$, $\mathcal{Z}(\mathfrak{m})$ is a point, and thus nonempty. $\square$

Over an algebraically closed field, maximal ideals in $k[x_1, \ldots, x_d]$ correspond to points in $\mathbb{A}^d$. So we can start from the solution set — a point — and recover an ideal that corresponds to it. What if we start with some non-maximal ideal I, and consider its solution set $\mathcal{Z}_k(I)$ — can we recover I in some way?

Example 3.21. Many ideals define the same solution set. For example, in $R = k[x]$, the ideals $I_n = (x^n)$, for any $n \geqslant 1$, all define the same solution set $\mathcal{Z}_k(I_n) = \{0\}$.

To attack this question, we will need an observation on inequations.

Remark 3.22 (Rabinowitz's trick)**.** Observe that, if $f(\underline{x})$ is a polynomial and $\underline{a} \in \mathbb{A}^d$, $f(\underline{a}) \neq 0$ if and only if $f(\underline{a}) \in k$ is invertible; equivalently, if there is a solution $y = b \in k$ to $yf(\underline{a}) - 1 = 0$. In particular, a system of polynomial equations and inequations

$$\begin{cases} f_1(\underline{x}) = 0 \\ \vdots \\ f_m(\underline{x}) = 0 \end{cases} \text{and} \begin{cases} g_1(\underline{x}) \neq 0 \\ \vdots \\ g_n(\underline{x}) \neq 0 \end{cases}$$

has a solution $\underline{x} = \underline{a}$ if and only if the system

$$\begin{cases} f_1(\underline{x}) = 0 \\ \vdots \\ f_m(\underline{x}) = 0 \end{cases} \text{and} \begin{cases} y_1 g_1(\underline{x}) - 1 = 0 \\ \vdots \\ y_n g_n(\underline{x}) - 1 = 0 \end{cases}$$

has a solution $(\underline{x}, \underline{y}) = (\underline{a}, \underline{b})$. In fact, this is equivalent to a system in one extra variable:

$$\begin{cases} f_1(\underline{x}) = 0 \\ \vdots \\ f_m(\underline{x}) = 0 \\ y g_1(\underline{x}) \cdots g_n(\underline{x}) - 1 = 0 \end{cases}$$

Theorem 3.23 (Strong Nullstellensatz)**.** *Let k be an algebraically closed field, and $R = k[x_1, \ldots, x_d]$ be a polynomial ring. Let $I \subseteq R$ be an ideal. The polynomial f vanishes on $\mathcal{Z}_k(I)$ if and only if $f^n \in I$ for some $n \in \mathbb{N}$.*

Proof. Suppose that $f^n \in I$. For each $\underline{a} \in \mathcal{Z}_k(I)$, $f(\underline{a}) \in k$ satisfies $f(\underline{a})^n = 0 \in k$. Since k is a field, $f(\underline{a}) = 0$. Thus, $f \in \mathcal{Z}_k(I)$ as well.

Suppose that f vanishes along $\mathcal{Z}_k(I)$. This means that given any solution $\underline{a} \in \mathbb{A}^d$ to the system determined by I, $f(\underline{a}) = 0$. In other words, the system

$$\begin{cases} g(\underline{x}) = 0 \text{ for all } g \in I \\ f \neq 0 \end{cases}$$

has no solutions. By the discussion above, $\mathcal{Z}_k(I + (yf - 1)) = \varnothing$ in a polynomial ring in one more variable. By the Weak Nullstellensatz, we have $IR[y] + (yf - 1) = R[y]$, and equivalently $1 \in IR[y] + (yf - 1)$. Write $I = (g_1(\underline{x}), \ldots, g_m(\underline{x}))$, and

$$1 = r_0(\underline{x}, y)(1 - yf(\underline{x})) + r_1(\underline{x}, y)g_1(\underline{x}) + \cdots + r_m(\underline{x}, y)g_m(\underline{x}).$$

We can map y to $1/f$ to get

$$1 = r_1(\underline{x}, 1/f)g_1(\underline{x}) + \cdots + r_m(\underline{x}, 1/f)g_m(\underline{x})$$

in the fraction field of $R[y]$. Since each r_i is polynomial, there is a largest negative power of f occurring; say that f^n serves as a common denominator. We can multiply by f^n to obtain f^n as a polynomial combination of the g's. $\square$

Definition 3.24. The **radical** of an ideal I in a ring R is the ideal

$$\sqrt{I} := \{f \in R \mid f^n \in I \text{ for some } n\}.$$

An ideal is a **radical ideal** if $I = \sqrt{I}$.

To see that $\sqrt{I}$ is an ideal, note that if $f^m, g^n \in I$, then

$$\begin{aligned}(f+g)^{m+n-1} &= \sum_{i=0}^{m+n-1} \binom{m+n-1}{i} f^i g^{m+n-1-i} \\ &= f^m \left(f^{n-1} + \binom{m+n-1}{1} f^{n-2} g + \cdots + \binom{m+n-1}{n-1} g^{n-1} \right) \\ &+ g^n \left(\binom{m+n-1}{n} f^{m-1} + \binom{m+n-1}{n+1} f^{m-2} g + \cdots + g^{m-1} \right) \in I,\end{aligned}$$

and $(rf)^m = r^m f^m \in I$.

Example 3.25. Prime ideals are radical.

Exercise 11. A nonzero element $f \in R$ is **nilpotent** if $f^n = 0$ for some $n > 1$; a ring R is **reduced** if it has no nilpotent elements. If R is a ring and I an ideal, then R/I is reduced if and only if I is a radical ideal.

Using this terminology, we can rephrase the Strong Nullstellensatz: if $k = \overline{k}$, then $f \in \mathcal{I}(\mathcal{Z}_k(I))$ if and only if $f \in \sqrt{I}$. Given any ideal I in $k[x_1, \dots, x_d]$, $\mathcal{I}(\mathcal{Z}(I)) = \sqrt{I}$.

We can now associate a ring to each subvariety of $\mathbb{A}^d$.

Definition 3.26. Let k be an algebraically closed field, and $X = \mathcal{Z}_k(I) \subseteq \mathbb{A}^d$ be a subvariety of $\mathbb{A}^d$. The **coordinate ring** of X is the ring $k[X] := k[x_1, \dots, x_d]/\mathcal{I}(X)$.

Since $k[X]$ is obtained from the polynomial ring on the ambient $\mathbb{A}^d$ by quotienting out by exactly those polynomials that are zero on X, we interpret $k[X]$ as the ring of polynomial functions on X. Note that every reduced finitely generated k-algebra is a coordinate ring of some zero set X.

Remark 3.27. We showed before that $\mathcal{Z}(IJ) = \mathcal{Z}(I \cap J)$, despite the fact that we often have $IJ \neq I \cap J$. The Strong Nullstellensatz implies that $\sqrt{IJ} \neq \sqrt{I+J}$.

Remark 3.28. Observe that $\mathcal{Z}_k(\sqrt{J}) = \mathcal{Z}_k(J)$ whether or not k is algebraically closed, by the same proof we used above. The containment $\subseteq$ is immediate since $J \subseteq \sqrt{J}$ from the definition. Moreover, if $f^n(\underline{a}) = 0$ then $f(\underline{a}) = 0$, so if $\underline{a} \in \mathcal{Z}_k(J)$ and $f \in \sqrt{J}$ then $f(\underline{a}) = 0$, and the equality of sets follows.

What might fail when the field is not algebraically closed is that $\mathcal{I}(\mathcal{Z}(I))$ is not necessarily $\sqrt{I}$. For example, $\mathcal{Z}_{\mathbb{R}}(x^2+1) = \varnothing$, so

$$\mathcal{I}(\mathcal{Z}_{\mathbb{R}}(x^2+1)) = \mathcal{I}(\emptyset) = \mathbb{R}[x] \neq \sqrt{(x^2+1)} = (x^2+1).$$

In fact, the ingredient that is missing is precisely the fact that the Weak Nullstellensatz is not satisfied over non-algebraically closed fields. If k is not algebraically closed, there exists some irreducible polynomial $f \in k[x]$ with no roots, so $\mathcal{Z}(f) = \varnothing$.

Remark 3.29. Note that if I is a radical ideal and $I \subsetneq J$, then $\mathcal{Z}(J) \subsetneq \mathcal{Z}(I)$. Indeed, there is some $f \in J$ such that $f \notin \sqrt{I} = I$, and thus $\mathcal{Z}(I) \nsubseteq \mathcal{Z}(f)$. Since $\mathcal{Z}(J) \subseteq \mathcal{Z}(f)$, we conclude that $\mathcal{Z}(I) \nsubseteq \mathcal{Z}(J)$.

Each variety corresponds to a unique radical ideal.

Corollary 3.30. *Let k be an algebraically closed field and $R = k[x_1, \ldots, x_d]$ a polynomial ring. There is an order-reversing bijection between the collection of subvarieties of $\mathbb{A}_k^d$ and the collection of radical ideals of R:*

$$\begin{array}{ccc} \{\textit{subvarieties of } \mathbb{A}_k^d\} & \longleftrightarrow & \{\textit{radical ideals } I \subseteq R\} \\ X & \xmapsto{\mathcal{I}} & \{f \in R \mid X \subseteq \mathcal{Z}_k(f)\} \\ \mathcal{Z}_k(I) & \xleftarrow{\mathcal{Z}} & I \end{array}$$

In particular, given ideals I and J, we have $\mathcal{Z}_k(I) = \mathcal{Z}_k(J)$ if and only if $\sqrt{I} = \sqrt{J}$.

Proof. The Strong Nullstellensatz says that $\mathcal{I}(\mathcal{Z}(J)) = \sqrt{J}$ for any ideal J, hence $\mathcal{I}(\mathcal{Z}(J)) = J$ for a radical ideal J. Conversely, given X we can write $X = \mathcal{Z}_k(J)$ for some ideal J, and we without loss of generality we can assume J is radical, since $\mathcal{Z}_k(J) = \mathcal{Z}_k(\sqrt{J})$. Then $\mathcal{Z}(\mathcal{I}(X)) = \mathcal{Z}(\mathcal{I}(\mathcal{Z}(J))) = \mathcal{Z}(J) = X$.

This shows that $\mathcal{I}$ and $\mathcal{Z}$ are inverse operations, and we are done. $\square$

Under this bijection, irreducible varieties correspond to prime ideals.

Lemma 3.31. *A variety $X \subseteq \mathbb{A}_k^d$ is irreducible if and only if $\mathcal{I}(X)$ is prime.*

Proof. Suppose that X is reducible, say $X = V_1 \cup V_2$ for two varieties V_1 and V_2 such that $V_1, V_2 \subsetneq X$. Note that this implies that $\mathcal{I}(X) \subsetneq \mathcal{I}(V_1)$, $\mathcal{I}(X) \subsetneq \mathcal{I}(V_2)$, and $\mathcal{I}(X) = \mathcal{I}(V_1) \cap \mathcal{I}(V_2)$. Then we can find $f \in \mathcal{I}(V_1)$ such that $f \notin \mathcal{I}(V_2)$, and $g \in \mathcal{I}(V_2)$ such that $g \notin \mathcal{I}(V_1)$. Notice that by construction $fg \in \mathcal{I}(V_1) \cap \mathcal{I}(V_2)\mathcal{I}(X)$, while $f \notin \mathcal{I}(X)$ and $g \notin \mathcal{I}(X)$. Therefore, $\mathcal{I}(X)$ is not prime.

Now assume that $\mathcal{I}(X)$ is not prime, and fix $f, g \notin \mathcal{I}(X)$ with $fg \in \mathcal{I}(X)$. Then

$$X \subseteq \mathcal{Z}(fg) = \mathcal{Z}(f) \cup \mathcal{Z}(g).$$

The intersections

$$V_f = \mathcal{Z}(f) \cap X = \mathcal{Z}(\mathcal{I}(X) + (f))$$

and

$$V_g = \mathcal{Z}(g) \cap X = \mathcal{Z}(\mathcal{I}(X) + (g))$$

are varieties, and $X = V_f \cup V_g$. Finally, since $f \notin \mathcal{I}(X)$, then $X \nsubseteq V_f$. Similarly, $X \nsubseteq V_g$. Thus X is reducible. $\square$

Given a variety X, we can decompose it in irreducible components by writing it as a union $X = V_1 \cup \cdots \cup V_n$. We can do this decomposition algebraically, by considering the radical ideal $I = \mathcal{I}(X)$ and writing it as an intersection of its minimal primes.

Definition 3.32. A prime P is a **minimal prime** of an ideal I if the only prime Q with $I \subseteq Q \subseteq P$ is $Q = P$. The set of minimal primes over I is denoted $\operatorname{Min}(I)$.

Soon we will show that

$$\sqrt{I} = \bigcap_{\substack{P \text{ prime} \\ P \supseteq I}} P = \bigcap_{P \in \operatorname{Min}(I)} P.$$

Later, we will prove that the set of minimal primes of an ideal in a Noetherian ring is finite, so in particular we can write $\mathcal{I}(X)$ as a finite intersection of prime ideals, say

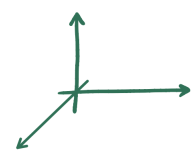
$$\mathcal{I}(X) = P_1 \cap \cdots \cap P_k.$$

Then

$$X = \mathcal{Z}(P_1) \cup \cdots \cup \mathcal{Z}(P_k)$$

is a decomposition of X into irreducible components.

Example 3.33. In $k[x, y, z]$, the radical ideal $I = (xy, xz, yz)$ corresponds to the variety X given by the union of the three coordinate axes.



Each of these axes is a variety in its own right, corresponding to the ideals (x, y), (x, z) and (y, z). The three axes are the irreducible components of X. And indeed, (x, y), (x, z) and (y, z) are the three minimal primes over I, and

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

We will come back to this decomposition when we discuss primary decomposition.

In summary, Nullstellensatz gives us a dictionary between varieties and ideals:

<u>Algebra</u>	$\longleftrightarrow$	<u>Geometry</u>
algebra of ideals	$\longleftrightarrow$	geometry of varieties
algebra of $R = k[x_1, \ldots, x_d]$	$\longleftrightarrow$	geometry of $\mathbb{A}^d$
radical ideals	$\longleftrightarrow$	varieties
prime ideals	$\longleftrightarrow$	irreducible varieties
maximal ideals	$\longleftrightarrow$	points
(0)	$\longleftrightarrow$	variety $\mathbb{A}^d$

$k[x_1, \dots, x_d]$	$\longleftrightarrow$	variety $\varnothing$
$(x_1 - a_1, \dots, x_d - a_d)$	$\longleftrightarrow$	point $\{(a_1, \dots, a_d)\}$
smaller ideals	$\longleftrightarrow$	larger varieties
larger ideals	$\longleftrightarrow$	smaller varieties

We now know that the subvarieties $\mathcal{Z}(I)$ of $\mathbb{A}^d$ satisfy the following properties:

- The sets $\varnothing$ and $\mathbb{A}^d$ are varieties.
- The finite union of varieties is a variety.
- The arbitrary intersection of varieties is a variety.

These are the axioms of closed sets in a topology. So there is a topology on $\mathbb{A}^d$ whose closed sets are precisely all the subvarieties $\mathcal{Z}(I)$ of $\mathbb{A}^d$. This topology is called the **Zariski topology**.

As a consequence, every variety inherits the Zariski topology, and this is the topology that algebraic geometers usually consider.

Exercise 12. A topological space X is **Noetherian** if it satisfies the descending chain condition for closed subsets: any descending chain of closed subsets

$$X_1 \supseteq X_2 \supseteq \cdots$$

stabilizes. Show that a variety with the Zariski topology is Noetherian.

Exercise 13. Show that if X is a Noetherian topological space, every open subset of X is quasicompact.

This topology is a little weird. In particular, it is *never* Hausdorff, unless the space we are considering is finite. The word *compact* is usually taken to include *Hausdorff*, so algebraic geometers say **quasicompact** to mean compact but maybe not Hausdorff.

Exercise 14. Show that $\mathbb{A}^d_{\mathbb{C}}$ with the Zariski topology is T_1 but not Hausdorff.

3.4 The prime spectrum of a ring

In modern algebraic geometry, on e often studies schemes instead of varieties. We will now introduce the simplest scheme: the spectrum $\operatorname{Spec}(R)$ of a ring R. The spectra of rings are to schemes as $\mathbb{R}^n$ is to manifolds: while a manifold is a topological space that locally looks like $\mathbb{R}^n$, a scheme is, roughly speaking, a topological space that locally looks like $\operatorname{Spec}(R)$ for some ring R.

The *maximal spectrum* of a ring R, denoted $\operatorname{mSpec}(R)$, is the set of maximal ideals of R endowed with the topology with closed sets given by

$$V_{\operatorname{Max}}(I) := \{\mathfrak{m} \in \operatorname{Max}(R) \mid \mathfrak{m} \supseteq I\}$$

as I varies over all the ideals in R. By the Nullstellensatz, for polynomial rings S over an algebraically closed field k, this space mSpec(S) has a natural homeomorphism to $\mathbb{A}^n$ with its Zariski topology, and for an ideal I in $S = k[x_1, \ldots, x_d]$, mSpec(S/I) has a natural homeomorphism to $\mathcal{Z}_k(I) \subseteq \mathbb{A}^n$ with the subspace topology coming from the Zariski topology. Moreover, this is functorial: for any map of finitely generated k-algebras, there is an induced map on maximal ideals.

This is not quite the right notion to deal with general rings, for at least two reasons. First, there are many many interesting rings with only one maximal ideal! The topological space with one element, in contrast, is not that exciting. Second, we would like to have a geometric space that is assigned *functorially* to a ring, meaning that ring homomorphisms induce continuous maps of spaces (in the other direction). For the inclusion $A = k[x, y] = k[x-1, y]$ into $B = k(x)[y] = k(x-1)[y]$, what maximal ideal in A would we assign to $(y) \subseteq B$? How could one of (x, y) or $(x - 1, y)$ have a better claim than the other?

Definition 3.34. Let R be a ring. The **prime spectrum**, or **spectrum** of R is the set of prime ideals of R, denoted Spec(R). This is naturally a poset, partially ordered by inclusion. We also endow it with the topology with closed sets

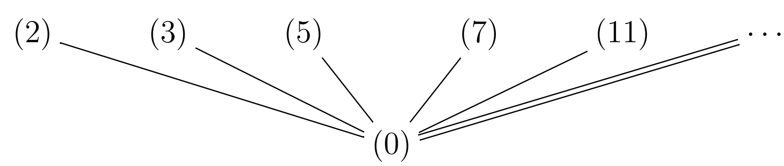
$$V(I) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid \mathfrak{p} \supseteq I\}$$

for (not necessarily proper) ideals $I \subseteq R$. In particular, $\varnothing = V(R)$ is closed.

Soon we will justify that this indeed forms a topology.

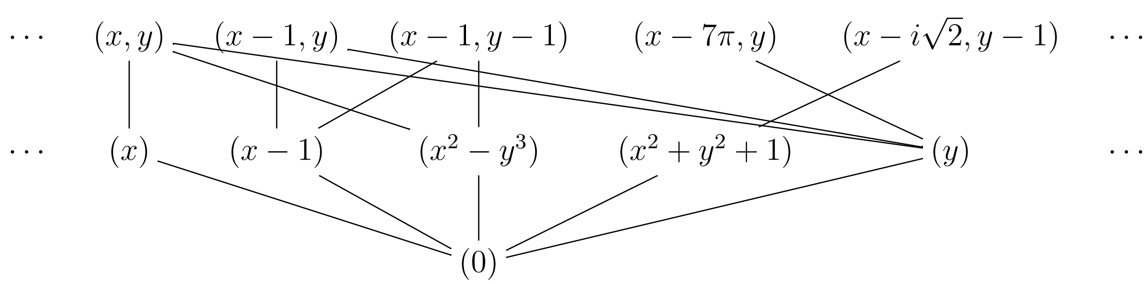
We will illustrate posets with Hasse diagrams: if an element is below something with a line connecting them, the higher element is $\geqslant$ the lower one.

Example 3.35. The spectrum of $\mathbb{Z}$ is the following poset:



The closed sets are of the form $V((n))$, which are the whole space when $n = 0$, the empty set with $n = 1$, and any finite union of things in the top row.

Example 3.36. Here are a few elements in $\mathbb{C}[x, y]$:



Note that Max(R) is a subspace of Spec(R) (that may be neither closed nor open).

Proposition 3.37. *Let R be a ring, and let I, J, and I_λ be ideals (possibly improper).*

a) *If $I \subseteq J$, then $V(J) \subseteq V(I)$.*

b) $V(I) \cup V(J) = V(I \cap J) = V(IJ)$.

c) $\bigcap_\lambda V(I_\lambda) = V(\sum_\lambda I_\lambda)$.

d) $\operatorname{Spec}(R)$ *has a basis given by open sets of the form*

$$D(f) := \operatorname{Spec}(R) \smallsetminus V(f) = \{\mathfrak{p} \in \operatorname{Spec}(R) \mid f \notin \mathfrak{p}\}.$$

e) $\operatorname{Spec}(R)$ *is quasicompact.*

Proof.

a) Clear from the definition.

b) To see $V(I) \cup V(J) \subseteq V(I \cap J)$, just observe that if $\mathfrak{p} \supseteq I$ or $\mathfrak{p} \supseteq J$, then $\mathfrak{p} \supseteq I \cap J$. Since $IJ \subseteq I \cap J$, we have $V(I \cap J) \subseteq V(IJ)$. To show $V(IJ) \subseteq V(I) \cup V(J)$, if $\mathfrak{p} \not\supseteq I, J$, let $f \in I \smallsetminus \mathfrak{p}$, and $g \in J \smallsetminus \mathfrak{p}$. Then $fg \in IJ \smallsetminus \mathfrak{p}$ since $\mathfrak{p}$ is prime.

c) Ideals are closed for sums, so if $\mathfrak{p} \supseteq I_\lambda$ for all λ, then $\mathfrak{p} \supseteq \sum_\lambda I_\lambda$. Moreover, if $\mathfrak{p} \supseteq \sum_\lambda I_\lambda$, then in particular $\mathfrak{p} \supseteq I_\lambda$.

d) We can write any open set as the complement of $V(\{f_\lambda\}_\lambda) = \bigcap_\lambda V(f_\lambda)$, which is the union of $D(f_\lambda)$.

e) Given a sequence of ideals I_λ, if $\sum_\lambda I_\lambda = R$, then 1 is in the sum on the left, and thus 1 can be realized in such a sum over finitely many indices, so

$$R = \sum_\lambda I_\lambda = I_{\lambda_1} + \cdots + I_{\lambda_t}.$$

Thus, if we have a family of closed sets with empty intersection,

$$\varnothing = \bigcap_\lambda V(I_\lambda) = V\left(\sum_\lambda I_\lambda\right) = V(I_{\lambda_1} + \cdots + I_{\lambda_t}) = V(I_{\lambda_1}) \cap \cdots \cap V(I_{\lambda_t}),$$

so some finite subcollection has an empty intersection. □

Definition 3.38 (Induced map on Spec)**.** Given a homomorphism of rings $R \xrightarrow{\varphi} S$, we obtain a map on spectra

$$\begin{aligned} \operatorname{Spec}(S) &\xrightarrow{\varphi^*} \operatorname{Spec}(R) \\ \mathfrak{p} &\longrightarrow \varphi^{-1}(\mathfrak{p}) \end{aligned}$$

The key point here is that the preimage of a prime ideal is also prime. We will often write $\mathfrak{p} \cap R$ for $\varphi^{-1}(\mathfrak{p})$, even if the map is not necessarily an inclusion.

This is not only an order-preserving map, but also continuous: if $U \subseteq \operatorname{Spec}(R)$ is open, say U is the complement of $V(I)$ for some ideal I, then for a prime $\mathfrak{q}$ of S,

$$\mathfrak{q} \in (\varphi^*)^{-1}(U) \iff \mathfrak{q} \cap R \not\supseteq I \iff \mathfrak{q} \not\supseteq IS \iff \mathfrak{q} \notin V(IS).$$

So $(\varphi^*)^{-1}(U)$ is the complement of $V(IS)$, and thus open.

Example 3.39. Let $R \xrightarrow{\pi} R/I$ be the canonical projection. Then

$$\operatorname{Spec}(R/I) \xrightarrow{\pi^*} \operatorname{Spec}(R)$$

corresponds to the inclusion of $V(I)$ into $\operatorname{Spec}(R)$, since primes of R/I correspond to primes of R containing I.

We can use the spectrum of a ring to give an analogue of the strong Nullstellensatz that is valid for any ring. To prepare for this, we need a notion that we will use later.

Definition 3.40. A subset $W \subseteq R$ of a ring R is **multiplicatively closed** if $1 \in W$ and $a, b \in W \Rightarrow ab \in W$.

Lemma 3.41. *Let R be a ring, I an ideal, and W a multiplicatively closed subset. If $W \cap I = \varnothing$, then there is a prime ideal $\mathfrak{p}$ with $\mathfrak{p} \supseteq I$ and $\mathfrak{p} \cap W = \varnothing$.*

Proof. Consider the family of ideals $\mathcal{F} := \{J \mid J \supseteq I, J \cap W = \varnothing\}$ ordered with inclusion. This is nonempty, since it contains I, and any chain $J_1 \subseteq J_2 \subseteq \cdots$ has an upper bound $\cup_i J_i$. Therefore, $\mathcal{F}$ has some maximal element $\mathbb{A}$ by a basic application of Zorn's Lemma. We claim $\mathbb{A}$ is prime. Suppose $f, g \notin \mathbb{A}$. By maximality, $\mathbb{A} + (f)$ and $\mathbb{A} + (g)$ both have nonempty intersection with W, so there exist $r_1 f + a_1$, $r_2 g + a_2 \in W$, with $a_1, a_2 \in \mathbb{A}$. If $fg \in \mathbb{A}$, then

$$\underbrace{(r_1 f + a_1)}_{\in W}\underbrace{(r_2 g + a_2)}_{\in W} = r_1 r_2 fg + r_1 f \underbrace{a_2}_{\in \mathbb{A}} + r_2 g \underbrace{a_1}_{\in \mathbb{A}} + \underbrace{a_1 a_2}_{\in \mathbb{A}} \in W \cap \mathbb{A},$$

a contradiction. $\square$

Proposition 3.42 (Spectrum analogue of strong Nullstellensatz)**.** *Let R be a ring, and I be an ideal. For $f \in R$,*

$$V(I) \subseteq V(f) \Longleftrightarrow f \in \sqrt{I}.$$

Equivalently,

$$\bigcap_{\mathfrak{p} \in V(I)} \mathfrak{p} = \sqrt{I}.$$

Proof. First to justify the equivalence of the two statements we observe:

$$V(I) \subseteq V(f) \Longleftrightarrow f \in \mathfrak{p} \text{ for all } \mathfrak{p} \in V(I) \Longleftrightarrow f \in \bigcap_{\mathfrak{p} \in V(I)} \mathfrak{p}.$$

We prove that $\displaystyle\bigcap_{\mathfrak{p} \in V(I)} \mathfrak{p} = \sqrt{I}$.

($\supseteq$): It suffices to show that $\mathfrak{p} \supseteq I$ implies $\mathfrak{p} \supseteq \sqrt{I}$, and indeed

$$f^n \in I \subseteq \mathfrak{p} \implies f \in \mathfrak{p}.$$

($\subseteq$): If $f \notin \sqrt{I}$, consider the multiplicatively closed set $W = \{1, f, f^2, f^3, \ldots\}$. We have $W \cap I = \varnothing$ by hypothesis. By the previous lemma, there is a prime $\mathfrak{p}$ in $V(I)$ that does not intersect W, and hence does not contain f. $\square$

The following corollary follows in exactly the same way as the analogous statement for subvarieties of $\mathbb{A}^n$, Corollary 3.30.

Corollary 3.43. *Let R a ring. There is an order-reversing bijection*

$$\{\textit{closed subsets of } \operatorname{Spec}(R)\} \longleftrightarrow \{\textit{radical ideals } I \subseteq R\}$$

In particular, for two ideals I, J, $V(I) = V(J)$ if and only if $\sqrt{I} = \sqrt{J}$.

Chapter 4

Local Rings

The study of local rings is central to commutative algebra. As we will see, life is easier in a local ring, so much so that we often want to *localize* so we can be in a local ring. A lot of the things we will say in this chapter also apply to $\mathbb{N}$-graded k-algebras and their homogenous maximal ideal — with some appropriate changes, such as considering only homogeneous ideals.

4.1 Local rings

Definition 4.1. A ring R is a **local ring** if it has exactly one maximal ideal. We often use the notation $(R, \mathfrak{m})$ to denote R and its maximal ideal, or $(R, \mathfrak{m}, k)$ to also specify the residue field $k = R/\mathfrak{m}$. Some people reserve the term *local ring* for a Noetherian local ring, and call what we have defined a **quasilocal ring**; we will not follow this convention here.

Lemma 4.2. *A ring R is local if and only if the set of nonunits of R forms an ideal.*

Proof. If the set of nonunits is an ideal, that must be the only maximal ideal. $\square$

Example 4.3.

a) The ring $\mathbb{Z}/(p^n)$ is local with maximal ideal (p).

b) The ring $\mathbb{Z}_{(p)} = \{\frac{a}{b} \in \mathbb{Q} \mid p \nmid b \text{ when in lowest terms}\}$ is a local ring with maximal ideal (p).

c) The ring of power series $k[\![\underline{x}]\!]$ over a field k is local. Indeed, a power series has an inverse if and only if its constant term is nonzero. The complement of this set of units is the ideal $(\underline{x})$.

d) More generally, $k[\![x_1, \ldots, x_d]\!]$ is local with maximal ideal $(x_1, \ldots, x_d)$.

e) The ring of complex power series holomorphic at the origin, $\mathbb{C}\{\underline{x}\}$, is local. In the above setting, one proves that the series inverse of a holomorphic function at the origin is convergent on a neighborhood of 0.

f) A polynomial ring over a field is certainly not local; we have seen it has so many maximal ideals!

We start with a comment about the characteristic of local rings.

Definition 4.4. The **characteristic** of a ring R is, if it exists, the smallest positive integer n such that

$$\underbrace{1 + \cdots + 1}_{n \text{ times}} = 0.$$

If no such n exists, we say that R has characteristic 0. Equivalently, the characteristic of R is the integer $n \geqslant 0$ such that

$$(n) = \ker \begin{pmatrix} \mathbb{Z} \longrightarrow R \\ a \longmapsto a \cdot 1_R \end{pmatrix}.$$

Proposition 4.5. *Let $(R, \mathfrak{m}, k)$ be a local ring. Then one of the following holds:*

a) $\operatorname{char}(R) = \operatorname{char}(k) = 0$*. We say that R has* **equal characteristic zero***.*

b) $\operatorname{char}(R) = 0$*,* $\operatorname{char}(k) = p$ *for a prime p, so R has* **mixed characteristic** $(0, p)$*.*

c) $\operatorname{char}(R) = \operatorname{char}(k) = p$ *for a prime p, so R has* **equal characteristic** p*.*

d) $\operatorname{char}(R) = p^n$*,* $\operatorname{char}(k) = p$ *for a prime p and an integer $n > 1$.*

If R is reduced, then one of the first three cases holds.

Proof. Since k is a quotient of R, the characteristic of R must be a multiple of the characteristic of k, since the map $\mathbb{Z} \longrightarrow k$ factors through R. We must think of 0 as a multiple of any integer for this to make sense. Now k is a field, so its characteristic is 0 or p for a prime p. If $\operatorname{char}(k) = 0$, then necessarily $\operatorname{char}(R) = 0$. If $\operatorname{char}(k) = p$, we claim that $\operatorname{char}(R)$ must be either 0 or a power of p. Indeed, if we write $\operatorname{char}(R) = p^n \cdot a$ with a coprime to p, note that $p \in \mathfrak{m}$, so if $a \in \mathfrak{m}$, we have $1 \in (p, a) \subseteq \mathfrak{m}$, which is a contradiction. Since R is local, this means that a is a unit. But then, $p^n a = 0$ implies $p^n = 0$, so the characteristic must be p^n. $\square$

Remark 4.6. If R is an $\mathbb{N}$-graded k-algebra with $R_0 = k$, and $\mathfrak{m} = \bigoplus_{n>0} R_0$ is the homogeneous maximal ideal, R and $\mathfrak{m}$ behave a lot like a local ring and its maximal ideal, and we sometimes use the suggestive notation $(R, \mathfrak{m})$ to refer to it. Many properties of local rings also apply to the graded setting, so given a statement about local rings, you might take it as a suggestion that there might be a corresponding statement about graded rings — a statement that, nevertheless, still needs to be proved. There are usually some changes one needs to make to the statement; for example, if a theorem makes assertions about the ideals in a local ring, the corresponding graded statement will likely only apply to homogeneous ideals, and a theorem about finitely generated modules over a local ring will probably translate into a theorem about graded modules in the graded setting.

4.2 Localization

Recall that a multiplicative subset of a ring R is a set $W \ni 1$ that is closed for products. The three most important classes of multiplicative sets are the following:

Example 4.7. Let R be a ring.

a) For any $f \in R$, the set $W = \{1, f, f^2, f^3, \dots\}$ is a multiplicative set.
b) If $\mathfrak{p} \subseteq R$ is a prime ideal, the set $W = R \smallsetminus \mathfrak{p}$ is multiplicative: this is an immediate translation of the definition.
c) The set of *nonzerodivisors* in R — elements that are not zerodivisors — forms a multiplicatively closed subset.

Remark 4.8. An arbitrary intersection of multiplicatively closed subsets is multiplicatively closed. In particular, for any family of primes $\{\mathfrak{p}_\lambda\}$, the complement of $\bigcup_\lambda \mathfrak{p}_\lambda$ is multiplicatively closed.

Definition 4.9 (Localization of a ring)**.** Let R be a ring, and W be a multiplicative set with $0 \notin W$. The **localization** of R at W is the ring

$$W^{-1}R := \left\{ \frac{r}{w} \;\middle|\; r \in R, w \in W \right\} / \sim$$

where $\sim$ is the equivalence relation

$$\frac{r}{w} \sim \frac{r'}{w'} \text{ if there exists } u \in W : u(rw' - r'w) = 0.$$

The operations are given by

$$\frac{r}{v} + \frac{s}{w} = \frac{rw + sv}{vw} \qquad \text{and} \qquad \frac{r}{v}\frac{s}{w} = \frac{rs}{vw}.$$

The zero in $W^{-1}R$ is $\frac{0}{1}$ and the identity is $\frac{1}{1}$. There is a canonical ring homomorphism

$$\begin{aligned} R &\longrightarrow W^{-1}R\,. \\ r &\longmapsto \tfrac{r}{1} \end{aligned}$$

Given an ideal I in $W^{-1}R$, we write $I \cap R$ for its preimage of I in R via the canonical map $R \longrightarrow W^{-1}R$.

Note that we write elements in $W^{-1}R$ in the form $\frac{r}{w}$ even though they are equivalence classes of such expressions.

Remark 4.10. Observe that if R is a domain, the equivalence relation simplifies to $rw' = r'w$, so $R \subseteq W^{-1}R \subseteq \operatorname{Frac}(R)$, and in particular $W^{-1}R$ is a domain too. In particular, $\operatorname{Frac}(R)$ is a localization of R.

In the localization of R at W, every element of W becomes a unit. The following universal property says roughly that $W^{-1}R$ is the smallest R-algebra in which every element of W is a unit.

Proposition 4.11. *Let R be a ring, and W a multiplicative set with $0 \notin W$. Let S be an R-algebra in which every element of W is a unit. Then there is a unique homomorphism α such that the following diagram commutes:*

$$\begin{tikzcd} R \arrow[r] \arrow[d] & W^{-1}R \arrow[dl, "\alpha"] \\ S & \end{tikzcd}$$

where the vertical map is the structure homomorphism and the horizontal map is the canonical homomorphism.

Example 4.12 (Most important localizations)**.** Let R be a ring.

a) For $f \in R$ and $W = \{1, f, f^2, f^3, \dots\}$, we usually write R_f for $W^{-1}R$.
b) For a prime ideal p in R, we generally write $R_\mathfrak{p}$ for $(R \smallsetminus \mathfrak{p})^{-1}R$, and call it **the localization of R at $\mathfrak{p}$**. Given an ideal I in R, we sometimes write $I_\mathfrak{p}$ to refer to $IR_\mathfrak{p}$, the image of I via the canonical map $R \to R_\mathfrak{p}$.
c) When W is the set of nonzerodivisors on R, we call $W^{-1}R$ the **total ring of fractions** of R. When R is a domain, this is just the fraction field of R, and in this case this coincides with the localization at the prime (0).

Remark 4.13. Notice that when we localize at a prime $\mathfrak{p}$, the resulting ring is a local ring $(R_\mathfrak{p}, \mathfrak{p}_\mathfrak{p})$. We can think of the process of localization at $\mathfrak{p}$ as *zooming in* at the prime $\mathfrak{p}$. Many properties of an ideal I can be checked *locally*, by checking them for $IR_\mathfrak{p}$ for each prime $\mathfrak{p} \in V(I)$.

If R is not a domain, the canonical map $R \to W^{-1}R$ is not necessarily injective.

Example 4.14. Consider $R = k[x,y]/(xy)$. The canonical maps $R \longrightarrow R_{(x)}$ and $R \longrightarrow R_y$ are not injective, since in both cases y is invertible in the localization, and thus

$$x \mapsto \frac{x}{1} = \frac{xy}{y} = \frac{0}{y} = \frac{0}{1}.$$

We can now add some more local rings to our list of examples.

Example 4.15.

a) A local ring one often encounters is $k[x_1, \dots, x_d]_{(x_1,\dots,x_d)}$. We can consider this as the ring of rational functions that in lowest terms have a denominator with nonzero constant term. Note that we can talk about lowest terms since the polynomial ring is a UFD.
b) Extending the following example, we have local rings like $(k[x_1, \dots, x_d]/I)_{(x_1,\dots,x_d)}$. If k is algebraically closed and I is a radical ideal, then $k[x_1, \dots, x_d]/I = k[X]$ is the coordinate ring of some affine variety, and $(x_1, \dots, x_d) = \mathfrak{m}_{\underline{0}}$ is the ideal defining the origin (as a point in $X \subseteq \mathbb{A}^d$). Then we call

$$k[X]_{\mathfrak{m}_{\underline{0}}} := (k[x_1, \dots, x_d]/I)_{(x_1,\dots,x_d)}$$

the **local ring of the point** $\underline{0} \in X$; some people write $\mathcal{O}_{X,\underline{0}}$. The radical ideals of this ring consist of radical ideals of $k[X]$ that are contained in $\mathfrak{m}_{\underline{0}}$, which by the Nullstellensatz correspond to subvarieties of X that contain $\underline{0}$. Similarly, we can define the local ring at any point $\underline{a} \in X$.

We state an analogous definition for modules, and for module homomorphisms.

Definition 4.16. Let R be a ring, W be a multiplicative set, and M an R-module. The **localization** of M at W is the $W^{-1}R$-module

$$W^{-1}M := \left\{ \frac{m}{w} \;\middle|\; m \in M, w \in W \right\} / \sim$$

where $\sim$ is the equivalence relation $\dfrac{m}{w} \sim \dfrac{m'}{w'}$ if $u(mw' - m'w) = 0$ for some $u \in W$. The operations are given by

$$\frac{m}{v} + \frac{n}{w} = \frac{mw + nv}{vw} \qquad \text{and} \qquad \frac{r}{v}\frac{m}{w} = \frac{rm}{vw}.$$

If $M \xrightarrow{\alpha} N$ is an R-module homomorphism, then there is a $W^{-1}R$-module homomorphism $W^{-1}M \xrightarrow{W^{-1}\alpha} W^{-1}N$ given by the rule $W^{-1}\alpha(\frac{m}{w}) = \frac{\alpha(m)}{w}$.

We will use the notations M_f and $M_{\mathfrak{p}}$ analogously to R_f and $R_{\mathfrak{p}}$.

To understand localizations of rings and modules, we will want to understand better how they are built from R. First, we take a small detour to talk about colons and annihilators.

Definition 4.17. The **annihilator** of a module M is the ideal

$$\operatorname{ann}(M) := \{r \in R \mid rm = 0 \text{ for all } m \in M\}.$$

Definition 4.18. Let I and J be ideals in a ring R. The **colon** of I and J is the ideal

$$(J : I) := \{r \in R \mid rI \subseteq J\}.$$

More generally, if M and N are submodules of some R-modules A, the colon of N and M is

$$(N :_R M) := \{r \in R \mid rM \subseteq N\}.$$

Exercise 15. The annihilator of M is an ideal in R, and

$$\operatorname{ann}(M) = (0 :_R M).$$

Moreover, any colon $(N :_R M)$ is an ideal in R.

Remark 4.19. If $M = Rm$ is a one-generated R-module, then $M \cong R/I$ for some ideal I. Notice that $I \cdot (R/I) = 0$, and that given an element $g \in R$, we have $g(R/I) = 0$ if and only if $g \in I$. Therefore, $M \cong R/\operatorname{ann}(M)$.

Remark 4.20. Let M be an R-module. If I is an ideal in R such that $I \subseteq \operatorname{ann}(M)$, then $IM = 0$, and thus M has is naturally an R/I-module with the *same* structure it has as an R-module, meaning

$$(r + I) \cdot m = rm$$

for each $r \in R$.

Remark 4.21. If $N \subseteq M$ are R-modules, then $\operatorname{ann}(M/N) = (N :_R M)$.

Lemma 4.22. *Let M be an R-module, and W a multiplicative set. The class*

$$\frac{m}{w} \in W^{-1}M \quad \textit{is zero} \quad \Longleftrightarrow \quad vm = 0 \textit{ for some } v \in W \quad \Longleftrightarrow \quad \operatorname{ann}_R(m) \cap W \neq \varnothing.$$

Note in particular that this holds for $w = 1$.

Proof. For the first equivalence, we use the equivalence relation defining $W^{-1}R$ to note that $\frac{m}{w} = \frac{0}{1}$ in $W^{-1}M$ if and only if there exists some $v \in W$ such that $0 = v(1m - 0w) = vm$. The second equivalence just comes from the definition of the annihilator. $\square$

Remark 4.23. It follows from this lemma that if $N \xrightarrow{\alpha} M$ is injective, then $W^{-1}\alpha$ is also injective, since

$$0 = W^{-1}\alpha\left(\frac{n}{w}\right) = \frac{\alpha(n)}{w} \;\Rightarrow\; 0 = u\alpha(n) = \alpha(un) \text{ for some } u \in W \Rightarrow\; un = 0 \Rightarrow \frac{n}{w} = 0.$$

We want to collect one more lemma for later.

Lemma 4.24. *Let M be a module, and $N_1, \dots, N_t$ be a finite collection of submodules. Let W be a multiplicative set. Then,*

$$W^{-1}(N_1 \cap \cdots \cap N_t) = W^{-1}N_1 \cap \cdots \cap W^{-1}N_t \subseteq W^{-1}M.$$

Proof. The containment $W^{-1}(N_1 \cap \cdots \cap N_t) \subseteq W^{-1}N_1 \cap \cdots \cap W^{-1}N_t$ is clear. Elements of $W^{-1}N_1 \cap \cdots \cap W^{-1}N_t$ are of the form $\frac{n_1}{w_1} = \cdots = \frac{n_t}{w_t}$; we can find a common denominator to realize this in $W^{-1}(N_1 \cap \cdots \cap N_t)$. $\square$

Later we will show that localization has good homological properties: it's an exact functor.

Theorem 4.25. *Given a short exact sequence of R-modules*

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

and a multiplicative set W, the sequence

$$0 \longrightarrow W^{-1}A \longrightarrow W^{-1}B \longrightarrow W^{-1}C \longrightarrow 0$$

is also exact.

Remark 4.26. Given a submodule N of M, we can apply the statement above to the short exact sequence

$$0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$$

and conclude that that $W^{-1}(M/N) \cong W^{-1}M/W^{-1}N$.

Proposition 4.27. *Let W be multiplicatively closed in R.*

a) If I is an ideal in R, then $W^{-1}I \cap R = \{r \in R \mid wr \in I$ for some $w \in W\}$.

b) If J is an ideal in $W^{-1}R$, then $W^{-1}(J \cap R) = J$.

c) If $\mathfrak{p}$ is prime and $W \cap \mathfrak{p} = \varnothing$, then $W^{-1}\mathfrak{p} = \mathfrak{p}(W^{-1}R)$ is prime.

d) The map $\operatorname{Spec}(W^{-1}R) \to \operatorname{Spec}(R)$ *is injective, with image*

$$\{\mathfrak{p} \in \operatorname{Spec}(R) \mid \mathfrak{p} \cap W = \varnothing\}.$$

Proof.

a) Since $W^{-1}(R/I) \cong W^{-1}R/W^{-1}I$, we have $\ker(R \to W^{-1}(R/I)) = R \cap W^{-1}I$. The equality is then clear.

b) The containment $W^{-1}(J \cap R) \subseteq J$ holds for general reasons: given any map f, and a subset J of the target of f, $f(f^{-1}(J)) \subseteq J$. On the other hand, if $\frac{a}{w} \in J$, then $\frac{a}{1} \in J$, since its a unit multiple of an element of J, and thus $a \in J \cap R$, so $\frac{a}{w} \in W^{-1}(J \cap R)$.

c) First, since $W \cap \mathfrak{p} = \varnothing$, and $\mathfrak{p}$ is prime, no element of W kills $\bar{1} = 1 + \mathfrak{p}$ in $R/\mathfrak{p}$, so $\bar{1}/1$ is nonzero in $W^{-1}(R/\mathfrak{p})$. Thus, $W^{-1}R/W^{-1}\mathfrak{p} \cong W^{-1}(R/\mathfrak{p})$ is nonzero, and a localization of a domain, hence is a domain. Thus, $W^{-1}\mathfrak{p}$ is prime.

d) First, by part b), the map $\mathfrak{p} \mapsto W^{-1}\mathfrak{p}$, for $S = \{\mathfrak{p} \in \operatorname{Spec}(R) \mid \mathfrak{p} \cap W = \varnothing\}$ sends primes to primes. We claim that

$$\begin{array}{ccc} \operatorname{Spec}(W^{-1}R) & & S \\ \mathfrak{q} & \longmapsto & \mathfrak{q} \cap R \\ W^{-1}\mathfrak{p} & \longleftarrow\!\shortmid & \mathfrak{p} \end{array}$$

are inverse maps.

We have already seen that $J = (J \cap R)W^{-1}R$ for any ideal J in $W^{-1}R$.

If $W \cap \mathfrak{p} = \varnothing$, then using part a) and the definition of prime, we have that

$$W^{-1}\mathfrak{p} \cap R = \{r \in R \mid rw \in \mathfrak{p} \text{ for some } w \in W\} = \{r \in R \mid r \in \mathfrak{p}\} = \mathfrak{p}. \qquad \square$$

Corollary 4.28. *Let R be a ring and $\mathfrak{p}$ be a prime ideal in R. The map on Spectra induced by the canonical map $R \to R_{\mathfrak{p}}$ corresponds to the inclusion*

$$\{\mathfrak{q} \in \operatorname{Spec}(R) \mid \mathfrak{q} \subseteq \mathfrak{p}\} \subseteq \operatorname{Spec}(R).$$

4.3 NAK

We will now show a very simple but extremely useful result known as Nakayama's Lemma. As noted in [Mat89, page 8], Nakayama himself claimed that this should be attributed to Krull and Azumaya, but it's not clear which of the three actually had the commutative ring statement first. So some authors (eg, Matsumura) prefer to refer to it as NAK. There are actually a range of statements, rather than just one, that go under the banner of Nakayama's Lemma a.k.a. NAK.

Proposition 4.29. *Let R be a ring, I an ideal, and M a finitely generated R-module. If $IM = M$, then*

a) there is an element $r \in 1 + I$ such that $rM = 0$, and

b) there is an element $a \in I$ such that $am = m$ for all $m \in M$.

Proof. Let $M = Rm_1 + \cdots + Rm_s$. By assumption, we have equations

$$m_1 = a_{11}m_1 + \cdots + a_{1s}m_s \ , \ \ldots \ , \ m_s = a_{s1}m_1 + \cdots + a_{ss}m_s,$$

with $a_{ij} \in I$. Setting $A = [a_{ij}]$ and $v = [x_i]$ we have a matrix equations $Av = v$. By the determinantal trick, Lemma 1.35, the element $\det(I_{s\times s} - A) \in R$ kills each m_i, and hence M. Since $\det(I_{s\times s} - A) \equiv \det(I_{s\times s}) \equiv 1 \bmod I$, this determinant is the element r we seek for the first statement.

For the latter statement, set $a = 1 - r$; this is in I and satisfies $am = m - rm = m$ for all $m \in M$. □

Proposition 4.30. *Let $(R, \mathfrak{m}, k)$ be a local ring, and M be a finitely generated module. If $M = \mathfrak{m}M$, then $M = 0$.*

Proof. By the Proposition 4.29, there exists an element $r \in 1+\mathfrak{m}$ that annihilates M. Notice that $1 \notin \mathfrak{m}$, so any such r must be outside of $\mathfrak{m}$, and thus a unit. Multiplying by its inverse, we conclude that 1 annihilates M, or equivalently, that $M = 0$. □

Proposition 4.31. *Let $(R, \mathfrak{m}, k)$ be a local ring, and M be a finitely generated module, and N a submodule of M. If $M = N + \mathfrak{m}M$, then $M = N$.*

Proof. By taking the quotient by N, we see that

$$M/N = (N + \mathfrak{m}M)/N = \mathfrak{m}\,(M/N)\,.$$

By Proposition 4.30, $M = N$. □

Proposition 4.32. *Let $(R, \mathfrak{m}, k)$ be a local ring, and M be a finitely generated module. For $m_1, \ldots, m_s \in M$,*

$$m_1, \ldots, m_s \text{ generate } M \Longleftrightarrow \overline{m_1}, \ldots, \overline{m_s} \text{ generate } M/\mathfrak{m}M.$$

Thus, any generating set for M consists of at least $\dim_k(M/\mathfrak{m}M)$ elements.

Proof. The implication ($\Rightarrow$) is clear. If $m_1, \ldots, m_s \in M$ are such that $\overline{m_1}, \ldots, \overline{m_s}$ generate $M/\mathfrak{m}M$, let $N = Rm_1 + \cdots + Rm_s \subseteq M$. By Proposition 4.30, $M/N = 0$ if and only if $M/N = \mathfrak{m}(M/N)$. The latter statement is equivalent to $M = \mathfrak{m}M + N$, which is equivalent to saying that $M/\mathfrak{m}M$ is generated by the image of N. □

Remark 4.33. Since $R/\mathfrak{m}$ is a field, $M/\mathfrak{m}M$ is a vector space over the field $R/\mathfrak{m}$.

Definition 4.34. Let $(R, \mathfrak{m})$ be a local ring, and M a finitely generated module. A set of elements $\{m_1, \ldots, m_t\}$ is a **minimal generating set** of M if the images of $m_1, \ldots, m_t$ form a basis for the $R/\mathfrak{m}$ vector space $M/\mathfrak{m}M$.

As a consequence of basic facts about basis for vector spaces, we conclude that any generating set for M contains a minimal generating set, and that every minimal generating set has the same cardinality.

Definition 4.35. Let $(R, \mathfrak{m})$ be a local ring, and N an R-module. The **minimal number of generators** of M is

$$\mu(M) := \dim_{R/\mathfrak{m}} (M/\mathfrak{m}M) .$$

Equivalently, this is the number of elements in a minimal generating set for M.

We commented before that graded rings behave a lot like local rings, so now we want to give graded analogues for the results above.

Proposition 4.36. *Let R be an $\mathbb{N}$-graded ring, and M a $\mathbb{Z}$-graded module such that $M_{<a} = 0$ for some a. If $M = (R_+)M$, then $M = 0$.*

Proof. On the one hand, the homogeneous elements in M live in degrees at least a, but $(R_+)M$ lives in degrees strictly bigger than a. If M has a nonzero element, it has a nonzero homogeneous element, and we obtain a contradiction. □

This condition includes all finitely generated $\mathbb{Z}$-graded R-modules.

Remark 4.37. If M is finitely generated, then it can be generated by finitely many homogeneous elements, the homogeneous components of some finite generating set. If a is the smallest degree of a homogeneous element in a homogeneous generating set, since R lives only in positive degrees we must have $M \subseteq RM_{\geqslant a} \subseteq M_{\geqslant a}$, so $M_{<a} = 0$.

Just as above, we obtain the following:

Proposition 4.38. *Let R be an $\mathbb{N}$-graded ring, with R_0 a field, and M a $\mathbb{Z}$-graded module such that $M_{<a} = 0$ for some degree a. A set of elements of M generates M if and only if their images in $M/(R_+)M$ spans as a vector space. Since M and $(R_+)M$ are graded, $M/(R_+)M$ admits a basis of homogeneous elements.*

In particular, if k is a field, R is a positively graded k-algebra, and I is a homogeneous ideal, then I has a minimal generating set by homogeneous elements, and this set is unique up to k-linear combinations.

Definition 4.39. Let R be an $\mathbb{N}$-graded ring with R_0 a field, and M a finitely generated $\mathbb{Z}$-graded R-module. The **minimal number of generators** of M is

$$\mu(M) := \dim_{R/R_+} (M/R_+M) .$$

We can use Macaulay2 to compute (the) minimal (number of) generators of graded modules over graded k-algebras, using the commands `mingens` and `numgens`.

Note that we can use NAK to prove that certain modules are finitely generated in the graded case; in the local case, we cannot.

Chapter 5

Decomposing ideals

We will consider a few ways of decomposing ideals into pieces, in three ways with increasing detail. The first is the most directly geometric: for any ideal I in a Noetherian ring, we aim to write $V(I)$ as a finite union of $V(\mathfrak{p}_i)$ for prime ideals $\mathfrak{p}_i$.

5.1 Minimal primes and support

Recall the definition of minimal primes that we mentioned before.

Definition 5.1. The primes that contain I and are minimal with the property of containing I are called the **minimal primes** of I. That is, the minimal primes of I are the minimal elements of $V(I)$. We write $\operatorname{Min}(I)$ for this set.

Exercise 16. Let R be a ring, and I an ideal. Every prime $\mathfrak{p}$ that contains I contains a minimal prime of I. Consequently,

$$\sqrt{I} = \bigcap_{\mathfrak{p} \in \operatorname{Min}(I)} \mathfrak{p}.$$

Remark 5.2. If $\mathfrak{p}$ is prime, then $\operatorname{Min}(\mathfrak{p}) = \{\mathfrak{p}\}$. Also, since $V(I) = V(\sqrt{I})$, we have $\operatorname{Min}(I) = \operatorname{Min}(\sqrt{I})$.

As a special case, the nilpotent elements of a ring R are exactly the elements in every minimal prime of R, or equivalently, in every minimal prime of the ideal (0). The radical of (0) is often called the **nilradical** of R, denoted $\mathcal{N}(R)$.

Lemma 5.3. *Whenever* $I = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n$ *for some* $\mathfrak{p}_i \not\subseteq \mathfrak{p}_j$ *for each* i, j, *we have* $\operatorname{Min}(I) = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\}$.

Proof. If $\mathfrak{q}$ is a prime containing I, then $\mathfrak{q} \supseteq (\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n)$. But if $\mathfrak{q} \not\supseteq \mathfrak{p}_i$ for each i, then there are elements $f_i \in \mathfrak{p}_i$ such that $f_i \notin \mathfrak{q}$, and the product $f_1 \cdots f_n \in (\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n)$ but $f_1 \cdots f_n \notin \mathfrak{q}$. Therefore, any minimal prime of I must be one of the $\mathfrak{p}_i$. Since we assumed that the $\mathfrak{p}_i$ are incomparable, they are exactly all the minimal primes of I. $\square$

Remark 5.4. If $I = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n$ for some primes p_i, we can always delete unnecessary components until no component can be deleted. Therefore, $\operatorname{Min}(I) \subseteq \{\mathfrak{p}_1, \cdots, \mathfrak{p}_n\}$.

Theorem 5.5. *Let R be a Noetherian ring. Then any ideal I has finitely many minimal primes, and thus $\sqrt{I}$ is a finite intersection of primes.*

Proof. Let $S = \{\text{ideals } I \subseteq R \mid \operatorname{Min}(I) \text{ is infinite}\}$, and suppose, to obtain a contradiction, that $S \neq \varnothing$. Since R is Noetherian, S has a maximal element J, by Proposition 1.2. If J was a prime ideal, then $\operatorname{Min}(J) = \{J\}$ would be finite, by Remark 5.2, so J is not prime. However, $\operatorname{Min}(J) = \operatorname{Min}(\sqrt{J})$, and thus $\sqrt{J} \supseteq J$ is also in S, so we conclude that J is radical. Since J is not prime, we can find some $a, b \notin J$ with $ab \in J$. Then $J \subsetneq J + (a) \subseteq \sqrt{J + (a)}$ and $J \subsetneq \sqrt{J + (b)}$. Since J is maximal in S, we conclude that $\sqrt{J + (a)}$ and $\sqrt{J + (b)}$ have finitely many minimal primes, so we can write

$$J + (a) = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_a \text{ and } J + (b) = \mathfrak{p}_{a+1} \cap \cdots \cap \mathfrak{p}_b$$

for some prime ideals $\mathfrak{p}_i$. Let $f \in \sqrt{J + (a)} \cap \sqrt{J + (b)}$. Some sufficiently high power of f is in both $J + (a)$ and $J + (b)$, so there exist $n, m \geqslant 1$ such that

$$f^n = J + (a) \text{ and } f^m \in J + (b)$$

so

$$f^{n+m} \in (J + (a))(J + (b)) \subseteq J^2 + J(a) + J(b) + (\underbrace{ab}_{\in J}) \subseteq J.$$

Therefore, $f \in \sqrt{J} = J$. This shows that

$$J = (J + (a)) \cap (J + (b)) = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_a \cap \mathfrak{p}_{a+1} \cap \cdots \cap \mathfrak{p}_b.$$

By Lemma 5.3, we see that $\operatorname{Min}(J)$ must be a subset of $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_b\}$, so it is finite. □

Remark 5.6. Lemma 5.3, Theorem 5.5, and Exercise 16 imply that an ideal I is equal to a finite intersection of primes if and only if I is radical.

We now can describe the relationship between the poset structure of $\operatorname{Spec}(R)$ and the topology.

Proposition 5.7. *Let R be a ring, and $X = \operatorname{Spec}(R)$.*

a) The poset structure on X can be recovered from the topology: $\mathfrak{p} \subseteq \mathfrak{q} \iff \mathfrak{q} \in \overline{\{\mathfrak{p}\}}$.

b) If R is Noetherian, the topology on X can by recovered from the poset structure by the rule

$$Y \subseteq X \textit{ is closed} \Longleftrightarrow Y = \{\mathfrak{q} \in X \mid \mathfrak{p}_i \subseteq \mathfrak{q} \textit{ for some } i\} \textit{ for some } \mathfrak{p}_1, \ldots, \mathfrak{p}_n \in X.$$

Proof.

a) By definition of closure, we have

$$\overline{\{\mathfrak{p}\}} = \bigcap_{\mathfrak{p} \in V(I)} V(I).$$

If $\mathfrak{p} \in V(I)$ then $I \subseteq \mathfrak{p}$, which implies $V(\mathfrak{p}) \subseteq V(I)$. It follows that $\overline{\{\mathfrak{p}\}} = V(\mathfrak{p})$, and thus the claim.

b) If Y is closed, we have $Y = V(I) = V(\sqrt{I}) = V(\mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n) = V(\mathfrak{p}_1) \cup \cdots \cup V(\mathfrak{p}_n)$. For the converse, we can work backwards. $\square$

We now wish to understand modules in a similar way.

Definition 5.8. If M is an R-module, the **support** of M is

$$\operatorname{Supp}(M) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid M_{\mathfrak{p}} \neq 0\}.$$

Proposition 5.9. *Given M a finitely generated R-module over a ring R,*

$$\operatorname{Supp}(M) = V(\operatorname{ann}_R(M)).$$

In particular, $\operatorname{Supp}(R/I) = V(I)$.

Proof. Let $M = Rm_i + \cdots + Rm_n$. We have

$$\operatorname{ann}_R(M) = \bigcap_{i=1}^{n} \operatorname{ann}_R(m_i),$$

so

$$V(\operatorname{ann}_R(M)) = \bigcup_{i=1}^{n} V(\operatorname{ann}_R(m_i)).$$

Notice that we need finiteness here. Also, we claim that

$$\operatorname{Supp}(M) = \bigcup_{i=1}^{n} \operatorname{Supp}(Rm_i).$$

To show $(\supseteq)$, notice that $(Rm_i)_{\mathfrak{p}} \subseteq M_{\mathfrak{p}}$, so

$$\mathfrak{p} \in \operatorname{Supp}(Rm_i) \implies 0 \neq (Rm_i)_{\mathfrak{p}} \subseteq M_{\mathfrak{p}} \implies \mathfrak{p} \in \operatorname{Supp}(M).$$

On the other hand, the images of $m_1, \ldots, m_n$ in $M_{\mathfrak{p}}$ generate $M_{\mathfrak{p}}$ for each $\mathfrak{p}$, so $\mathfrak{p} \in \operatorname{Supp}(M)$ if and only if $\mathfrak{p} \in \operatorname{Supp}(Rm_i)$ for some m_i. Thus, we can reduce to the case of a cyclic module Rm. Now $\frac{m}{1} = 0$ in $M_{\mathfrak{p}}$ if and only if $(R \smallsetminus \mathfrak{p}) \cap \operatorname{ann}_R(m) \neq \varnothing$, which happens if and only if $\operatorname{ann}_R(m) \not\subseteq \mathfrak{p}$. $\square$

The finite generating hypothesis is necessary!

Example 5.10. Let k be a field, and $R = k[x]$. Take

$$M = R_x/R = \bigoplus_{i>0} k \cdot x^{-i}.$$

With this k-vector space structure, the action is given by multiplication in the obvious way, then killing any nonnegative degree terms.

On one hand, we claim that $\operatorname{Supp}(M) = \{(x)\}$. Indeed, any element of M is killed by a large power of x, so $W^{-1}M = 0$ whenever $x \in W$, so $\operatorname{Supp}(M) \subseteq \{(x)\}$. We will soon see that the support of a nonzero module is nonempty, and thus $\operatorname{Supp}(M) = \{(x)\}$.

On the other hand, the annihilator of the class of x^{-n} is x^n, so

$$\operatorname{ann}_R(M) \subseteq \bigcap_{n \geqslant 1} (x^n) = 0.$$

In particular, $V(\operatorname{ann}_R(M)) = \operatorname{Spec}(R)$.

Example 5.11. Let $R = \mathbb{C}[x]$, and $M = \bigoplus_{n\in\mathbb{Z}} R/(x-n)$.

First, note that $M_{\mathfrak{p}} = \bigoplus_{n\in\mathbb{Z}} (R/(x-n))_{\mathfrak{p}}$, so

$$\operatorname{Supp}(M) = \bigcup_{n\in\mathbb{Z}} \operatorname{Supp}(R/(x-n)) = \bigcup_{n\in\mathbb{Z}} V((x-n)) = \{(x-n) \mid n \in \mathbb{Z}\}.$$

On the other hand,

$$\operatorname{ann}_R(M) = \bigcap_{n\in\mathbb{Z}} \operatorname{ann}_R(R/(x-n)) = \bigcap_{n\in\mathbb{Z}} (x-n) = 0.$$

Note that in this example the support is not even closed.

Lemma 5.12. *Let R be a ring, M an R-module, and $m \in M$. The following are equivalent:*

1) *$m = 0$ in M.*
2) *$\frac{m}{1} = 0$ in $M_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{Spec}(R)$.*
3) *$\frac{m}{1} = 0$ in $M_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{mSpec}(R)$.*

Proof. The implications 1) $\implies$ 2) $\implies$ 3) are clear. If $m \neq 0$, its annihilator is a proper ideal, which is contained in a maximal ideal, so $V(\operatorname{ann}_R m) = \operatorname{Supp}(Rm)$ contains a maximal ideal, so $\frac{m}{1} \neq 0$ in $M_{\mathfrak{p}}$ for some maximal ideal $\mathfrak{p}$. □

Lemma 5.13. *Let R be a ring, L, M, N be modules. If*

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0$$

is exact, then $\operatorname{Supp}(L) \cup \operatorname{Supp}(N) = \operatorname{Supp}(M)$.

Proof. Localization is exact, by Theorem 4.25, so for any $\mathfrak{p}$,

$$0 \longrightarrow L_{\mathfrak{p}} \longrightarrow M_{\mathfrak{p}} \longrightarrow N_{\mathfrak{p}} \longrightarrow 0$$

is exact. If $\mathfrak{p} \in \operatorname{Supp}(L) \cup \operatorname{Supp}(N)$, then $L_{\mathfrak{p}}$ or $N_{\mathfrak{p}}$ is nonzero, so $M_{\mathfrak{p}}$ must be nonzero as well. On the other hand, if $\mathfrak{p} \notin \operatorname{Supp}(L) \cup \operatorname{Supp}(N)$, then $L_{\mathfrak{p}} = N_{\mathfrak{p}} = 0$, so $M_{\mathfrak{p}} = 0$. □

Remark 5.14. As a corollary, $\operatorname{Supp}(L) \subseteq \operatorname{Supp}(M)$ for any submodule L of M.

Corollary 5.15. *If M is a finitely generated R-module,*

1) *$M = 0$.*
2) *$M_{\mathfrak{p}} = 0$ in $M_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{Spec}(R)$.*
3) *$M_{\mathfrak{p}} = 0$ in $M_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{mSpec}(R)$.*

Proof.

The implications $\Rightarrow$ are clear. To show the last implies the first, we show the contrapositive. If $m \neq 0$, consider $Rm \subseteq M$. By Lemma 5.12, there is a maximal ideal in $\operatorname{Supp}(Rm)$, and by Lemma 5.13 applied to the inclusion $Rm \subseteq M$, this maximal ideal is in $\operatorname{Supp}(M)$ as well. □

So we conclude that $\operatorname{Supp}(M) \neq \varnothing$ for any R-module $M \neq 0$.

5.2 Associated primes

Remark 5.16. Let R be a ring, I be an ideal in R, and M be an R-module. To give an R-module homomorphism $R \longrightarrow M$ is the same as choosing an element m of M (the image of 1 via our map) or equivalently, to choose a cyclic submodule of M (the submodule generated by m).

To give an R-module homomorphism $R/I \longrightarrow M$ is the same as giving an R-module homomorphism $R \longrightarrow M$ whose image is killed by I. Thus giving an R-module homomorphism $R/I \longrightarrow M$ is to choose an element $m \in M$ that is killed by I, meaning $I \subseteq \operatorname{ann}(m)$.

Definition 5.17. Let R be a ring, and M a module. We say that $\mathfrak{p} \in \operatorname{Spec}(R)$ is an **associated prime** of M if $\mathfrak{p} = \operatorname{ann}_R(m)$ for some $m \in M$. Equivalently, $\mathfrak{p}$ is associated to M if there is an injective homomorphism $R/\mathfrak{p} \longrightarrow M$. We write $\operatorname{Ass}_R(M)$ for the set of associated primes of M.

If I is an ideal, by the **associated primes** of I we (almost always) mean the associated primes of R/I. To avoid confusion, we will try to write $\operatorname{Ass}_R(R/I)$.

Lemma 5.18. *Let R be a Noetherian ring and M be an R-module. A prime P is associated to M if and if and only if $P_P \in \operatorname{Ass}(M_P)$.*

Proof. Localization is exact, so any inclusion $R/P \subseteq M$ localizes to an inclusion $R_P/P_P \subseteq M_P$. Conversely, suppose that $P_P = \operatorname{ann}(\frac{m}{w})$ for some $\frac{m}{w} \in M_P$. Let $P = (f_1, \ldots, f_n)$. Since $\frac{f_i}{1}\frac{m}{r} = \frac{0}{1}$, there exists $u_i \notin P$ such that $u_i f_i m = 0$. Then $u = u_1 \cdots u_n$ is not in P, since P is prime, and $uf_i m = 0$ for all i. Since the f_i generate P, we have $P(um) = 0$. On the other hand, if $r \in \operatorname{ann}(um)$, then $\frac{ru}{1} \in \operatorname{ann}(\frac{m}{w}) = P_P$. We conclude that $ru \in P_P \cap R = P$. Since $u \notin P$, we conclude that $r \in P$. $\square$

Lemma 5.19. *If $\mathfrak{p}$ is prime, $\operatorname{Ass}_R(R/\mathfrak{p}) = \{\mathfrak{p}\}$.*

Proof. For any nonzero $\bar{r} \in R/\mathfrak{p}$, we have $\operatorname{ann}_R(\bar{r}) = \{s \in R \mid rs \in \mathfrak{p}\} = \mathfrak{p}$ by definition of prime ideal. $\square$

Let's recall the definition of zerodivisors on M.

Definition 5.20. Let M be an R-module. An element $r \in R$ is a **zerodivisor** on M if $rm = 0$ for some $m \in M$. We sometimes write the set of zerodivisors of M as $\mathcal{Z}(M)$.

Lemma 5.21. *If R is Noetherian, and M is an arbitrary R-module, then*

1) *For any nonzero $m \in M$, $\operatorname{ann}_R(m)$ is contained in an associated prime of M.*
2) $\operatorname{Ass}(M) = \varnothing \iff M = 0$, *and*
3) $\displaystyle\bigcup_{\mathfrak{p} \in \operatorname{Ass}(M)} \mathfrak{p} = \mathcal{Z}(M)$.

Additionally, if R and M are $\mathbb{Z}$-graded and $M \neq 0$, M has an associated prime that is homogeneous.

Proof. Even if R is not Noetherian, $M = 0$ implies $\operatorname{Ass}(M) = \varnothing$ by definition. So we focus on the case when $M \neq 0$.

First, suppose that we have shown 1. If $M \neq 0$, then M contains a nonzero element m, and $\operatorname{ann}(m)$ is contained in an associated prime of M. In particular, $\operatorname{Ass}(M) \neq 0$, and 2 holds. Now if $r \in \mathcal{Z}(M)$, then by definition we have $r \in \operatorname{ann}(m)$ for some nonzero $m \in M$. Since $\operatorname{ann}(m)$ is contained in some associated prime of M, so is r. On the other hand, if $\mathfrak{p}$ is an associated prime of M, then by definition all elements in $\mathfrak{p}$ are zerodivisors on M. This shows that 3 holds. So all that is left is to prove 1.

Now we show 1 for any $M \neq 0$. The set of ideals $S := \{\operatorname{ann}_R(m) \mid m \in M, m \neq 0\}$ is nonempty, and any element in S is contained in a maximal element, by Noetherianity. Note in fact that any element in S must be contained in a maximal element of S. Let $I = \operatorname{ann}(m)$ be any maximal element, and let $rs \in I$, $s \notin I$. We always have $\operatorname{ann}(sm) \supseteq \operatorname{ann}(m)$, and equality holds by the maximality of $\operatorname{ann}(m)$ in S. Then $r(sm) = (rs)m = 0$, so $r \in \operatorname{ann}(sm) = \operatorname{ann}(m) = I$. We conclude that I is prime, and therefore it is an associated prime of M.

For the graded case, replace the set of zerovisisors with the annihilators of homogeneous elements. Such annihilator is homogeneous, since if m is homogeneous, and $fm = 0$, writing $f = f_{a_1} + \cdots + f_{a_b}$ as a sum of homogeneous elements of different degrees a_i, then $0 = fm = f_{a_1}m + \cdots + f_{a_b}m$ is a sum of homogeneous elements of different degrees, so $f_{a_i}m = 0$ for each i. The same argument above works if we take $\{\operatorname{ann}_R(m) \mid m \in M, m \neq 0 \text{ homogeneous }\}$, using the following lemma. $\square$

Lemma 5.22. *If R is $\mathbb{Z}$-graded, an ideal with the property*

$$\textit{for any homogeneous elements } r, s \in R \quad rs \in I \Rightarrow r \in I \textit{ or } s \in I$$

is prime.

Proof. We need to show that this property implies that for any $a, b \in R$ not necessarily homogeneous, $ab \in I$ implies $a \in I$ or $b \in I$. We induce on the number of nonzero homogeneous components of a plus the number of nonzero homogeneous components of b. The base case is when this is two, which means that both a and b are homogeneous, and thus the hypotheses already gives us this case. Otherwise, write $a = a' + a_m$ and $b = b' + b_n$, where a_m, b_n are the nonzero homogeneous components of a and b of largest degree, respectively. We have $ab = (a'b' + a_m b' + b_n a') + a_m b_n$, where $a_m b_n$ is either the largest homogeneous component of ab or else it is zero. Either way, $a_m b_n \in I$, so $a_m \in I$ or $b_n \in I$; without loss of generality, we can assume $a_m \in I$. Then $ab = a'b + a_m b$, and $ab, a_m b \in I$, so $a'b \in I$, and the total number of homogeneous pieces of $a'b$ is smaller, so by induction, either $a' \in I$ so that $a \in I$, or else $b \in I$. $\square$

Lemma 5.23. *If*

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0$$

is an exact sequence of R-modules, then $\operatorname{Ass}(L) \subseteq \operatorname{Ass}(M) \subseteq \operatorname{Ass}(L) \cup \operatorname{Ass}(N)$.

Proof. If $R/\mathfrak{p}$ includes in L, then composition with the inclusion $L \hookrightarrow M$ gives an inclusion $R/\mathfrak{p} \hookrightarrow M$. So $\operatorname{Ass}(L) \subseteq \operatorname{Ass}(M)$. Let $\mathfrak{p} \in \operatorname{Ass}(M)$, say $\mathfrak{p} = \operatorname{ann}(m)$. First, note that $\mathfrak{p} \subseteq \operatorname{ann}(rm)$ for all $r \in R$.

Thinking of L as a submodule of N, suppose that there exists $r \notin \mathfrak{p}$ such that $rm \in L$. Then

$$s(rm) = 0 \Longleftrightarrow (sr)m = 0 \implies sr \in \mathfrak{p} \implies s \in p.$$

So $\mathfrak{p} = \operatorname{ann}(rm)$, and thus $\mathfrak{p} \in \operatorname{Ass}(L)$.

If $rm \notin L$ for all $r \notin \mathfrak{p}$, let n be the image of m in N. Thinking of N as M/L, if $rn = 0$, then we must have $rm \in L$, and by assumption this implies $r \in \mathfrak{p}$. Since $\mathfrak{p} = \operatorname{ann}(m) \subseteq \operatorname{ann}(n)$, we conclude that $\mathfrak{p} = \operatorname{ann}(n)$. Therefore, $\mathfrak{p} \in \operatorname{Ass}(N)$. □

Note that the inclusions in Lemma 5.23 are not necessarily equalities.

Example 5.24. If M is a module with at least two associated primes, and $\mathfrak{p}$ is an associated prime of M, then

$$0 \longrightarrow R/\mathfrak{p} \longrightarrow M$$

is exact, but $\{\mathfrak{p}\} = \operatorname{Ass}(R/\mathfrak{p}) \subsetneq \operatorname{Ass}(M)$.

Example 5.25. Let $R = k[x]$, where k is a field, and consider the short exact sequence of R-modules

$$0 \longrightarrow (x) \longrightarrow R \longrightarrow R/(x) \longrightarrow 0\,.$$

Then one can check that:

- $\operatorname{Ass}(R/(x)) = \operatorname{Ass}(k) = \{(x)\}$.
- $\operatorname{Ass}(R) = \operatorname{Ass}((x)) = \{(0)\}$.

In particular, $\operatorname{Ass}(R) \subsetneq \operatorname{Ass}(R/(x)) \cup \operatorname{Ass}((x))$.

Corollary 5.26. *Let A and B be R-modules. Then* $\operatorname{Ass}(A \oplus B) = \operatorname{Ass}(A) \cup \operatorname{Ass}(B)$.

Proof. Apply Lemma 5.23 to the short exact sequence

$$0 \longrightarrow A \longrightarrow A \oplus B \longrightarrow B \longrightarrow 0\,.$$

We obtain $\operatorname{Ass}(A) \subseteq \operatorname{Ass}(A \oplus B) \subseteq \operatorname{Ass}(A) \cup \operatorname{Ass}(B)$. Repeat with

$$0 \longrightarrow B \longrightarrow A \oplus B \longrightarrow A \longrightarrow 0\,.$$ □

We will need a bit of notation for graded modules to help with the next statement; we saw a simple use of this notation back in Example 2.13.

Definition 5.27. Let R and M be T-graded, and $t \in T$. The **shift** of M by t is the graded R-module $M(t)$ with graded pieces $M(t)_i := M_{t+i}$. This is isomorphic to M as an R-module, when we forget about the graded structure.

Theorem 5.28. *Let R be a Noetherian ring, and M is a finitely generated module. There exists a* **filtration** *of M*

$$M = M_t \supsetneq M_{t-1} \supsetneq M_{t-2} \supsetneq \cdots \supsetneq M_1 \supsetneq M_0 = 0$$

such that $M_i/M_{i-1} \cong R/\mathfrak{p}_i$ for primes $\mathfrak{p}_i \in \operatorname{Spec}(R)$. Such a filtration is called a **prime filtration** *of M.*

If R and M are $\mathbb{Z}$-graded, there exists a prime filtration as above where the quotients $M_i/M_{i-1} \cong (R/\mathfrak{p}_i)(t_i)$ are graded modules, the $\mathfrak{p}_i$ are homogeneous primes, and the t_i are integers.

Proof. If $M \neq 0$, then M has at least one associated prime, so there is an inclusion $R/\mathfrak{p}_1 \hookrightarrow M$. Let M_1 be the image of this inclusion. If $M/M_1 \neq 0$, it has an associated prime, so there is an $M_2 \subseteq M$ such that $R/\mathfrak{p}_2 \cong M_2/M_1 \subseteq R/M_1$. Continuing this process, we get a strictly ascending chain of submodules of M where the successive quotients are of the form $R/\mathfrak{p}_i$. If we do not have $M_t = M$ for some t, then we get an infinite strictly ascending chain of submodules of M, which contradicts that M is a Noetherian module.

In the graded case, if $\mathfrak{p}_i$ is the annihilator of an element m_i of degree t_i, we have a degree-preserving map $(R/\mathfrak{p}_i)(t_i) \cong Rm_i$ sending the class of 1 to m_i. $\square$

Example 5.29. Let's build a prime filtration for $M = R/I$, where $I = (x^2, yz)$ and $R = \mathbb{Q}[x, y, z]$. With a little help from Macaulay2, we find that

```
i4 : associatedPrimes M

o4 = {ideal (y, x), ideal (z, x)}

o4 : List
```

So our first goal is to find $m \in M$ such that $\text{ann}(m) = (x, z)$ or $\text{ann}(m) = (x, z)$. Let's start from (x, z). To find such an element, we can start by searching for all the elements killed by (x, z):

```
i5 : I : ideal"x,z"

                            2
o5 = ideal (y*z, x*y, x )

o5 : Ideal of R
```

Now yz and x^2 are both 0 in M, so the submodule of M generated by xy is precisely the set of elements killed by (x, z). Is $\text{ann}(R \cdot xy) = (x, z)$?

```
i6 : ann ((I + ideal(x*y))/I)

o6 = ideal (z, x)

o6 : Ideal of R
```

Yes, it is! So our prime filtration starts with

$$M_0 = 0 \subseteq M_1 = \frac{I + (xy)}{I},$$

where our computations so far show that $\text{ann}(M_1) = (x, z)$. For step 2, we start from scratch, and compute the associated primes of $M/M_1 \cong R/(I + (y))$:

```
i7 : associatedPrimes (R^1/(I + ideal"xy"))

o7 = {ideal (y, x), ideal (z, x)}

o7 : List
```

Unfortunately, we will again have to find another element killed by (x, z). So we repeat the process:

```
i8 : (I + ideal"xy") : ideal"x,z"

                    2
o8 = ideal (y, x )

o8 : Ideal of R

i9 : ann((I + ideal"y")/(I + ideal"xy"))

o9 = ideal (z, x)

o9 : Ideal of R
```

So in M_1, $\mathrm{ann}(y) = (x, z)$, so we can take the submodule generated by y for our next step, so our prime filtration for now looks like

$$M_0 = 0 \underset{R/(x,z)}{\subseteq} M_1 = \frac{I + (xy)}{I} \underset{R/(x,z)}{\subseteq} M_2 = \frac{I + (y)}{I}.$$

So now we repeat the process with $M/M_2 \cong R/(I + (y))$:

```
i10 : associatedPrimes (R^1/(I + ideal"y"))

o10 = {ideal (y, x)}

o10 : List

i11 : (I + ideal"y") : ideal"x,y"

o11 = ideal (y, x)

o11 : Ideal of R

i12 : ann((I + ideal"x")/(I + ideal"y"))

o12 = ideal (y, x)

o12 : Ideal of R
```

This gives us

$$M_0 = 0 \underset{R/(x,z)}{\subseteq} M_1 = \frac{I + (xy)}{I} \underset{R/(x,z)}{\subseteq} M_2 = \frac{I + (y)}{I} \underset{R/(x,y)}{\subseteq} M_3 = \frac{I + (x, y)}{I}.$$

Next, we take $M/M_3 \cong R/(I + (x, y))$ and find that

```
i13 : associatedPrimes (R^1/(I + ideal"x,y"))

o13 = {ideal (y, x)}

o13 : List

i14 : (I + ideal"x,y") : ideal"x,y"

o14 = ideal 1

o14 : Ideal of R
```

This last computation actually says we are done: since (x, y) kills everything inside M/M_3, we can now complete our prime filtration with

$$0 \underset{R/(x,z)}{\subseteq} M_1 = \frac{I+(xy)}{I} \underset{R/(x,z)}{\subseteq} M_2 = \frac{I+(y)}{I} \underset{R/(x,y)}{\subseteq} M_3 = \frac{I+(x,y)}{I} \underset{R/(x,y)}{\subseteq} M_4 = R/I.$$

Prime filtrations often allow us to reduce statements about finitely generated modules to statements about quotients of R that are also domains: modules of the form $R/\mathfrak{p}$ for primes $\mathfrak{p}$.

Corollary 5.30. *If R is a Noetherian ring, and M is a finitely generated module, and*

$$M = M_t \supsetneq M_{t-1} \supsetneq M_{t-2} \supsetneq \cdots \supsetneq M_1 \supsetneq M_0 = 0$$

is a prime filtration of M with $M_i/M_{i-1} \cong R/\mathfrak{p}_i$, then

$$\operatorname{Ass}_R(M) \subseteq \{\mathfrak{p}_1, \ldots, \mathfrak{p}_t\}.$$

Therefore,

- $\operatorname{Ass}_R(M)$ *is finite.*
- *If M is graded, then $\operatorname{Ass}_R(M)$ is a finite set of homogeneous primes.*

Proof. For each i, we have a short exact sequence

$$0 \longrightarrow M_{i-1} \longrightarrow M_i \longrightarrow M_i/M_{i-1} \longrightarrow 0\,.$$

By Lemma 5.23, $\operatorname{Ass}(M_i) \subseteq \operatorname{Ass}(M_{i-1}) \cup \operatorname{Ass}(M_i/M_{i-1}) = \operatorname{Ass}(M_{i-1}) \cup \{\mathfrak{p}_i\}$. Inductively, we have $\operatorname{Ass}(M_i) \subseteq \{\mathfrak{p}_1, \ldots, \mathfrak{p}_i\}$, and $\operatorname{Ass}_R(M) = \operatorname{Ass}_R(M_t) \subseteq \{\mathfrak{p}_1, \ldots, \mathfrak{p}_t\}$. This immediately implies that $\operatorname{Ass}(M)$ is finite. In the graded case, Theorem 5.28 gives us a filtration where all the $\mathfrak{p}_i$ are homogeneous primes, and those include all the associated primes. $\square$

Example 5.31. Any subset $X \subseteq \operatorname{Spec}(R)$ (for any R) can be realized as $\operatorname{Ass}(M)$ for some M: take $M = \bigoplus_{\mathfrak{p} \in X} R/\mathfrak{p}$. However, M is not finitely generated when X is infinite.

Example 5.32. If R is not Noetherian, then there may be modules (or ideals even) with no associated primes. Let $R = \bigcup_{n \in \mathbb{N}} \mathbb{C}[\![x^{1/n}]\!]$ be the ring of nonnegatively-valued Puiseux series. We claim that $R/(x)$ is a cyclic module with no associated primes, i.e., the ideal (x) has no associated primes. First, observe that any element of R can be written as a unit times $x^{m/n}$ for some m, n, so any associated prime of $R/(x)$ must be the annihilator of $x^{m/n} + (x)$ for some $m \leqslant n$. Hpwever, we claim that these are never prime. Indeed, we have $\operatorname{ann}(x^{m/n} + (x)) = (x^{1-m/n})$, which is not prime since $(x^{1/2-m/2n})^2 \in (x^{1-m/n})$ but $x^{1/2-m/2n} \notin (x^{1-m/n})$.

In a Noetherian ring, associated primes localize.

Theorem 5.33 (Associated primes localize in Noetherian rings)**.** *Let R be a Noetherian ring, W a multiplicative set, and M a module. Then*

$$\operatorname{Ass}_{W^{-1}R}(W^{-1}M) = \{W^{-1}\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Ass}_R(M),\, \mathfrak{p} \cap W = \varnothing\}.$$

Proof. Given $\mathfrak{p} \in \operatorname{Ass}_R(M)$ such that $\mathfrak{p} \cap W = \varnothing$, $W^{-1}\mathfrak{p}$ is a prime in $W^{-1}R$. Then $W^{-1}R/W^{-1}\mathfrak{p} \cong W^{-1}(R/\mathfrak{p}) \hookrightarrow W^{-1}M$ by exactness, so $W^{-1}\mathfrak{p}$ is an associated prime of $W^{-1}M$.

Suppose that $Q \in \operatorname{Spec}(W^{-1}R)$ is associated to $W^{-1}M$. We know this is of the form $W^{-1}\mathfrak{p}$ for some prime $\mathfrak{p}$ in R such that $\mathfrak{p} \cap W = \varnothing$. Since R is Noetherian, $\mathfrak{p}$ is finitely generated, say $\mathfrak{p} = (f_1, \dots, f_n)$ in R, and so $Q = \left(\frac{f_1}{1}, \dots, \frac{f_n}{1}\right)$.

By assumption, $Q = \operatorname{ann}(\frac{r}{w})$ for some $r \in R$, $w \in W$. Since w is a unit in $W^{-1}R$, we can also write $Q = \operatorname{ann}(\frac{r}{1})$. By definition, this means that for each i

$$\frac{f_i}{1}\frac{r}{1} = \frac{0}{1} \iff u_i f_i r = 0 \text{ for some } u_i \in W.$$

Let $u = u_1 \cdots u_n \in W$. Then $u f_i r = 0$ for all i, and thus $\mathfrak{p} u r = 0$. We claim that in fact $\mathfrak{p} = \operatorname{ann}(ur)$ in R. Consider $v \in \operatorname{ann}(ur)$. Then $u(vr) = 0$, and since $u \in W$, this implies that $\frac{vr}{1} = 0$. Therefore, $\frac{v}{1} \in \operatorname{ann}(\frac{r}{1}) = W^{-1}\mathfrak{p}$, and $vw \in \mathfrak{p}$ for some $w \in W$. But $\mathfrak{p} \cap W = \varnothing$, and thus $v \in \mathfrak{p}$. Thus $\mathfrak{p} \in \operatorname{Ass}(M)$. □

Corollary 5.34. *Let R be Noetherian, and M be an R-module.*

a) $\operatorname{Supp}_R(M) = \displaystyle\bigcup_{\mathfrak{p} \in \operatorname{Ass}_R(M)} V(\mathfrak{p})$.

b) If M is a finitely generated R-module, then $\operatorname{Min}(\operatorname{ann}_R(M)) \subseteq \operatorname{Ass}_R(M)$. *In particular,* $\operatorname{Min}(I) \subseteq \operatorname{Ass}_R(R/I)$.

Proof.

a) Let $\mathfrak{p} \in \operatorname{Ass}_R(M)$ and let $\mathfrak{p} = \operatorname{ann}_R(m)$ for $m \in M$. Let $\mathfrak{q} \in V(\mathfrak{p})$, which in particular implies that $\mathfrak{q} \in \operatorname{Supp}(R/\mathfrak{p})$, by Proposition 5.9. Since $0 \to R/\mathfrak{p} \xrightarrow{m} M$ is exact, so is $0 \to (R/\mathfrak{p})_\mathfrak{q} \to M_\mathfrak{q}$. Since $(R/\mathfrak{p})_\mathfrak{q} \neq 0$, we must also have $M_\mathfrak{q} \neq 0$, and thus $\mathfrak{q} \in \operatorname{Supp}(M)$.

Suppose that $\mathfrak{q} \notin \bigcup_{\mathfrak{p} \in \operatorname{Ass}_R(M)} V(\mathfrak{p})$, so that $\mathfrak{q}$ does not contain any associated prime of M. Then there is no associated prime of M that does not intersect $R \smallsetminus \mathfrak{q}$, so by Theorem 5.33, $\operatorname{Ass}_{R_\mathfrak{q}}(M_\mathfrak{q}) = \varnothing$. By Lemma 5.21, $M_\mathfrak{q} = 0$.

b) We have that $V(\operatorname{ann}_R(M)) = \operatorname{Supp}_R(M) = \bigcup_{\mathfrak{p} \in \operatorname{Ass}_R(M)} V(\mathfrak{p})$, so the minimal elements of both sets agree. In particular, the right hand side has the minimal primes of $\operatorname{ann}_R(M)$ as minimal elements, and they must be associated primes of M, or else this would contradict minimality.

$\square$

So the minimal primes of a module M are all associated to M, and they are precisely the minimal elements in the support of M.

Definition 5.35. If I is an ideal, then an associated prime of I that is not a minimal prime of I is called an **embedded prime** of I.

5.3 Prime Avoidance

We take a quick detour to discuss an important lemma.

Lemma 5.36 (Prime avoidance)**.** *Let R be a ring, $I_1, \ldots, I_n, J$ be ideals, and suppose that I_i is prime for $i > 2$.*[1] *If $J \nsubseteq I_i$ for all i, then $J \nsubseteq \bigcup_i I_i$. Equivalently, if $J \subseteq \bigcup_i I_i$, then $J \subseteq I_i$ for some i.*

Moreover, if R is $\mathbb{N}$-graded, and all of the ideals are homogeneous, all I_i are prime, and $J \nsubseteq I_i$ for all i, then there is a homogeneous element in J that is not in $\bigcup_i I_i$.

Proof. We proceed by induction on n. If $n = 1$, there is nothing to show.

By induction hypothesis, we can find elements a_i such that

$$a_i \notin \bigcup_{j \neq i} I_j \text{ and } a_i \in J$$

for each i. If some $a_i \notin I_i$, we are done, so let's assume that $a_i \in I_i$ for each i. Consider $a = a_n + a_1 \cdots a_{n-1} \in J$. Notice that $a_1 \cdots a_{n-1} = a_i(a_1 \cdots \widehat{a_i} \cdots a_{n-1}) \in I_i$. If $a \in I_i$ for $i < n$, then we also have $a_n \in I_i$, a contradiction. If $a \in I_n$, then we also have $a_1 \cdots a_{n-1} = a - a_n \in I_n$, since $a_n \in I_n$. If $n = 2$, this says $a_1 \in I_2$, a contradiction. If $n > 2$, our assumption is that I_n is prime, so one of $a_1, \ldots, a_{n-1} \in I_n$, which is a contradiction. So a is the element we were searching for, meaning $a \notin I_i$ for all i.

If all I_i are homogeneous and prime, then we proceed as above but replacing a_n and $a_1, \ldots, a_{n-1}$ with suitable powers so that $a_n + a_1 \cdots a_{n-1}$ is homogeneous. For example, we could take

$$a := a_n^{\deg(a_1) + \cdots \deg(a_{n-1})} + (a_1 \cdots a_{n-1})^{\deg(a_n)}.$$

The primeness assumption guarantees that noncontainments in ideals is preserved. $\square$

Corollary 5.37. *Let I be an ideal and M a finitely generated module over a Noetherian ring R. If I consists of zerodivisors on M, then $Im = 0$ for some nonzero $m \in M$.*

[1] So all the ideals are prime, except we may allow two of them to not be prime.

Proof. The assumption says that

$$I \subseteq \bigcup_{\mathfrak{p} \in \operatorname{Ass}(M)} (\mathfrak{p}).$$

By the assumptions, Corollary 5.30 applies, and it guarantees that this is a finite set of primes. By prime avoidance, $I \subseteq \mathfrak{p}$ for some $\mathfrak{p} \in \operatorname{Ass}(M)$. Equivalently, $I \subseteq \operatorname{ann}_R(m)$ for some nonzero $m \in M$. □

We will also need a slightly stronger version of Prime Avoidance.

Theorem 5.38. *Let R be a ring, $P_1, \ldots, P_n$ prime ideals, $x \in R$ and I be an ideal in R. If $(x) + I \nsubseteq P_i$ for each i, then there exists $y \in J$ such that*

$$x + y \notin \bigcup_{i=1}^{n} P_i.$$

Proof. We proceed by induction on n. When $n = 1$, if every element of the form $x + y$ with $y \in R$ is in $P = P_1$, then multiplying by $r \in R$ we conclude that every $rx + y \in P$, meaning $(x) + I \subseteq P$.

Now suppose $n > 1$ and that we have shown the statement for $n - 1$ primes. If $P_i \subseteq P_j$ for some $i \neq j$, then we might as well exclude P_i from our list of primes, and the statement follows by induction. So assume that all our primes P_i are incomparable.

If $x \notin P_i$ for all i, we are done, since we can take $x + 0$ for the element we are searching for. So suppose x is in some P_i, which we assume without loss of generality to be P_n. Our induction hypothesis says that we can find $y \in I$ such that $x + y \notin P_1 \cup \cdots \cup P_{n-1}$. If $x + y \notin P_n$, we are done, so suppose $x + y \in P_n$. Since we assumed $x \in P_n$, we must have $I \nsubseteq P_n$, or else we would have had $(x) + I \subseteq P_n$. Now P_n is a prime ideal that does not contain $P_1, \ldots, P_{n-1}$, nor I, so

$$P \nsupseteq I P_1 \cdots P_{n-1}.$$

Choose $z \in I P_1 \cdots P_{n-1}$ not in P_n. Then $x + y + z \notin P_n$, since $z \notin P_n$ but $x + y \in P_n$. Moreover, for all $i < n$ we have $x + y + z \notin P_i$, since $z \in P_i$ and $x + y \notin P_i$. □

5.4 Primary decomposition

We refine our decomposition theory once again, and introduce primary decompositions of ideals. One of the fundamental classical results in commutative algebra is the fact that every ideal in any noetherian ring has a primary decomposition. This can be thought of as a generalization of the Fundamental Theorem of Arithmetic:

Theorem 5.39 (Fundamental Theorem of Arithmetic)**.** *Every integer $n \in \mathbb{Z}$ can be written as a product of primes: there are distinct prime integers $p_1, \ldots, p_n$ and integers $a_1, \ldots, a_n \geqslant 1$ such that*

$$n = p_1^{a_1} \cdots p_n^{a_n}.$$

Moreover, such a a product is unique up to sign and the order of the factors.

We will soon discover that such a product *is* a primary decomposition, perhaps after some light rewriting. But before we get to the *what* and the *how* of primary decomposition, it is worth discussing the *why*. If we wanted to extend the Fundamental Theorem of Arithmetic to other rings, our first attempt might involve irreducible elements. Unfortunately, we don't have to go far to find rings where we *cannot* write elements as a unique product of irreducibles up to multiplying by a unit.

Example 5.40. In $\mathbb{Z}[\sqrt{-5}]$,

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

are two *different* ways to write 6 as a product of irreducible elements. In fact, we cannot obtain 2 or 3 by multiplying $1 + \sqrt{-5}$ or $1 + \sqrt{-5}$ by a unit.

Instead of writing *elements* as products of irreducibles, we will write *ideals* in terms of *primary ideals*.

Definition 5.41. We say that an ideal is **primary** if

$$xy \in I \implies x \in I \text{ or } y \in \sqrt{I}.$$

We say that an ideal is **$\mathfrak{p}$-primary**, where $\mathfrak{p}$ is prime, if I is primary and $\sqrt{I} = \mathfrak{p}$.

Remark 5.42. Note that a primary ideal has indeed a prime radical: if Q is primary, and $xy \in \sqrt{Q}$, then $x^n y^n \in Q$ for some n. If $y \notin \sqrt{Q}$, then we must have $x^n \in Q$, so $x \in \sqrt{Q}$. Thus, every primary ideal Q is $\sqrt{Q}$-primary.

Example 5.43.

a) Any prime ideal is also primary.

b) If R is a UFD, we claim that a principal ideal is primary if and only if it is generated by a power of a prime element. Indeed, if $a = f^n$, with f irreducible, then

$$xy \in (f^n) \iff f^n | xy \iff f^n | x \text{ or } f | y \iff x \in (f^n) \text{ or } y \in \sqrt{(f^n)} = (f).$$

Conversely, if a is not a prime power, then $a = gh$, for some g, h nonunits with no common factor, then take $gh \in (a)$ but $g \notin a$ and $h \notin \sqrt{(a)}$.

c) As a particular case of the previous example, the nonzero primary ideals in $\mathbb{Z}$ are of the form (p^n) for some prime p and some $n \geqslant 1$. This example is a bit misleading, as it suggests that primary ideals are the same as powers of primes. We will soon see that it not the case.

d) In $R = k[x, y, z]$, the ideal $I = (y^2, yz, z^2)$ is primary. Give R the grading with weights $|y| = |z| = 1$, and $|x| = 0$. If $g \notin \sqrt{I} = (y, z)$, then g has a degree zero term. If $f \notin I$, then f has a term of degree zero or one. The product fg has a term of degree zero or one, so is not in I.

If the radical of an ideal is prime, that does not imply that ideal is primary.

Example 5.44. In $R = k[x, y, z]$, the ideal $\mathfrak{q} = (x^2, xy)$ is not primary, even though $\sqrt{\mathfrak{q}} = (x)$ is prime. The offending product is xy.

The definition of primary can be reinterpreted in many forms.

Proposition 5.45. *If R is Noetherian, the following are equivalent :*

(1) $\mathfrak{q}$ is primary.

(2) Every zerodivisor in $R/\mathfrak{q}$ is nilpotent on $R/\mathfrak{q}$.

(3) $\operatorname{Ass}(R/\mathfrak{q})$ is a singleton.

(4) $\mathfrak{q}$ has exactly one minimal prime, and no embedded primes.

(5) $\sqrt{\mathfrak{q}} = \mathfrak{p}$ is prime and for all $r, w \in R$ with $w \notin \mathfrak{p}$, $rw \in \mathfrak{q}$ implies $r \in \mathfrak{q}$.

(6) $\sqrt{\mathfrak{q}} = \mathfrak{p}$ is prime, and $\mathfrak{q}R_{\mathfrak{p}} \cap R = \mathfrak{q}$.

Proof. (1) $\iff$ (2): y is a zerodivisor mod $\mathfrak{q}$ if there is some $x \notin \mathfrak{q}$ with $xy \in \mathfrak{q}$; the primary assumption translates to a power of y is in $\mathfrak{q}$.

(2) $\iff$ (3): On the one hand, (2) says that the set of zerodivisors on $R/\mathfrak{q}$ and coincide with the elements in the nilradical of $R/\mathfrak{q}$. By Lemma 5.21 and Exercise 16, respectively, these agree with the union of all the associated primes and the intersection of all the minimal primes.

$$\bigcup_{\mathfrak{p} \in \operatorname{Ass}(R/\mathfrak{q})} \mathfrak{p} = \mathcal{Z}(R/\mathfrak{q}) = \{r \in R \mid r + \mathfrak{q} \in \mathcal{N}(R/\mathfrak{q})\} = \bigcap_{\mathfrak{p} \in \operatorname{Min}(\mathfrak{q})} \mathfrak{p} = \bigcap_{\mathfrak{p} \in \operatorname{Ass}(R/\mathfrak{q})} \mathfrak{p}.$$

This holds if and only if there is only one associated prime.

(3) $\iff$ (4) is clear, since each statement is just a restatement of the other one.

(1) $\iff$ (5): Given the observation that the radical of a primary ideal is prime, this is just a rewording of the definition.

(5) $\iff$ (6): We secretly already know this from the discussion on behavior of ideals in localizations, in Proposition 4.27, which says that

$$\mathfrak{q}R_{\mathfrak{p}} \cap R = \{r \in R \mid rs \in \mathfrak{q} \text{ for some } s \notin \mathfrak{p}\}.$$

□

If the radical of an ideal is maximal, that *does* imply the ideal is primary.

Remark 5.46. Let I be an ideal with $\sqrt{I} = \mathfrak{m}$ a maximal ideal. If R is Noetherian, then $\operatorname{Ass}_R(R/I)$ is nonempty and contained in $\operatorname{Supp}(R/I) = V(I) = \{\mathfrak{m}\}$, so $\operatorname{Ass}_R(R/I) = \mathfrak{m}$, and hence I is primary.

Note that the assumption that $\mathfrak{m}$ is maximal was necessary here. Indeed, having a prime radical does not guarantee an ideal is primary, as we saw in Example 5.44. Moreover, even the powers of a prime ideal may fail to be primary.

Example 5.47. Let $R = k[x, y, z]/(xy - z^n)$, where k is a field and $n \geqslant 2$ is an integer. Consider the prime ideal $P = (x, z)$ in R, and note that $y \notin P$. On the one hand, $xy = z^n \in P^n$, while $x \notin P^n$ and $y \notin \sqrt{P^n} = P$. Therefore, P^n is not a primary ideal, even though its radical is the prime P.

The contraction of primary ideals is always primary.

Remark 5.48. Given any ring map $R \xrightarrow{f} S$, and a primary ideal Q in S, then the contraction of Q in R (via f) $Q \cap R$ is always primary. Indeed, if $xy \in Q \cap R$, and $x \notin Q \cap R$, then $f(x) \notin Q$, so $f(y^n) = f(y)^n \in Q$ for some n. Therefore, $y^n \in Q \cap R$, and $Q \cap R$ is indeed primary.

Lemma 5.49. *If $I_1, \ldots, I_t$ are ideals, then*

$$\operatorname{Ass}\left(R/\bigcap_{j=1}^{t} I_j\right) \subseteq \bigcup_{j=i}^{t} \operatorname{Ass}(R/I_j).$$

In particular, a finite intersection of $\mathfrak{p}$-primary ideals is $\mathfrak{p}$-primary.

Proof. There is an inclusion $R/(I_1 \cap I_2) \subseteq R/I_1 \oplus R/I_2$. Hence, by Lemma 5.23, $\operatorname{Ass}(R/(I_1 \cap I_2)) \subseteq \operatorname{Ass}(R/I_1) \cup \operatorname{Ass}(R/I_2)$; the statement for larger t is an easy induction.

If the I_j are all $\mathfrak{p}$-primary, then

$$\operatorname{Ass}(R/(\bigcap_{j=1}^{t} I_j)) \subseteq \bigcup_{j=i}^{t} \operatorname{Ass}(R/I_j) = \{\mathfrak{p}\}.$$

On the other hand, $\bigcap_{j=1}^{t} I_j \subseteq I_1 \neq R$, so $R/(\bigcap_{j=1}^{t} I_j) \neq 0$. Thus $\operatorname{Ass}(R/(\bigcap_{j=1}^{t} I_j))$ is nonempty, and therefore the singleton $\{\mathfrak{p}\}$. Then $\bigcap_{j=1}^{t} I_j$ is $\mathfrak{p}$-primary by the characterization of primary in Proposition 5.45 (3) above. $\square$

Definition 5.50 (Primary decomposition)**.** A **primary decomposition** of an ideal I is an expression of the form

$$I = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t,$$

with each $\mathfrak{q}_i$ primary. A **minimal primary decomposition** of an ideal I is a primary decomposition as above in which $\sqrt{\mathfrak{q}_i} \neq \sqrt{\mathfrak{q}_j}$ for $i \neq j$, and $\mathfrak{q}_i \not\supseteq \bigcap_{j \neq i} \mathfrak{q}_j$ for all i.[2]

Remark 5.51. By the previous lemma, we can turn any primary decomposition into a minimal one by combining the terms with the same radical, then removing redundant terms.

Example 5.52 (Primary decomposition in $\mathbb{Z}$)**.** Given a decomposition of $n \in \mathbb{Z}$ as a product of distinct primes, say $n = p_1^{a_1} \cdots p_k^{a_k}$, then the primary decomposition of the ideal (n) is $(n) = (p_1^{a_1}) \cap \cdots \cap (p_k^{a_k})$. However, this example can be deceiving, in that it suggests that primary ideals are just powers of primes; as we saw in Example 5.47 they are not!

The existence of primary decompositions was first shown by Emanuel Lasker (yes, the chess champion!) for polynomial rings and power series rings in 1905 [Las05], and then extended to Noetherian rings (which weren't called that yet at the time) by Emmy Noether in 1921 [Noe21].

[2]Some authors use the term *irredundant* instead of minimal.

Theorem 5.53 (Existence of primary decompositions)**.** *If R is Noetherian, then every ideal of R admits a primary decomposition.*

Proof. We will say that an ideal is irreducible if it cannot be written as a proper intersection of larger ideals. If R is Noetherian, we claim that any ideal of R can be expressed as a finite intersection of irreducible ideals. If the set of ideals that are not a finite intersection of irreducibles were non-empty, then by Noetherianity there would be an ideal maximal with the property of not being an intersection of irreducible ideals. Such a maximal element must be an intersection of two larger ideals, each of which are finite intersections of irreducibles, giving a contradiction.

Next, we claim that every irreducible ideal is primary. To prove the contrapositive, suppose that $\mathfrak{q}$ is not primary, and take $xy \in \mathfrak{q}$ with $x \notin \mathfrak{q}$, $y \notin \sqrt{\mathfrak{q}}$. The ascending chain of ideals

$$(\mathfrak{q} : y) \subseteq (\mathfrak{q} : y^2) \subseteq (\mathfrak{q} : y^3) \subseteq \cdots$$

stabilizes for some n, since R is Noetherian. This means that $y^{n+1} f \in \mathfrak{q} \implies y^n f \in \mathfrak{q}$. We will show that

$$(\mathfrak{q} + (y^n)) \cap (\mathfrak{q} + (x)) = \mathfrak{q},$$

proving that $\mathfrak{q}$ is not irreducible.

The containment $\mathfrak{q} \subseteq (\mathfrak{q} + (y^n)) \cap (\mathfrak{q} + (x))$ is clear. On the other hand, if

$$a \in (\mathfrak{q} + (y^n)) \cap (\mathfrak{q} + (x)),$$

we can write $a = q + by^n$ for some $q \in \mathfrak{q}$, and

$$a \in \mathfrak{q} + (x) \implies ay \in \mathfrak{q} + (xy) = \mathfrak{q}.$$

So

$$by^{n+1} = ay - aq \in \mathfrak{q} \implies b \in (\mathfrak{q} : y^{n+1}) = (\mathfrak{q} : y^n).$$

By definition, this means that $by^n \in \mathfrak{q}$, and thus $a = q + by^n \in \mathfrak{q}$. This shows that $\mathfrak{q}$ is not irreducible, concluding the proof. $\square$

Primary decompositions, even minimal ones, are not unique.

Example 5.54. Let $R = k[x, y]$, where k is a field, and $I = (x^2, xy)$. We can write

$$I = (x) \cap (x^2, xy, y^2) = (x) \cap (x^2, y).$$

These are two different minimal primary decompositions of I. To check this, we just need to see that each of the ideals (x^2, xy, y^2) and (x^2, y) are primary. Observe that each has radical $\mathfrak{m} = (x, y)$, which is maximal, so by an earlier remark, these ideals are both primary. In fact, our ideal I has infinitely many minimal primary decompositions: given any $n \geqslant 1$,

$$I = (x) \cap (x^2, xy, y^n)$$

is a minimal primary decomposition. One thing all of these have in common is the radicals of the primary components: they are always (x) and (x, y).

In the previous example, the fact that all our minimal primary decompositions had primary components always with the same radical was not an accident. Indeed, there are some aspects of primary decompositions that are unique, and this is one of them.

Theorem 5.55 (First uniqueness theorem for primary decompositions)**.** *Suppose I is an ideal in a Noetherian ring R. Given any minimal primary decomposition of I, say*

$$I = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t,$$

we have

$$\{\sqrt{\mathfrak{q}_1}, \ldots, \sqrt{\mathfrak{q}_t}\} = \operatorname{Ass}(R/I).$$

In particular, this set is the same for all minimal primary decompositions of I.

Proof. For any primary decomposition, minimal or not, we have

$$\operatorname{Ass}(R/I) \subseteq \bigcup_i \operatorname{Ass}(R/\mathfrak{q}_i) = \{\sqrt{\mathfrak{q}_1}, \ldots, \sqrt{\mathfrak{q}_t}\}$$

from the lemma on intersections we proved, Lemma 5.49. We just need to show that in a minimal decomposition as above, every $\mathfrak{p}_j := \sqrt{\mathfrak{q}_j}$ is an associated prime.

So fix j, and let

$$I_j = \bigcap_{i \neq j} \mathfrak{q}_i \supseteq I.$$

Since the decomposition is minimal, the module I_j/I is nonzero, hence by Lemma 5.21 it has an associated prime $\mathfrak{a}$. Let $\mathfrak{a}$ be such an associated prime, and fix $x_j \in R$ such that $\mathfrak{a}$ is the annihilator of $\overline{x_j}$ in I_j/I. Since

$$\mathfrak{q}_j x_j \subseteq \mathfrak{q}_j \cdot \bigcap_{i \neq j} \mathfrak{q}_i \subseteq \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_n = I,$$

we conclude that $\mathfrak{q}_j$ is contained in the annihilator of $\overline{x_j}$, meaning $\mathfrak{q}_j \subseteq \mathfrak{a}$. Since $\mathfrak{p}_j$ is the unique minimal prime of $\mathfrak{q}_j$ and $\mathfrak{a}$ is a prime containing $\mathfrak{q}_j$, we must have $\mathfrak{p}_j \subseteq \mathfrak{a}$. On the other hand, if $r \in \mathfrak{a}$, we have $rx_j \in I \subseteq \mathfrak{q}_j$, and since $x_j \notin \mathfrak{q}_j$, we must have $r \in \mathfrak{p}_j = \sqrt{\mathfrak{q}_j}$ by the definition of primary ideal. Thus $\mathfrak{a} \subseteq \mathfrak{p}_j$, so $\mathfrak{a} = \mathfrak{p}_j$. This shows that $\mathfrak{p}_j$ is an associated prime of R/I. $\square$

We note that if we don't assume that R is Noetherian, we may or may not have a primary decomposition for a given ideal. It is true that if an ideal I in a general ring has a primary decomposition, then the primes occurring are the same in any minimal decomposition. However, they are not the associated primes of I in general; rather, they are the primes that occur as radicals of annihilators of elements.

There is also a partial uniqueness result for the actual primary ideals that occur in a minimal decomposition.

Theorem 5.56 (Second uniqueness theorem for primary decompositions)**.** *If I is an ideal in a Noetherian ring R, then for any minimal primary decomposition of I, say $I = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$, the set of* minimal components $\{\mathfrak{q}_i \mid \sqrt{\mathfrak{q}_i} \in \operatorname{Min}(R/I)\}$ *is the same. Namely, $\mathfrak{q}_i = IR_{\sqrt{\mathfrak{q}_i}} \cap R$.*

Proof. We observe that a localization $\mathfrak{q}_{\mathbb{A}}$ of a $\mathfrak{p}$-primary ideal $\mathfrak{q}$ at a prime $\mathbb{A}$ is either the unit ideal (if $\mathfrak{p} \not\subseteq \mathbb{A}$), or a $\mathfrak{p}_{\mathbb{A}}$-primary ideal; this follows from the fact that the associated primes of $R/\mathfrak{q}$ localize, Theorem 5.33.

Now, since finite intersections commute with localization, then for any prime $\mathbb{A}$,

$$I_{\mathbb{A}} = (\mathfrak{q}_1)_{\mathbb{A}} \cap \cdots \cap (\mathfrak{q}_t)_{\mathbb{A}}$$

is a primary decomposition, although not necessarily minimal. In a minimal decomposition, choose a minimal prime $\mathbb{A} = \mathfrak{p}_i$. Then when we localize at $\mathbb{A}$, all the other components become the unit ideal since their radicals are not contained in $\mathfrak{p}_i$, and thus $I_{\mathfrak{p}_i} = (\mathfrak{q}_i)_{\mathfrak{p}_i}$t. We can then contract to R to get $I_{\mathfrak{p}_i} \cap R = (\mathfrak{q}_i)_{\mathfrak{p}_i} \cap R = \mathfrak{q}_i$, since $\mathfrak{q}_i$ is $\mathfrak{p}_i$-primary. $\square$

It is relatively easy to give a primary decomposition for a radical ideal:

Example 5.57. If R is Noetherian, and I is a radical ideal, then we have seen that I coincides with the intersection of its minimal primes $\mathfrak{p}_i$, meaning $I = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_t$. This is the *only* primary decomposition of a radical ideal.

For a more concrete example, take the ideal $I = (xy, xz, yz)$ in $k[x, y, z]$. This ideal is radical, so we just need to find its minimal primes. And indeed, one can check that $(xy, xz, yz) = (x, y) \cap (x, z) \cap (y, z)$. More generally, the radical monomial ideals are precisely those that are squarefree, and the primary components of a monomial ideal are also monomial.

Example 5.58. Let's get back to our motivating example in $\mathbb{Z}[\sqrt{-5}]$, where some elements can be written as products of irreducible elements in more than one way. For example, we saw that

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

So $(6) = (2) \cap (3)$, but while (2) is primary, (3) is not. In fact, (3) has two distinct minimal primes, and the following is a minimal primary decomposition for (6):

$$(6) = (2) \cap (3, 1 + \sqrt{-5}) \cap (3, 1 - \sqrt{-5}).$$

In fact, all of these come components are minimal, and so this primary decomposition is unique. Primary decomposition saves the day!

Finally, we note that the primary decompositions of powers of ideals are especially interesting.

Definition 5.59 (Symbolic power)**.** If $\mathfrak{p}$ is a prime ideal in a ring R, the n**th symbolic power** of $\mathfrak{p}$ is $\mathfrak{p}^{(n)} := \mathfrak{p}^n R_{\mathfrak{p}} \cap R$.

This admits equivalent characterizations.

Proposition 5.60. *Let R be Noetherian, and $\mathfrak{p}$ a prime ideal of R.*

a) $\mathfrak{p}^{(n)} = \{r \in R \mid rs \in \mathfrak{p}^n \text{ for some } s \notin \mathfrak{p}\}$.

b) $\mathfrak{p}^{(n)}$ *is the unique smallest* $\mathfrak{p}$*-primary ideal containing* $\mathfrak{p}^n$.

c) $\mathfrak{p}^{(n)}$ *is the* $\mathfrak{p}$*-primary component in any minimal primary decomposition of* $\mathfrak{p}^n$.

Proof. The first characterization follows from the definition, and the fact that expanding and contraction to/from a localization is equivalent to saturating with respect to the multiplicative set, which we proved in Proposition 4.27.

We know that $\mathfrak{p}^{(n)}$ is $\mathfrak{p}$-primary from one of the characterizations of primary we gave in Proposition 5.45. Any $\mathfrak{p}$-primary ideal satisfies $\mathfrak{q}R_{\mathfrak{p}} \cap R = \mathfrak{q}$, and if $\mathfrak{q} \supseteq \mathfrak{p}^n$, then $\mathfrak{p}^{(n)} = \mathfrak{p}^n R_{\mathfrak{p}} \cap R \subseteq \mathfrak{q}R_{\mathfrak{p}} \cap R = \mathfrak{q}$. Thus, $\mathfrak{p}^{(n)}$ is the unique smallest $\mathfrak{p}$-primary ideal containing $\mathfrak{p}^n$.

The last characterization follows from the second uniqueness theorem, Theorem 5.56. $\square$

In particular, note that $\mathfrak{p}^n = \mathfrak{p}^{(n)}$ if and only if $\mathfrak{p}^n$ is primary.

Example 5.61.

a) In $R = k[x, y, z]$, the prime $\mathfrak{p} = (y, z)$ satisfies $\mathfrak{p}^{(n)} = \mathfrak{p}^n$ for all n. This follows along the same lines as Example 5.43 d.

b) In $R = k[x, y, z] = (xy - z^n)$, where $n \geqslant 2$, we have seen in Example 5.47 that the square of $\mathfrak{p} = (y, z)$ is not primary, and therefore $\mathfrak{p}^{(2)} \neq \mathfrak{p}^2$. Indeed, $xy = z^n \in \mathfrak{p}^2$, and $x \notin \mathfrak{p}$, so $y \in \mathfrak{p}^{(2)}$ but $y \notin \mathfrak{p}^2$.

c) Let $X = X_{3\times 3}$ be a 3×3 matrix of indeterminates, and $k[X]$ be a polynomial ring over a field k. Let $\mathfrak{p} = I_2(X)$ be the ideal generated by 2×2 minors of X. Write $\Delta_{\substack{i|k\\ j|l}}$ for the determinant of the submatrix with rows i, j and columns k, l. We find

$$\begin{aligned} x_{11}\det(X) &= x_{11}x_{31}\Delta_{\substack{1|2\\2|3}} - x_{11}x_{32}\Delta_{\substack{1|1\\2|3}} + x_{11}x_{33}\Delta_{\substack{1|1\\2|2}} \\ &= (x_{11}x_{31}\Delta_{\substack{1|2\\2|3}} - x_{11}x_{32}\Delta_{\substack{1|1\\2|3}} + x_{11}x_{33}\Delta_{\substack{1|1\\2|2}}) \\ &\quad - (x_{11}x_{31}\Delta_{\substack{1|2\\2|3}} - x_{12}x_{31}\Delta_{\substack{1|1\\2|3}} + x_{13}x_{31}\Delta_{\substack{1|1\\2|2}}) \\ &= -\Delta_{\substack{1|1\\3|2}}\Delta_{\substack{1|1\\2|3}} + \Delta_{\substack{1|1\\3|3}}\Delta_{\substack{1|1\\2|2}} \in I_2(X)^2. \end{aligned}$$

Note that in the second row, we subtracted the Laplace expansion of the determinant of the matrix with row 3 replaced by another copy of row 1. That is, we subtracted zero.

While we will not discuss symbolic powers in detail, they are ubiquitous in commutative algebra. They show up as tools to prove various important theorems of different flavors, and they are also interesting objects in their own right. In particular, symbolic powers can be interpreted from a geometric perspective, via the Zariski–Nagata Theorem [Zar49, NM91]. Roughly, this theorem says that when we consider symbolic powers of prime ideals over $\mathbb{C}[x_1, \ldots, x_d]$, the polynomials in $\mathfrak{p}^{(n)}$ are precisely the polynomials that vanish *to order* n on the variety corresponding to $\mathfrak{p}$. This result can be made sense of more generally, for any radical ideal in $\mathbb{C}[x_1, \ldots, x_d]$ over any perfect field k [EH79, FMS14], and even when $k = \mathbb{Z}$ [DSGJ20].

5.5 The Krull Intersection Theorem

Lemma 5.62. *Let R be a ring. If $I \subseteq J$ are ideals, $J \subseteq \sqrt{I}$, and J is finitely generated, then there is some n with $J^n \subseteq I$. Therefore, if R is Noetherian, for every ideal I, there is some n with $\sqrt{I}^n \subseteq I$.*

Proof. Write $J = (f_1, \dots, f_m)$. By definition, each $f_i^{a_i} \in I$ for some $a_1, \dots, a_m$. Let $n := a_1 + \cdots + a_m + 1$. Now J^n is generated by products of the form $f_1^{b_1} \cdots f_m^{b_m}$ with $b_1 + \cdots + b_m = n$. By the Pigeonhole Principle, at least one b_i satisfies $b_i \geqslant a_i$, so $f_1^{b_1} \cdots f_m^{b_m} \in I$.

The second statement is a consequence of the first, since $\sqrt{I}$ is a finitely generated ideal with $\sqrt{I} \supseteq I$. $\square$

Theorem 5.63 (Krull intersection theorem)**.** *Let $(R, \mathfrak{m}, k)$ be a Noetherian local ring. Then*

$$\bigcap_{n \geqslant 1} \mathfrak{m}^n = 0.$$

Proof. Let $J = \bigcap_{n \in \mathbb{N}} \mathfrak{m}^n$. First, we claim that $J \subseteq \mathfrak{m}J$.

Let $\mathfrak{m}J = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$ be a primary decomposition. To show that $J \subseteq \mathfrak{m}J$, it is sufficient to prove that $J \subseteq \mathfrak{q}_i$ for each i. If $\sqrt{\mathfrak{q}_i} \neq \mathfrak{m}$, pick $x \in \mathfrak{m}$ such that $x \notin \sqrt{\mathfrak{q}_i}$. Then $xJ \subseteq \mathfrak{m}J \subseteq \mathfrak{q}_i$, but $x \notin \sqrt{\mathfrak{q}_i}$, so $J \subseteq \mathfrak{q}_i$ by definition of primary. If instead $\sqrt{\mathfrak{q}_i} = \mathfrak{m}$, there is some N with $\mathfrak{m}^N \subseteq \mathfrak{q}_i$ by Lemma 5.62. By definition of J, we have $J \subseteq \mathfrak{m}^N \subseteq \mathfrak{q}_i$, and we are done.

We showed that $J \subseteq \mathfrak{m}J$, hence $J = \mathfrak{m}J$, and thus $J = 0$ by NAK 4.30. $\square$

Remark 5.64. As an easy corollary, we obtain that

$$\bigcap_{n \geqslant 1} I^n = 0$$

for any proper ideal I in a Noetherian local ring $(R, \mathfrak{m})$, since $I^n \subseteq \mathfrak{m}^n$ for all n.

In the non-local setting, it is not true in general that $\bigcap_{n \geqslant 1} I^n = 0$.

Exercise 17. Let k be a field and let $R = k \times k$ be the product of k with itself. Show that the ideal $I = \{(a, 0) \mid a \in k\}$ is **idempotent**, meaning $I^2 = I$, and thus $\bigcap_{n \geqslant 1} I^n = I \neq 0$.

But it is true if R is a domain.

Theorem 5.65 (Krull Intersection Theorem for domains)**.** *If R is a domain, then*

$$\bigcap_{n \geqslant 1} I^n = 0.$$

for any proper ideal I in R.

Proof. Exercise. $\square$

Chapter 6

Dimension theory

6.1 Dimension and height

Definition 6.1. A **chain of primes** of **length** n in a ring R is a chain

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n \quad \text{with } \mathfrak{p}_i \in \operatorname{Spec}(R).$$

We say a chain of primes is **saturated** if for each i, there is no $\mathfrak{q} \in \operatorname{Spec}(R)$ with $\mathfrak{p}_i \subsetneq \mathfrak{q} \subsetneq \mathfrak{p}_{i+1}$. The **dimension** or **Krull dimension** of a ring R is the supremum of the lengths of chains of primes in R. Equivalently, it is the supremum of the lengths of saturated chains of primes in R. We denote the Krull dimension of R by $\dim(R)$.

The **height** of a prime $\mathfrak{p}$ is the supremum of the lengths of chains of primes in R that end in $\mathfrak{p}$, i.e., with $\mathfrak{p} = \mathfrak{p}_n$ above. Equivalently, it is the supremum of the lengths of saturated chains of primes in R that end in $\mathfrak{p}$. We denote the height of $\mathfrak{p}$ by $\operatorname{ht}(\mathfrak{p})$. The **height** of an ideal I is the infimum of the heights of the minimal primes of I:

$$\operatorname{ht}(I) := \inf \{\operatorname{ht}(\mathfrak{p}) \mid \mathfrak{p} \in \operatorname{Min}(I)\}$$

To get a feel for these definitions, here are some easy observations.

Remark 6.2.

a) If $\mathfrak{p}$ is prime, then $\dim(R/\mathfrak{p})$ is the supremum of the lengths of (saturated) chains of primes in R

$$\mathfrak{q}_0 \subsetneq \mathfrak{q}_1 \subsetneq \cdots \subsetneq \mathfrak{q}_n$$

with each $\mathfrak{q}_i \in V(\mathfrak{p})$.

b) If I is an ideal, then $\dim(R/I)$ is the supremum of the lengths of (saturated) chains of primes in R

$$\mathfrak{q}_0 \subsetneq \mathfrak{q}_1 \subsetneq \cdots \subsetneq \mathfrak{q}_n$$

with each $\mathfrak{q}_i \in V(I)$.

c) If W is a multiplicative set, then $\dim(W^{-1}R) \leqslant \dim(R)$.

d) If $\mathfrak{p}$ is prime, then $\dim(R_{\mathfrak{p}}) = \operatorname{ht}(\mathfrak{p})$.

e) If $\mathfrak{q} \supseteq \mathfrak{p}$ are primes, then $\dim(R_{\mathfrak{q}}/\mathfrak{p}R_{\mathfrak{q}})$ is the supremum of the lengths of (saturated) chains of primes in R
$$\mathfrak{p} = \mathfrak{a}_0 \subsetneq \mathfrak{a}_1 \subsetneq \cdots \subsetneq \mathfrak{a}_n = \mathfrak{q}.$$

f) $\dim(R) = \sup\{\text{height}(\mathfrak{m}) \mid \mathfrak{m} \in \text{mSpec}(R)\}$.

g) $\dim(R) = \sup\{\dim(R/\mathfrak{p}) \mid \mathfrak{p} \in \text{Min}(R)\}$.

h) If $\mathfrak{p}$ is prime, $\dim(R/\mathfrak{p}) + \text{height}(\mathfrak{p}) \leqslant \dim(R)$.

i) If I is an ideal, $\dim(R/I) + \text{height}(I) \leqslant \dim(R)$.

j) The ideal (0) has height 0.

k) A prime has height zero if and only if it is a minimal prime.

We will need a few theorems before we compute the height and dimension of many examples, but we can handle a few basic cases.

Example 6.3.

a) The dimension of a field is zero.

b) A ring is zero-dimensional if and only if every minimal prime is maximal.

c) The ring of integers $\mathbb{Z}$ has dimension 1: there is one minimal prime (0) and every other prime is maximal. Likewise, a principal ideal domain has dimension 1.

d) In a UFD, I is a prime of height 1 if and only if $I = (f)$ for a prime element f.
To see this, note that if $I = (f)$ with f irreducible, and $0 \subsetneq \mathfrak{p} \subseteq I$, then $\mathfrak{p}$ contains some nonzero multiple of f, say af^n with a and f coprime. Since $a \notin I$, $a \notin \mathfrak{p}$, so we must have $f \in \mathfrak{p}$, so $\mathfrak{p} = (f)$. Thus, I has height one. On the other hand, if I is a prime of height one, we claim I contains an irreducible element. Indeed, I is nonzero, so it contains some $f \neq 0$, and primeness implies one of the prime factors of f is contained in I. Thus, any nonzero prime contains a prime ideal of the form (f), so a height one prime must be of this form.

e) If k is a field, then $\dim(k[x_1, \ldots, x_d]) \geqslant d$, since there is a saturated chain of primes $(0) \subsetneq (x_1) \subsetneq (x_1, x_2) \subsetneq \cdots \subsetneq (x_1, \ldots, x_d)$.

We pose a related definition for modules.

Definition 6.4. The **dimension** of an R-module M is defined as $\dim(R/\operatorname{ann}_R(M))$.

Note that if M is finitely generated, $\dim(M)$ is the same as the supremum of the lengths of chains of primes in $\text{Supp}_R(M)$.

Definition 6.5. A ring is **catenary** if for every pair of primes $\mathfrak{q} \supseteq \mathfrak{p}$ in R, every saturated chain of primes
$$\mathfrak{p} = P_0 \subsetneq P_1 \subsetneq \cdots \subsetneq P_n = \mathfrak{q}$$
has the same length. A ring is **equidimensional** if every maximal ideal has the same finite height, or equivalently $\dim(R/P)$ is the same finite number for every minimal prime P.

Here are some examples of what can go wrong.

Example 6.6. Consider the ring

$$R = \frac{k[x,y,z]}{(xy,xz)}.$$

We can find the minimal primes of R by computing $\mathrm{Min}((xy, xz))$ in $k[x, y, z]$: (x) and (y, z) are prime, and $(x) \cap (y, z) = (xy, xz)$. Therefore, $\mathrm{Min}(R) = \{(x), (y, z)\}$. Now, the height of $(x-1, y, z)$ is one: it contains the minimal prime (y, z), and any saturated chain from (y, z) to $(x-1, y, z)$ corresponds to a saturated chain from (0) to $(x-1)$ in $K[x]$, which must have length 1 since this is a PID. The height of $(x, y-1, z)$ is at least 2, as witnessed by the chain $(x) \subseteq (x, y-1) \subseteq (x, y-1, z)$. So R is not equidimensional.

Even domains may fail to be equidimensional.

Example 6.7. The ring $\mathbb{Z}_{(2)}[x]$ is a domain that is not equidimensional. On the one hand, the maximal ideal $(2, x)$ has height at least two, which we see from the chain

$$(0) \subsetneq (x) \subseteq (n, 2).$$

On the other hand, the maximal ideal $(2x-1)$ has height 1; this is maximal since the quotient is $\mathbb{Q}$!

Remark 6.8.

a) If R is a finite dimensional domain, and $f \neq 0$, then $\dim(R/(f)) < \dim(R)$.

b) If R is equidimensional, then $\dim(R/(f)) < \dim(R)$ if and only if $f \notin \bigcup\limits_{\mathfrak{p} \in \mathrm{Min}(R)} \mathfrak{p}$.

c) In general, $\dim(R/(f)) < \dim(R)$ if and only if $f \notin \bigcup\limits_{\substack{\mathfrak{p} \in \mathrm{Min}(R) \\ \dim(R/\mathfrak{p}) = \dim(R)}} \mathfrak{p}$.

d) $f \notin \bigcup\limits_{\mathfrak{p} \in \mathrm{Min}(R)} \mathfrak{p}$ if and only if $\dim(R/(\mathfrak{p} + (f))) < \dim(R/\mathfrak{p})$ for all $\mathfrak{p} \in \mathrm{Min}(R)$.

Before we get too optimistic, know that there are Noetherian rings of infinite dimension, as the following example due to Nagata [Nag62, Appendix, Example 1] shows.

Example 6.9. The ring $R = k[x_1, x_2, \ldots]$ is infinite-dimensional. Let

$$W = R \smallsetminus ((x_1) \cup (x_2, x_3) \cup (x_4, x_5, x_6) \cdots)$$

and $S = W^{-1}R$. This ring has primes of arbitrarily large height, given by the images of those primes we cut out from W. Thus, it has infinite dimension. The work is to show that this ring is Noetherian. We omit this argument here.

Note also that a ring might have finite dimension but not be Noetherian.

Example 6.10. Let $R = k[x_1, x_2, \ldots]/(x_1^2, x_2^2, \ldots)$. On the one hand, R is not Noetherian, since

$$(x_1) \subseteq (x_1, x_2) \subseteq (x_1, x_2, x_3) \subseteq \cdots$$

is an infinite ascending chain of ideals. On the other hand, R has only one prime ideal, and thus $\dim(R) = 0$.

6.2 Artinian rings

To prepare for our next big theorems in dimension theory, we need to understand the structure of zero-dimensional Noetherian rings. In order to do that, we will take a theorem on primary decomposition for certain ideals in not necessarily Noetherian rings.

Theorem 6.11. *Let R be a ring, not necessarily Noetherian. Let I be an ideal such that $V(I) = \{\mathfrak{m}_1, \ldots, \mathfrak{m}_t\}$ is a finite set of maximal ideals. There is a primary decomposition $I = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$, and moreover $I = \mathfrak{q}_1 \cdots \mathfrak{q}_t$ and $R/I \cong R/\mathfrak{q}_1 \times \cdots \times R/\mathfrak{q}_t$.*

Proof. First, we claim that $IR_{\mathfrak{m}_i}$ is $\mathfrak{m}_i R_{\mathfrak{m}_i}$-primary. The local ring $(R/I)_{\mathfrak{m}_i} = R_{\mathfrak{m}_i}/IR_{\mathfrak{m}_i}$ has a unique maximal ideal $\mathfrak{m}_i R_{\mathfrak{m}_i}/IR_{\mathfrak{m}_i}$, so if $x, y \in R_{\mathfrak{m}_i}$ are such that $xy \in IR_{\mathfrak{m}_i}$, and $x \notin \mathfrak{m}_i R_{\mathfrak{m}_i}$, then x is a unit modulo $I_{\mathfrak{m}_i}$, so $y \in IR_{\mathfrak{m}_i}$. Now the contraction of a primary ideal is primary, by Remark 5.48, so $\mathfrak{q}_i = IR_{\mathfrak{m}_i} \cap R$ is $\mathfrak{m}_i$-primary, and $I \subseteq \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$.

On the other hand, equality of these modules is a local property, so let's check it at each prime. When $\mathfrak{p} \notin V(I)$, then both I and $\mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$ are the unit ideal in $R_{\mathfrak{p}}$. On the other hand, for each $\mathfrak{m}_i \in V(I)$, $I_{\mathfrak{m}_i} = (\mathfrak{q}_i)_{\mathfrak{m}_i}$ in $R_{\mathfrak{m}_i}$. Therefore, $I = \mathfrak{q}_1 \cdots \mathfrak{q}_t$, and this is a primary decomposition, since each $\mathfrak{q}_i$ is $\mathfrak{m}_i$-primary.

Now notice that $\mathfrak{q}_i + \mathfrak{q}_j = R$ for each pair $\mathfrak{q}_i \neq \mathfrak{q}_j$, so Theorem 0.8 applies. Therefore, we obtain the fact that our intersection is a product and the quotient ring is a direct product as a consequence of Theorem 0.8. $\square$

Definition 6.12. A module $M \neq 0$ is **simple** if its only submodules are (0) and M.

Remark 6.13. If $\mathfrak{m}$ is a maximal ideal in R, then $R/\mathfrak{m}$ is a simple R-module. On the other hand, if M is any cyclic R-module, then given any nonzero element $m \in M$, Rm is a nonzero submodule of M, and thus it must be all of M. Therefore, any simple module must be cyclic. If I is a proper ideal contained in some maximal ideal $\mathfrak{m} \supsetneq I$, then $\mathfrak{m}/I$ is a proper nonzero submodule of R/I. Therefore, the simple modules of any ring R are precisely those that are isomorphic to $R/\mathfrak{m}$ for some maximal ideal $\mathfrak{m}$.

In particular, if R is a local, then R has only one simple module up to isomorphism: the residue field.

Definition 6.14. A module M has **finite length** if it has a filtration of the form

$$0 = M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots \subseteq M_n = M$$

with M_{i+1}/M_i simple for each i; such a filtration is called a **composition series** of **length** n. We say a composition series is **strict** if $M_i \neq M_{i+1}$ for all i. Two composition series are **equivalent** if the collections of composition factors M_{i+1}/M_i are the same up to reordering. The **length** of a finite length module M, denoted $\ell(M)$, is the minimum of the lengths of a composition series of M. If M has does not have finite length, we say that M has infinite length, or $\ell(M) = \infty$.

You may have seen the Jordan–Holder theorem in the context of groups:

Theorem 6.15 (Jordan–Holder theorem)**.** *Let M be a module of finite length.*

1) *Any proper submodule N of M has $\ell(N) < \ell(M)$.*
2) *Any filtration of M can be refined to a composition series.*
3) *All strict composition series for M are equivalent, and hence have the same length.*

Proof. If $n := \ell(M)$, consider a strict composition series of M of length n, say

$$0 = M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots \subseteq M_n = M.$$

1) Given a submodule N of M, consider the filtration

$$0 = M_0 \cap N \subseteq M_1 \cap N \subseteq M_2 \cap N \subseteq \cdots \subseteq M_n \cap N = N.$$

By the Second Isomorphism Theorem, its composition factors satisfy

$$(M_{i+1} \cap N)/(M_i \cap N) \cong (M_{i+1} \cap N + M_i)/M_i.$$

This is a submodule of M_{i+1}/M_i, which by assumption is simple. Then either

$$(M_{i+1} \cap N + M_i)/M_i = 0 \quad \text{or} \quad (M_{i+1} \cap N + M_i)/M_i = M_{i+1}/M_i.$$

The quotients that are zero correspond to terms that we can delete; the remaining ones are simple modules. The resulting filtration is a strict composition series for N, so this shows that $\ell(N) \leqslant n$. Moreover, if there are no zero coefficients to delete, then

$$(M_{i+1} \cap N + M_i)/M_i = M_{i+1}/M_i$$

for all i, and in particular when we take $i + 1 = n$ we obtain

$$N + M_{n-1} = M_n \cap N + M_{n-1} = M_n = M,$$

Since M/M_{n-1} is simple by assumption, we must have $N = M$. If N is a proper submodule, this cannot happen, and thus at least one of the terms can be deleted, so $\ell(N) < n$.

2) Let us use induction on n to show that any chain of submodules of M, say

$$N_0 \subsetneq N_1 \subseteq M_2 \subsetneq \cdots \subsetneq N_k,$$

has length at most n. If $n = 0$, then $M = 0$ and there is nothing to prove. Now assume that $n \geqslant 1$ and that the statement holds for modules of length $< n$. Since N_{k-1} must be a proper submodule of N, 1) tells us that $\ell(N_{k-1}) < n$, and thus $k - 1 \leqslant n - 1$. Therefore, $k \leqslant n$. This shows our claim that all chains of submodules of M have length at most n.

Now notice that if we are given a chain of submodules of length $< n$, then it cannot be a compositions series, since by definition the smallest composition series has length n. That means that some of the quotients are not simple, so we can extend the chain by adding a term, as follows: if N_{i+1}/N_i is not simple, and N'/N_i is a.proper nonzero submodule, then $N_i \subseteq N_{i+1}$ can be extended $N_i \subseteq N \subseteq N_{i+1}$. Therefore, every chain of submodules can be extended to a composition series.

3) Suppose that

$$0 = N_0 \subseteq N_1 \subseteq \cdots \subseteq N_k = M$$

is a strict composition series of $M \neq 0$. Since $\ell(M)$ is the smallest possible length of a composition series, we have $\ell(M) \leqslant k$. Moreover, notice that for each $i \geqslant 1$, N_i is a proper submodule of N_{i+1}, and thus by 1) we have

$$\ell(M) > \ell(N_{k-1}) > \ell(N_{k-1}) > \cdots > \ell(N_1) > 0.$$

Therefore, $\ell(M) \geqslant k$, so we must have $\ell(M) = \ell(k)$. □

Let's collect some basic consequences of this theorem.

Lemma 6.16. *Length is associative on short exact sequences, that is, if*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is a short exact sequence of R-modules, then $\ell(B) = \ell(A) + \ell(B)$.

Proof. Given filtrations of lengths a and c for A and C, respectively, we can construct a filtration for B of length $a + c$, so $\ell(B) \geqslant \ell(A) + \ell(C)$. On the other hand, if

$$0 = B_0 \subseteq B_1 \subseteq B_2 \subseteq \cdots \subseteq B_n = B$$

is a filtration for B, then $B_i \cap A$ and $g(B_i)$ are filtrations for A and C, respectively. Suppose that both $g(B_i) = g(B_{i+1}$ and $B_i \cap A = B_{i+1} \cap A$, and let $b \in B_{i+1}$. Then $g(b) \in g(B_i)$, so there is $b' \in B_i$ such that $b - b' \in \ker g = \operatorname{im} f$. Since b and b' are both in B_{i+1}, we conclude that $b - b' \in B_{i+1} \cap A = B_i \cap A$. But $b' \in B_i$, so we conclude that $b \in B_i$. Therefore, $B_i = B_{i+1}$. This shows that sum of the lengths of the filtrations $B_i \cap A$ and $g(B_i)$ is at most the length of the filtration B_i. We conclude that $\ell(B) \leqslant \ell(A) + \ell(C)$. □

Using an homological trick we haven't seen yet, one can actually show that if

$$0 \longrightarrow A_1 \longrightarrow \cdots \longrightarrow A_n \longrightarrow 0\ ,$$

is an exact sequence, then $\displaystyle\sum_{i=1}^{n} \ell(A_i) = 0$.

Remark 6.17.

a) Given a chain of submodules $0 = M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots \subseteq M_n = M$,

$$\ell(M) = \sum_{i=0}^{n-1} \ell(M_{i+1}/M_i).$$

b) If $M \subseteq N$, then $\ell(M) \leqslant \ell(N)$, with equality only when $M = N$.

Remark 6.18. If M is annihilated by a maximal ideal $\mathfrak{m}$, so that M is an $R/\mathfrak{m}$-module, then $\ell(M) = \dim_{R/\mathfrak{m}}(M/\mathfrak{m}M)$.

Example 6.19. Let $R = \mathbb{R}[x,y]_{(x,y)}$. Then $M = R/\mathfrak{m}^2$ has length 3, since we have a composition series $0 \subseteq xM \subseteq (x,y)M \subseteq M$. However, M is not an $R/\mathfrak{m}$-vector space.

Back when we discussed Noetherian rings, we could have also considered the dual notion of Artinian rings. The reason we have waited so long to do so is that as we will soon show, Artinian rings are just Noetherian rings of dimension 0.

Definition 6.20. A ring is **Artinian** if every descending chain of ideals eventually stabilizes. A module is **Artinian** if every descending chain of submodules eventually stabilizes.

Adapting the proofs of the analogous statements for Noetherian rings and modules, one can easily show the following:

Exercise 18.

a) If R is an Artinian ring, then R/I is Artinian for any ideal I of R.
b) If R is an Artinian ring, then any nonempty family of ideals has a minimal element.
c) If M is an Artinian module, and $N \subseteq M$, then N and M/N are Artinian.

Lemma 6.21. *A module M has finite length if and only if it is both Noetherian and Artinian.*

Proof. If M has finite length, then all chains of submodules of M must have length at most $\ell(M)$, and thus in particular all ascending and descending chains of submodules must stabilize.

On the other hand, suppose that M is both Noetherian and Artinian. If $M = 0$, there is nothing to show, so we might as well assume $M \neq 0$. The set of proper submodules of M is then be nonempty, and thus it has a maximal element M_1 by Noetherianity. This forces M/M_1 to be simple, so we can start constructing a composition series for M by taking $M \supseteq M_1$. At each step, if we have constructed modules

$$M_0 = M \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_k$$

such that M_i/M_{i+1} is simple, either $M_k = 0$ and we can stop, or $M_k \neq 0$ and it has a proper submodule. Repeating the initial construction for M_k, which is again Noetherian, we can continue to build a descending chain of submodules of M. But M is Artinian, and thus this process must eventually stop, since M is Artinian. $\square$

Over a field, things are a bit easier.

Remark 6.22. If M is a k-vector space, M is Artinian if and only it is Noetherian if and only if it has finite length.

Lemma 6.23. *Let R be a Noetherian ring. An R-module M has finite length if and only if M is finitely generated and* $\dim(M) = \dim(R/\operatorname{ann}(M)) = 0$.

Proof. Suppose M has finite length. Then M is Noetherian, by Lemma 6.21, and in particular finitely generated. Moreover, consider a composition series

$$0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots \subsetneq M_n = M$$

for M. For each $i \geqslant 1$, $M_i/M_{i-1} \cong R/\mathfrak{m}_i$ for some maximal ideal $\mathfrak{m}_i$. Also, our composition series breaks into short exact sequences

$$0 \longrightarrow M_i \longrightarrow M_{i+1} \longrightarrow M_{i+1}/M_i \longrightarrow 0\,.$$

When $i = 1$, $M_1 \cong M_1/M_0 \cong R/\mathfrak{m}_1$. Using Lemma 5.23 repeatedly, we conclude that $\operatorname{Ass}(M_i) \subseteq \{\mathfrak{m}_1, \ldots, \mathfrak{m}_i\}$ for each i, and in particular $\operatorname{Ass}(M) \subseteq \{\mathfrak{m}_1, \ldots, \mathfrak{m}_i\}$. So all the minimal primes over $\operatorname{ann}(M)$ are maximal, and $\dim(R/\operatorname{ann}(M)) = 0$.

Now suppose that M is finitely generated and $\dim(R/\operatorname{ann}(M)) = 0$. Since M is finitely generated over a Noetherian ring, by Theorem 5.28 there exists a filtration of M

$$M = M_t \supsetneq M_{t-1} \supsetneq M_{t-2} \supsetneq \cdots \supsetneq M_1 \supsetneq M_0 = 0$$

such that $M_i/M_{i-1} \cong R/\mathfrak{p}_i$ for primes $\mathfrak{p}_i \in \operatorname{Spec}(R)$. For each i,

$$\operatorname{ann}(M)M_i = 0 \subseteq M_{i-1} \implies \operatorname{ann}(M) \subseteq (M_{i-1} :_R M_i) = \mathfrak{p}_i.$$

Since $\dim(R/\operatorname{ann}(M)) = 0$, all the primes containing $\operatorname{ann}(M)$ must be maximal, and thus the $\mathfrak{p}_i$ are all maximal ideals. So our prime filtration is a composition series, and M has finite length. □

Equivalently, an R-module M over a Noetherian ring has finite length if and only if it is finitely generated and all of its associated primes are maximal ideals of R.

Exercise 19. Let $(R, \mathfrak{m})$ be a Noetherian local ring. An R-module M has finite length if and only if M is finitely generated and $\mathfrak{m}^n M = 0$ for some n.

Example 6.24. Let $(R, \mathfrak{m})$ be a local ring. Then $M = (R/\mathfrak{m})^n$ is a finite length module for any $n \geqslant 1$. Note that $\ell(M) = \dim_{R/\mathfrak{m}}((R/m)^n) = n$, while $\mathfrak{m}M = 0$.

Finally, we can show that Artinian rings are just zero-dimensional Noetherian rings.

Theorem 6.25. *The following are equivalent:*

a) R is Noetherian of dimension zero.

b) R is a finite product of local Noetherian rings of dimension zero.

c) R has finite length as an R-module.

d) R is Artinian.

Proof. (1)⇒(2): Since R is Noetherian of dimension zero, every prime is maximal and minimal. Since there are finitely many minimal primes in R, by Theorem 5.5, there are finitely many primes in R. By Theorem 6.11, R decomposes as a direct product of Noetherian local rings, which all must have dimension zero.

(2)⇒(3): It suffices to deal with the case when $(R, \mathfrak{m})$ is a local Noetherian ring of dimension 0. In this case, the maximal ideal is the unique minimal prime, so $\mathfrak{m} = \sqrt{(0)}$. Since R is Noetherian, Lemma 5.62 yields $\mathfrak{m}^n = 0$ for some n. Then R has finite length by Exercise 19.

(3)⇒(4): This follows by Lemma 6.21, noting that R is an Artinian R-module if and only if R is an Artinian ring.

(4)⇒(1):

First we show that R has dimension zero. If $\mathfrak{p}$ is any prime, then $R/\mathfrak{p}$ is Artinian, since the ideals of $R/\mathfrak{p}$ are in bijection with the ideals of R containing $\mathfrak{p}$. Pick $a \in R/\mathfrak{p}$ some nonzero element. The ideals

$$(a) \supseteq (a^2) \supseteq (a^3) \supseteq \cdots$$

stabilize, so $a^n = a^{n+1}b$ for some b. Since $R/\mathfrak{p}$ is a domain, $ab = 1$ in A, so a is a unit. Thus, $R/\mathfrak{p}$ is a field, so every prime is maximal. In particular, $\dim(R) = 0$.

Second, note that there are only finitely many maximal ideals. Otherwise, of $\mathfrak{m}_i$ are distinct primes for all $i \geqslant 1$, consider the chain

$$\mathfrak{m}_1 \supseteq \mathfrak{m}_1 \cap \mathfrak{m}_2 \supseteq \mathfrak{m}_1 \cap \mathfrak{m}_2 \cap \mathfrak{m}_3 \supseteq \cdots.$$

This stabilizes, since R is Artinian, so $\mathfrak{m}_{n+1} \supseteq \mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \supseteq \mathfrak{m}_1 \cdots \mathfrak{m}_n$. Since $\mathfrak{m}_{n+1}$ is prime, $\mathfrak{m}_{n+1} \supseteq \mathfrak{m}_i$ for some $i \leqslant n$, and since $\mathfrak{m}_i$ is maximal, we conclude that $\mathfrak{m}_i = \mathfrak{m}_{n+1}$. This contradicts the hypothesis that the $\mathfrak{m}_i$ are all distinct maximal ideals, so we conclude that R has finitely many primes, all maximal. Now, we apply Theorem 6.11 to conclude that R is a finite direct product of local rings of dimension zero. Each of the factors is a quotient of R, and thus each is Artinian. It suffices to show that each factor is Noetherian. So our proof will be complete if we show that any Artinian local ring $(R, \mathfrak{m})$ with only one prime must be a Noetherian ring.

The chain $\mathfrak{m} \supseteq \mathfrak{m}^2 \supseteq \mathfrak{m}^3 \supseteq \cdots$ stabilizes, so that $\mathfrak{m}^n = \mathfrak{m}^{n+1}$. Notice that we cannot apply NAK yet, since we don't know $\mathfrak{m}^n$ is finitely generated. If $\mathfrak{m}^n \neq 0$, consider the family S of ideals $I \subseteq \mathfrak{m}$ such that $I\mathfrak{m}^n \neq 0$. This family contains $\mathfrak{m}$, so in particular it is nonempty, and thus it must have a minimal element since R is Artinian. Take J minimal in S. For some $x \in J$, $x\mathfrak{m}^n \neq 0$, and $(x) \subseteq J \subseteq \mathfrak{m}$, so $J = (x)$ is principal by minimality. Now, $((x)\mathfrak{m}) \cdot \mathfrak{m}^n = (x)\mathfrak{m}^{n+1} = (x)\mathfrak{m}^n \neq 0$, so $(x)\mathfrak{m} \subseteq (x)$ is in S, and by minimality, $(x) = \mathfrak{m}(x)$. Now we can apply NAK 4.30, so $(x) = (0)$, contradicting that $\mathfrak{m}^n \neq 0$. Therefore,

$$0 = \mathfrak{m}^n \subseteq \mathfrak{m}^{n-1} \subseteq \cdots \subseteq \mathfrak{m} \subseteq R.$$

Each of the quotient modules $\mathfrak{m}^i/\mathfrak{m}^{i+1}$ has annihilator $\mathfrak{m}$, so it is a vector space over the field $R/\mathfrak{m}$. Since R is Artinian, so is $\mathfrak{m}^i$ and therefore also $\mathfrak{m}^i/\mathfrak{m}^{i+1}$. While this only shows that $\mathfrak{m}^i/\mathfrak{m}^{i+1}$ is Artinian as an R-module, its $R/\mathfrak{m}$ structure is the same, and thus $\mathfrak{m}^i/\mathfrak{m}^{i+1}$ is Artinian over the field $R/\mathfrak{m}$. As noted in Remark 6.18, this implies that $\mathfrak{m}^i/\mathfrak{m}^{i+1}$ is a finite length $R/\mathfrak{m}$-module, but again that implies it is also finite length over R.

As a consequence, we can stitch together composition series for each quotient and conclude that R has finite length as an R-module; this also follows by Lemma 6.16. Therefore, R is a Noetherian R-module by Lemma 6.21. We conclude that R is also a Noetherian ring. □

Example 6.26. Some Artinian local rings include $k[x,y]/(x^2, y^2)$, $k[x,y]/(x^2, xy, y^2)$, and $\mathbb{Z}/(p^n)$.

Note that $\dim(R) = 0$ does not imply R Artinian unless R is also Noetherian.

Example 6.27. As we saw in Example 6.10, there are rings of dimension 0 that are not Noetherian, and thus also not Artinian. In Example 6.10 we considered the ring $k[x_1, x_2, \ldots]/(x_1^2, x_2^2, \ldots)$, which in fact has only one prime ideal. Note however that

$$(x_1, x_2, \ldots) \supseteq (x_2, x_3, \ldots) \supseteq (x_3, x_4, \ldots) \supseteq \cdots$$

is an infinite descending chain.

Even though every Artinian ring is Noetherian and has finite length, it is not true that Artinian modules are always Noetherian or of finite length.

Example 6.28. Let $R = \mathbb{C}[\![x]\!]$, and $M = R[1/x]/R$. Note that $R[1/x]$ is the ring of Laurent series, so M is the module of "tails" of these functions. This module does not have finite length; it is not even finitely generated! Observe that any submodule N of M either contains $1/x^n$ for all n, or else there is a largest n for which $1/x^n \in N$, and $N = R \cdot 1/x^n$ for this n. The module $R \cdot 1/x^n \subseteq M$ has length n, so it is Artinian, Then every proper submodule of M is Artinian, and thus M itself is Artinian.

Definition 6.29. If $(R, \mathfrak{m}, k)$ is local, a **coefficient field** for R is a subfield $K \subseteq R$ such that the map $K \to R \to R/\mathfrak{m} \cong k$ is an isomorphism.

Rings like $K[\underline{x}]_{(\underline{x})}/I$ have coefficient fields: the copy of K. Some rings without coefficient fields are $\mathbb{Z}_{(p)}$ and $\mathbb{R}[x]_{(x^2+1)}$. Other rings have lots of coefficient fields: $\mathbb{C}[x, y]_{(x)}$ contains $\mathbb{C}(y)$ and $\mathbb{C}(x + y)$, which both are coefficient fields!

Remark 6.30. If $(R, \mathfrak{m}, k)$ is local with coefficient field K, then a finite length R-module M may not be a k-module (it may not be killed by $\mathfrak{m}$), but it is a K-vector space by restriction of scalars, and $\ell(M) = \dim_K(M)$.

6.3 Height and number of generators

Theorem 6.31 (Krull's Principal Ideal theorem)**.** *Let R be a Noetherian ring, and $f \in R$. Then, every minimal prime of (f) has height at most one.*

Note that this is stronger than the statement that the height of (f) is at most one: that would only mean that some minimal prime of (f) has height at most one.

Proof. Suppose the theorem is false, so that there is some ring R, a prime $\mathfrak{p}$, and an element f such that $\mathfrak{p}$ is minimal over (f) and $\operatorname{ht}(\mathfrak{p}) > 1$. If we localize at $\mathfrak{p}$ and then mod out by an appropriate minimal prime, we obtain a Noetherian local domain $(R, \mathfrak{m})$ of dimension at least two in which $\mathfrak{m}$ is the unique minimal prime of (f). Let's work over that Noetherian local domain $(R, \mathfrak{m})$. Note that $\overline{R} = R/(f)$ is zero-dimensional, since $\mathfrak{m}$ is the only minimal prime over (f). Back in R, let $\mathfrak{q}$ be a prime strictly in between (0) and $\mathfrak{m}$, and notice that we necessarily have $f \notin \mathfrak{q}$.

Consider the symbolic powers $\mathfrak{q}^{(n)}$ of $\mathfrak{q}$. We will show that these stabilize in R. Since $\overline{R} = R/(f)$ is Artinian, the descending chain of ideals

$$\mathfrak{q}\overline{R} \supseteq \mathfrak{q}^{(2)}\overline{R} \supseteq \mathfrak{q}^{(3)}\overline{R} \supseteq \cdots$$

stabilizes. We then have some n such that $\mathfrak{q}^{(n)}\overline{R} = \mathfrak{q}^{(m)}\overline{R}$ for all $m \geqslant n$, and in particular, $\mathfrak{q}^{(n)}\overline{R} = \mathfrak{q}^{(n+1)}\overline{R}$. Pulling back to R, we get $\mathfrak{q}^{(n)} \subseteq \mathfrak{q}^{(n+1)} + (f)$. Then any element $a \in \mathfrak{q}^{(n)}$ can be written as $a = b + fr$, where $b \in \mathfrak{q}^{(n+1)} \subseteq \mathfrak{q}^{(n)}$ and $r \in R$. Notice that this implies that $fr \in \mathfrak{q}^{(n)}$. Since $f \notin \mathfrak{q}$, we must have $r \in \mathfrak{q}^{(n)}$. This yields $\mathfrak{q}^{(n)} = \mathfrak{q}^{(n+1)} + f\mathfrak{q}^{(n)}$. Thus, $\mathfrak{q}^{(n)}/\mathfrak{q}^{(n+1)} = f(\mathfrak{q}^{(n)}/\mathfrak{q}^{(n+1)})$, so $\mathfrak{q}^{(n)}/\mathfrak{q}^{(n+1)} = \mathfrak{m}(\mathfrak{q}^{(n)}/\mathfrak{q}^{(n+1)})$. By NAK 4.30, $\mathfrak{q}^{(n)} = \mathfrak{q}^{(n+1)}$ in R. Similarly, we obtain $\mathfrak{q}^{(n)} = \mathfrak{q}^{(m)}$ for all $m \geqslant n$.

Now, if $a \in \mathfrak{q}$ is nonzero, we have $a^n \in \mathfrak{q}^n \subseteq \mathfrak{q}^{(n)} = \mathfrak{q}^{(m)}$ for all m, so

$$\bigcap_{m \geqslant 1} \mathfrak{q}^{(m)} = \bigcap_{m \geqslant n} \mathfrak{q}^{(m)} = \mathfrak{q}^{(n)}.$$

Notice that $\mathfrak{q}^n \neq 0$ because R is a domain, and so $\mathfrak{q}^{(n)} \supseteq \mathfrak{q}^n$ is also nonzero. So

$$\bigcap_{m \geqslant 1} \mathfrak{q}^{(m)} = \mathfrak{q}^{(n)} \neq 0.$$

On the other hand, $\mathfrak{q}^{(m)} = \mathfrak{q}^m R_{\mathfrak{q}} \cap R$ for all m, and

$$\bigcap_{m \geqslant 1} \mathfrak{q}^{(m)} R_{\mathfrak{q}} \subseteq \bigcap_{m \geqslant 1} \mathfrak{q}^m R_{\mathfrak{q}} = \bigcap_{m \geqslant 1} (\mathfrak{q} R_{\mathfrak{q}})^m = 0$$

by the Krull intersection theorem 5.63. Since R is a domain, the contraction of (0) in $R_{\mathfrak{q}}$ back in R is (0). This is the contradiction we seek. So no such $\mathfrak{q}$ exists, so that R has dimension 1, and in the original ring, all the minimal primes over f must have height at most 1. □

We want to generalize this, but it is not so straightforward to run an induction. We will need a lemma that allows us to control the chains of primes we get.

Lemma 6.32. *Let R be Noetherian, $\mathfrak{p} \subsetneq \mathfrak{q} \subsetneq \mathfrak{a}$ be primes, and $f \in \mathfrak{a}$. Then there is some $\mathfrak{q}'$ with $\mathfrak{p} \subsetneq \mathfrak{q}' \subsetneq \mathfrak{a}$ and $f \in \mathfrak{q}'$.*

Proof. If $f \in \mathfrak{p}$, there is nothing to prove, since we can simply take $\mathfrak{q}' = \mathfrak{q}$. Suppose $f \notin \mathfrak{p}$. After we quotient out by $\mathfrak{p}$ and localize at $\mathfrak{a}$, we may assume that $\mathfrak{a}$ is the maximal ideal. We want to find a nonzero prime $\mathfrak{q}' \subsetneq \mathfrak{a}$. Our assumption implies that $f \neq 0$, and then by the principal ideal theorem 6.31, minimal primes of (f) have height one, hence are not $\mathfrak{a}$ nor $\mathfrak{p}$. We can take $\mathfrak{q}'$ to be one of the minimal primes of f. □

Theorem 6.33 (Krull's Height Theorem)**.** *Let R be a Noetherian ring. If I is an ideal generated by n elements, then every minimal prime of I has height at most n.*

Proof. By induction on n. The case $n = 1$ is the Principal Ideal Theorem 6.31.

Let $I = (f_1, \dots, f_n)$ be an ideal, $\mathfrak{p}$ a minimal prime of I, and $\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_h = \mathfrak{p}$ be a saturated chain of length h ending at $\mathfrak{p}$. If $f_1 \in \mathfrak{p}_1$, then we can apply the induction hypothesis to the ring $\overline{R} = R/((f_1) + \mathfrak{p}_0)$ and the ideal $(f_2, \dots, f_n)\overline{R}$. Then by induction hypothesis, the chain $\mathfrak{p}_1\overline{R} \subsetneq \cdots \subsetneq \mathfrak{p}_h\overline{R}$ has length at most $n-1$, so $h - 1 \leqslant n - 1$ and $\mathfrak{p}$ has height at most n.

If $f_1 \notin \mathfrak{p}_1$, we use the previous lemma to replace our given chain with a chain of the same length but such that $f_1 \in \mathfrak{p}_1$. To do this, note that $f_1 \in \mathfrak{p}_i$ for some i; after all, $f_1 \in I \subseteq \mathfrak{p}$. So in the given chain, suppose that $f_1 \in \mathfrak{p}_{i+1}$ but $f_1 \notin \mathfrak{p}_i$. If $i > 0$, apply the previous lemma with $\mathfrak{a} = \mathfrak{p}_{i+1}$, $\mathfrak{q} = \mathfrak{p}_i$, and $\mathfrak{p} = \mathfrak{p}_{i-1}$ to find $\mathfrak{q}_i$ such that $f_1 \in \mathfrak{q}_i$. Replace the chain with

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_{i-1} \subsetneq \mathfrak{q}_i \subsetneq \mathfrak{p}_i \subsetneq \cdots \subsetneq \mathfrak{p}_h = \mathfrak{p}.$$

Repeat until $f_1 \in \mathfrak{p}_1$. □

Example 6.34.

a) The bound is certainly sharp: an ideal generated by n variables $(x_1, x_2, \dots, x_n)$ in a polynomial ring has height n. There are many other such ideals. For example, $(u^3 - xyz, x^2 + 2xz - 6y^5, vx + 7vy) \in k[u, v, w, x, y, z]$. An ideal of height n generated by n elements is called a **complete intersection**.

b) The ideal (xy, xz) in $k[x, y, z]$ has minimal primes of heights 1 and 2.

c) It is possible to have associated primes of height greater than the number of generators. For a cheap example, in $R = k[x, y]/(x^2, xy)$, the ideal generated by zero elements (the zero ideal) has an associated prime of height two, namely (x, y).

d) The same phenomenon can happen even in a nice polynomial ring. For example, consider the ideal $I = (x^3, y^3, x^2u + xyv + y^2w) \subseteq R = k[u, v, w, x, y]$. Note that $(u, v, w, x, y) = (I : x^2y^2)$, so I has an associated prime of height 5.

e) Noetherianity is necessary. Let $R = k[x, xy, xy^2, \dots] \subseteq k[x, y]$. For all $a \geqslant 1$, $xy^a \notin (x)$, since $y^a \notin R$, but $(xy^a)^2 = x \cdot xy^{2a} \in (x)$. Then (x) is not prime in R, and moreover $\mathfrak{m} = (x, xy, xy^2, \dots) \subseteq \sqrt{(x)}$. Since $\mathfrak{m}$ is a maximal ideal, we have equality, so $\operatorname{Min}(x) = \{\mathfrak{m}\}$. However, $\mathfrak{p} = (xy, xy^2, xy^3, \dots) = (y)k[x, y] \cap R$ is prime, and the chain $(0) \subsetneq \mathfrak{p} \subsetneq \mathfrak{m}$ shows that $\operatorname{ht}(\mathfrak{m}) > 1$.

Lemma 6.35. *Let R be a Noetherian ring, and I be an ideal. Let $f_1, \dots, f_t \in I$, and $J_i = (f_1, \dots, f_i)$ for each i. Suppose that for each i,*

$$f_i \notin \bigcup_{\substack{\mathfrak{a} \in \operatorname{Min}(J_{i-1}) \\ \mathfrak{a} \notin V(I)}} \mathfrak{a}.$$

Then any minimal prime of J_i either contains I or has height i.

Proof. We use induction on i. For $i = 0$, $J_0 = (0)$, and every minimal prime has height zero. Suppose know the statement holds for $i = m$, and consider a minimal prime $\mathfrak{q}$ of J_{m+1}. Since $J_m \subseteq J_{m+1}$, $\mathfrak{q}$ must contain some minimal prime of J_m, say $\mathfrak{p}$. If $\mathfrak{p} \supseteq I$, then $\mathfrak{q} \supseteq I$. If $\mathfrak{q}$ does not contain I, then neither does $\mathfrak{p}$. On the one hand, $f_{m+1} \in J_{m+1} \subseteq \mathfrak{q}$. On the other hand, since $\mathfrak{p} \in \operatorname{Min}(J_m)$ and $\mathfrak{p} \notin V(I)$, our assumption implies that $f_{m+1} \notin \mathfrak{p}$. In particular, $\mathfrak{p} \subsetneq \mathfrak{q}$. By the induction hypothesys, $\mathfrak{p}$ has height m, and thus the height of $\mathfrak{q}$ is at least $m+1$. But J_{m+1} is generated by $m+1$ elements, so by the Krull Height Theorem 6.33, the height of $\mathfrak{q}$ is then exactly $m+1$. □

Theorem 6.36. *Let R be a Noetherian ring of dimension d.*

a) If $\mathfrak{p}$ is a prime of height h, then there are h elements $f_1, \dots, f_h \in \mathfrak{p}$ such that $\mathfrak{p}$ is a minimal prime of $(f_1, \dots, f_h)$.

b) If I is any ideal in R, then there are (at most) $d+1$ elements $f_1, \dots, f_{d+1} \in I$ such that $\sqrt{I} = \sqrt{(f_1, \dots, f_{d+1})}$.

c) Suppose that R is either a local ring or an $\mathbb{N}$-graded ring with R_0 a field. Let I is an ideal in R, homogeneous in the graded case. There are d elements, which can be chosen to be homogeneous in the graded case, say $f_1, \dots, f_d \in I$, such that $\sqrt{I} = \sqrt{(f_1, \dots, f_d)}$.

Proof. We will use the notation from the previous lemma.

a) If $\mathfrak{p}$ is a minimal prime in R, then $\mathfrak{p}$ is minimal over the ideal generated by 0 elements, (0). Otherwise, we will use the recipe from the lemma above with $I = \mathfrak{p}$. First, we need to show that we can choose h elements satisfying the hypotheses. So we will show that starting from $J_0 = (0)$, we can find elements $f_1, \dots, f_h \in p$ such that $J_i = (f_1, \dots, f_i)$

$$\mathfrak{p} \not\subseteq \bigcup_{\substack{\mathfrak{a} \in \operatorname{Min}(J_i) \\ \mathfrak{a} \notin V(I)}} \mathfrak{a}$$

for $i = 0, \dots, h-1$. As long as the set on the right is nonempty,

$$(f_1, \dots, f_i) \subseteq \bigcup_{\substack{\mathfrak{a} \in \operatorname{Min}(J_i) \\ \mathfrak{a} \notin V(I)}} \mathfrak{a},$$

so the previous statement allows us to choose f_{i+1} as in the Lemma. So fix any $i \leqslant h-1$, and suppose we have constructed J_i. The Krull Height Theorem 6.33 implies that all the elements in $\operatorname{Min}(J_i)$ have height strictly less than h. Since $\mathfrak{p}$ has height h, that implies that the sets $\operatorname{Min}(J_i)$ and $V(\mathfrak{p})$ are disjoint. So we want to show that

$$\mathfrak{p} \not\subseteq \bigcup_{\substack{\mathfrak{a} \in \operatorname{Min}(f_1, \dots, f_i) \\ \mathfrak{a} \notin V(I)}} \mathfrak{a} = \bigcup_{\mathfrak{a} \in \operatorname{Min}(f_1, \dots, f_i)} \mathfrak{a}$$

This is immediate by prime avoidance 5.36, again because $\mathfrak{p}$ is not contained in a minimal prime of $(f_1, \dots, f_i)$. Thus, we can choose $(f_1, \dots, f_h) \subseteq \mathfrak{p}$ as in the lemma, and by the lemma its minimal primes either have height h or contain $\mathfrak{p}$. Since $(f_1, \dots, f_h) \subseteq \mathfrak{p}$, some minimal prime $\mathfrak{q}$ of J_h is contained in $\mathfrak{p}$. We know that this $\mathfrak{q}$ either contains $\mathfrak{p}$, and hence is $\mathfrak{p}$, or else is contained in and has the same height as $\mathfrak{p}$, so again must be equal to $\mathfrak{p}$. Therefore, $\mathfrak{p}$ is a minimal prime of $(f_1, \dots, f_h)$.

b) Again, we use the recipe from Lemma 6.35. We again need to see that we can do this. Inductively, we will choose elements inside of I, so each J_i is contained in I, and $V(I) \subseteq V(J_i)$. We start with $J_0 = (0)$.

If for some i we have $\operatorname{Min}(J_i) \smallsetminus V(I) = \varnothing$, then each minimal prime of J_i lies in $V(I)$, so $V(J_i) \subseteq V(I)$. Then $V(J_i) = V(I)$, so $\sqrt{J_i} = \sqrt{I}$. If $\operatorname{Min}(J_i) \smallsetminus V(I) \neq \varnothing$, then $I \not\subseteq \mathfrak{q}$ for any $\mathfrak{q} \in \operatorname{Min}(J_i) \smallsetminus V(I)$, and $I \not\subseteq \bigcup_{\operatorname{Min}(J_i) \smallsetminus V(I)} \mathfrak{q}$ by prime avoidance 5.36, so we can choose elements as in the lemma.

If $\sqrt{(f_1, \ldots, f_i)} = \sqrt{I}$ for $i \leqslant d$, we are done. Suppose not. Then we get elements $(f_1, \ldots, f_{d+1}) = J_{d+1} \subseteq I$ such that the minimal primes of J_{d+1} either contain I or have height at least $d+1$. By the assumption that $\dim(R) = d$, no prime has height $d+1$, so all the minimal primes of J_{d+1} must contain I. Since $J_{d+1} \subseteq I$, any minimal prime of J_{d+1} must also be minimal over I. Thus, $\mathrm{Min}(J_{d+1}) \subseteq \mathrm{Min}(I)$, so $V(J_{d+1}) \subseteq V(I)$, and equality holds, so the radicals are equal.

c) We again run the same argument, using homogeneous prime avoidance in the graded case. The point is that the only (homogeneous, in the graded case) ideal of height d already contains I. □

Corollary 6.37. *Let $(R, \mathfrak{m}, k)$ be a Noetherian local ring. Then*

$$\dim(R) = \min\{n \mid \sqrt{(f_1, \ldots, f_n)} = \mathfrak{m} \textit{ for some } f_1, \ldots, f_n\} \leqslant \mu(\mathfrak{m}).$$

In particular, a Noetherian local ring has finite dimension.

Proof. The dimension of a local ring is the height of its maximal ideal. Thus, by Krull's Height Theorem 6.33, the minimum n in the middle is at least $\dim(R)$, and Theorem 6.36 gives the other direction. Since $\mathfrak{m}$ is generated by $\mu(\mathfrak{m})$ elements, there are in particular $\mu(\mathfrak{m})$ elements whose radical is $\mathfrak{m}$. □

Definition 6.38. The **embedding dimension** of a local ring $(R, \mathfrak{m})$ is the minimal number of generators of $\mathfrak{m}$, $\mu(\mathfrak{m})$. We write $\mathrm{embdim}(R) := \mu(\mathfrak{m})$ for the embedding dimension of R.

So Corollary 6.37 can be restated as $\dim(R) \leqslant \mathrm{embdim}(R)$.

Corollary 6.39. *Let k be a field and $R = k[\![x_1, \ldots, x_d]\!]$. Then $\dim(R) = d$.*

Proof. Let $\mathfrak{m} = (x_1, \ldots, x_d)$. The strict chain of primes

$$(0) \subsetneq (x_1) \subsetneq (x_1, x_2) \subsetneq \cdots \subsetneq (x_1, \ldots, x_d)$$

shows that $\dim(R) \leqslant d$. On the other hand, the images of $x_1, \ldots, x_d$ in $\mathfrak{m}/\mathfrak{m}^2$ are linearly independent, so $\mu(\mathfrak{m}) = d$. By Corollary 6.37, $\dim(R) = d$. □

Rings whose dimension and embedding dimension agree are very nicely behaved.

Definition 6.40. A local ring $(R, \mathfrak{m})$ is **regular** if $\dim(R) = \mathrm{embdim}(R)$.

So we just showed that power series rings $k[\![x_1, \ldots, x_d]\!]$ are regular local rings.

In general, a ring is regular if all its localizations are regular local rings. In order for this definition to make sense, we need to first make sure that regularity localizes, meaning that if $(R, \mathfrak{m})$ is a regular local ring, then R_P is also regular for all primes P. But to do that, we need some homological algebra. However (spoiler alert!), things do work out alright, and as you might expect, polynomial rings over fields are also regular.

Chapter 7

Dimension theory II

7.1 Over, up and down

Given a ring homomorphism $R \xrightarrow{\varphi} S$, we want to study the behavior of chains of primes under φ, meaning how chains in R behave under expansion to S or chains in S behave under contraction to R.

First, we need a technical definition.

Definition 7.1. Let $R \xrightarrow{\varphi} S$ be a ring homomorphism, and consider a prime $\mathfrak{p}$ in R. The ring

$$\kappa_\phi(\mathfrak{p}) := (R \smallsetminus \mathfrak{p})^{-1}(S/\mathfrak{p}S)$$

is the *fiber ring* of ϕ over $\mathfrak{p}$. As a special case, we write $\kappa(\mathfrak{p})$ for the fiber of the identity map; this is $R_\mathfrak{p}/\mathfrak{p}R_\mathfrak{p}$, the residue field of the local ring $R_\mathfrak{p}$.

The point of this definition is that the primes ideals in this ring correspond to the primes in S that contract to $\mathfrak{p}$.

Lemma 7.2. *Let $R \xrightarrow{\varphi} S$ be a ring homomorphism, and $\mathfrak{p} \in \operatorname{Spec}(R)$. The set of primes in S that contract to $\mathfrak{p}$ correspond to the primes in $\kappa_\phi(\mathfrak{p})$. More precisely, $\operatorname{Spec}(\kappa_\phi(\mathfrak{p})) \cong (\varphi^*)^{-1}(\mathfrak{p})$.*

Proof. Consider the maps $S \xrightarrow{\pi} S/\mathfrak{p}S \xrightarrow{g} (R \smallsetminus \mathfrak{p})^{-1}(S/\mathfrak{p}S)$. In Example 3.39 we saw that the map on spectra induced by π can be identified with the inclusion of $V(\mathfrak{p}S)$ into $\operatorname{Spec}(S)$. For the second map, g, we saw in Proposition 4.27 that the map on spectra can be identified with the inclusion of the set of primes that do not intersect $R \smallsetminus \mathfrak{p}$, i.e., those whose contraction is contained in $\mathfrak{p}$. Together, these say $(g \circ \pi)^*$ is an inclusion, whose image is the set of primes in S that contract to $\mathfrak{p}$. $\square$

We have seen that taking $IS \cap R$ does not always recover the ideal I. When I is a prime ideal, we can characterize this in terms of the induced map on Spec.

Lemma 7.3 (Image criterion)**.** *Let $R \xrightarrow{\varphi} S$ be a ring homomorphism. For any $\mathfrak{p} \in \operatorname{Spec}(R)$, $\mathfrak{p} \in \operatorname{im}(\varphi^*)$ if and only if $\mathfrak{p}S \cap R = \mathfrak{p}$.*

Proof. If $\mathfrak{p}S \cap R = \mathfrak{p}$, then

$$\frac{R}{\mathfrak{p}} = \frac{R}{\mathfrak{p}S \cap R} \hookrightarrow \frac{S}{\mathfrak{p}S},$$

so localizing at $(R \smallsetminus \mathfrak{p})$, we get an inclusion $\kappa(\mathfrak{p}) \subseteq \kappa_\varphi(\mathfrak{p})$. Since $\kappa(\mathfrak{p})$ is nonzero, so is $\kappa_\varphi(\mathfrak{p})$, and thus its spectrum is nonempty. By Lemma 7.2, there is a prime mapping to $\mathfrak{p}$.

If $\mathfrak{p}S \cap R \neq \mathfrak{p}$, then $\mathfrak{p}S \cap R \supsetneq \mathfrak{p}$. If $\mathfrak{q} \cap R = \mathfrak{p}$, then $\mathfrak{q} \supseteq \mathfrak{p}S$, so $\mathfrak{q} \cap R \supsetneq \mathfrak{p}$. So no prime contracts to $\mathfrak{p}$. □

Note that $\mathfrak{p}S$ may not be prime, in general.

Example 7.4. Let $R = \mathbb{C}[x^n] \subseteq S = \mathbb{C}[x]$. The ideal $(x^n - 1)R$ is prime. On the other hand, if ζ is a primitive nth root of unity, then

$$(x^n - 1)S = \left(\prod_{i=0}^{n-1} x - \zeta^i \right) S,$$

which is not prime. However, each of its minimal primes $(x - \zeta^i)S$ contracts to $(x^n - 1)R$, so $(x^n - 1)S \cap R = (x^n - 1)R$. Similarly, the ideal $x^n R$ is prime, while $x^n S$ is not even radical.

Example 7.5. Consider the inclusion $R := k[xy, xz, yz] \xhookrightarrow{\varphi} S := k[x, y, z]$ and the prime $\mathfrak{p} = (xy)$ in R. Notice that $(xz)(yz) \in \mathfrak{p}S \cap R$, but *not* in $\mathfrak{p}$, so $\mathfrak{p}S \cap R \supsetneq \mathfrak{p}$, and thus $\mathfrak{p} \notin \operatorname{im}(\varphi^*)$. We can check this more directly, by noting that any prime Q in S contracting to $\mathfrak{p}$ would contain $\mathfrak{p}S = (x) \cap (y)$, so $Q \supseteq (x)$ or $Q \supseteq (y)$. But $(x) \cap R = (xy, xz) \supsetneq \mathfrak{p}$ and $(y) \cap R = (xy, yz) \supsetneq \mathfrak{p}$, so no prime in S contracts to $\mathfrak{p}$.

Corollary 7.6. *If $R \subseteq S$ is a direct summand, then* $\operatorname{Spec}(S) \to \operatorname{Spec}(R)$ *is surjective, so Lemma 7.3 says the map on Spec is surjective.*

Proof. By Lemma 2.17, we know $IS \cap R = I$ for all ideals in this case. □

We want to extend the idea of the last corollary to work for all integral extensions.

Definition 7.7. Let R be a ring, S an R-algebra, and I an ideal. An element r of R is **integral** over I if it satisfies an equation of the form

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0 \qquad \text{with } a_i \in I^i \text{ for all } i.$$

An element of S is **integral** over I if

$$s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n = 0 \qquad \text{with } a_i \in I^i \text{ for all } i.$$

The **integral closure** of I in R is the set of elements of R that are integral over I, denoted $\overline{I}$. Similarly, we write $\overline{I}^S$ for the integral closure of I in S.

The convention is that $I^0 = R$ for any ideal I of R.

Remark 7.8. Notice that $\overline{I}^S \cap R = \overline{I}$ is immediate from the definition.

Exercise 20. Let $R \subseteq S$, I be an ideal of S, and t be an indeterminate. Consider the rings $R[It] \subseteq R[t] \subseteq S[t]$. Here $R[It]$ is the subalgebra of $R[t]$ generated by elements of the form at for all $a \in I$. Notice that we can give this a structure of a graded ring by setting all elements in R to have degree 0 and t to have degree 1, so

$$R[It] = \bigoplus_{n \geqslant 0} I^n t^n.$$

This is usually called the **Rees algebra** of I.

a) $\overline{I}^S = \{s \in S \mid st \in S[t] \text{ is integral over the ring } R[It]\}$.

b) $\overline{I}^S$ is an ideal of S.

In older texts and papers (e.g., Atiyah–Macdonald [AM69] and [Kun69]) a different definition is given for integral closure of an ideal. The one we use here is now the more universally used notion.

Lemma 7.9 (Extension–contraction lemma for integral extensions)**.** *Let $R \subseteq S$ be integral, and I be an ideal of R. Then $IS \subseteq \overline{I}^S$, and hence $IS \cap R \subseteq \overline{I}$.*

Proof. Let $x \in IS$. We can write $x = a_1 s_1 + \cdots + a_t s_t$ for some $a_i \in I$. Moreover, taking $S' = R[s_1, \ldots, s_t]$, we also have $x \in IS'$. We will show that $x \in \overline{I}^{S'}$, so $x \in \overline{I}^S$ follows as a corollary. So we might as well replace S with S', so that $R \subseteq S$ is also integral and module-finite. By Corollary 1.37, the extension is also module-finite.

Let $S = Rb_1 + \cdots + Rb_n$. We can write

$$xb_i = \left(\sum_{k=1}^{t} a_k s_k \right) b_i = \sum_j a_{ij} b_j$$

with $a_{ij} \in I$. We can write these equations in the form $xv = Av$, where $v = (b_1, \ldots, b_u)$, and $A = [a_{ij}]$. By the determinantal trick, Lemma 1.35, we have $\det(xI - A)v = 0$. Since we can assume $b_1 = 1$, we have $\det(xI - A) = 0$. The fact that this is the type of equation we want follows from the monomial expansion of the determinant: any monomial is a product of n terms where some of them are copies of x, and the rest are elements of I. Since this is a product of n terms, a term in x^i has a coefficient coming from a product of $n - i$ elements of I.

So this shows that $IS \subseteq \overline{I}^S$. Now notice that $\overline{I}^S \cap R = \overline{I}$ is immediate from the definition, as noted in Remark 7.8. □

Theorem 7.10 (Lying over)**.** *If $R \subseteq S$ is an integral extension, then $\mathfrak{p}S \cap R = \mathfrak{p}$ for every $\mathfrak{p} \in \operatorname{Spec}(R)$, so the induced map $\operatorname{Spec}(S) \longrightarrow \operatorname{Spec}(R)$ is surjective.*

Proof. We claim that $\overline{I} \subseteq \sqrt{I}$. Indeed, if $r \in \overline{I}$, then

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$

for some n and some $a_i \in I^i$ for all i, so

$$r^n = -a_1 r^{n-1} - \cdots - a_{n-1} r - a_n \in I.$$

Therefore, if $\mathfrak{p}$ is a prime in R, by Lemma 7.9 we have $\mathfrak{p}S \cap R \subseteq \overline{\mathfrak{p}}$, and

$$\mathfrak{p}S \cap R \subseteq \overline{\mathfrak{p}} \subseteq \sqrt{\mathfrak{p}} = \mathfrak{p}.$$

Then $\mathfrak{p}S \cap R = \mathfrak{p}$, and by Lemma 7.3 we conclude that $\mathfrak{p}$ is in the image of the map on Spec. □

Example 7.11. We saw in Example 7.5 that the map induced on Spec by the inclusion $k[xy, xz, yz] \subseteq k[x, y, z]$ is not surjective. So Theorem 7.10 does not apply — indeed, this inclusion is not module-finite, and thus it is not integral. For example, the infinite set $\{1, x^n, y^n, z^n \mid n \geqslant 1\}$ is a minimal generating set for $k[x, y, z]$ over $k[xy, xz, yz]$

Both assumptions that the extension is integral and that it is an inclusion are needed in Theorem 7.10.

Example 7.12.

a) Suppose f is a regular element on R, but not a unit. Since f is regular, the map $R \longrightarrow R_f$ is an inclusion, but we claim it is not integral. If $\frac{1}{f}$ was integral over R, there would be $a_i \in R$ such that
$$\frac{1}{f^n} + \frac{a_{n-1}}{f^{n-1}} + \cdots + \frac{a_1}{f} + a_0 = 0.$$
After multiplying by f^n all terms are of the form $\frac{r}{1}$, and thus in R, since the localization map is injective. So
$$1 = a_{n-1}f + \cdots + a_1 f^{n-1} + a_0 f^n \in (f),$$
and f must be a unit.
So $R \longrightarrow R_f$ is an example of an inclusion that is not integral. Note that the image of the map on Spec is the complement of $V(f)$, so in particular the map is not surjective.

b) In contrast, the map $R \longrightarrow R/(f)$ is integral, but it is not an inclusion. The map on Spec is again not surjective: its image is $V(f)$.

Remark 7.13. Let I be an ideal in S. Suppose $R \longrightarrow S$ is an integral extension. There is an induced map $R/(I \cap R) \longrightarrow S/I$, and that map is integral: an equation of integral dependance for $s \in S$ over R give an equation for integral dependance of its class in S/I over $R/(I \cap R)$.

Lemma 7.14. *If $R \xrightarrow{\varphi} S$ is integral, $Q \cap R$ is maximal if and only if Q is maximal in S. If $R \subseteq S$ is an integral extension of domains, R is a field if and only S is a field.*

Proof. By Remark 7.13, the induced map $R/(Q \cap R) \subseteq S/Q$ is an integral extension of domains, and Q (respectively, $Q \cap R$) is maximal if and only if S/Q (respectively, $R/(Q \cap R)$) is a field. So it is sufficient to show the second statement, about inclusions of domains.

Now suppose $R \subseteq S$ is an integral extension of domains. Assume R is a field, and take any nonzero $s \in S$. Consider some equation of integral dependance of s over R, say

$$s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0.$$

Since a_0 is a unit in $R \subseteq S$, we can divide by a_0, so that

$$-s(s^{n-1} + a_{n-1}s^{n-2} + \cdots + a_1) = 1.$$

The s is a unit, and S is a field.

If S is a field, and $r \in R$ is nonzero, then there exists an inverse s for r in S, which is integral over R. Then

$$s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$$

for some $a_i \in R$, and multiplying through by r^{n-1} gives

$$s = -(a_{n-1} + a_{n-2}r \cdots + a_1 r^{n-2} + a_0 r^{n-1}) \in R.$$

Then R is a field. $\square$

Theorem 7.15 (Incomparability). *If $R \longrightarrow S$ is integral and $P \subseteq Q$ are such that $P \cap R = Q \cap R$, then $P = Q$.*

Proof. Since the map on spectra induced by $R \longrightarrow R/\ker(R)$ is injective, we can replace R by the quotient and assume φ is an integral inclusion.

So suppose $R \subseteq S$ is integral, and let $\mathfrak{p} = P \cap R = Q \cap R$. We claim that localizing at $(R \smallsetminus P)$ preserves integrality: if $x \in S$ and $w \in R \smallsetminus \mathfrak{p}$, then we have equations of the form

$$x^n + r_1 x^{n-1} + \cdots + r_n = 0 \implies \left(\frac{x}{w}\right)^n + \frac{r_1}{w}\left(\frac{x}{w}\right)^{n-1} + \cdots + \frac{r_n}{w^n} = 0.$$

By localizing R at $(R \smallsetminus \mathfrak{p})$, the image of $\mathfrak{p}$ is a maximal ideal. So we reduced to the situation where $R \cap P = R \cap Q$ is a maximal ideal. By Lemma 7.14, $P \subseteq Q$ are both maximal ideals. Therefore, $P = Q$. $\square$

Corollary 7.16. *Suppose $R \longrightarrow S$ is integral and that S is Noetherian. If S is Noetherian, then only finitely many primes contract to each $\mathfrak{p} \in \operatorname{Spec}(R)$.*

Proof. If $P \in \operatorname{Spec}(S)$ contracts to $\mathfrak{p}$, then $P \supseteq \mathfrak{p}S$, so in particular P contains some prime Q minimal over $\mathfrak{p}S$. Then

$$\mathfrak{p}S \subseteq Q \subseteq P \implies \mathfrak{p} \subseteq Q \cap R \subseteq P \cap R = \mathfrak{p},$$

so $Q \cap R = P \cap R$. By Theorem 7.15, $Q = P$. So all the primes contracting to $\mathfrak{p}$ are in $\operatorname{Min}(\mathfrak{p}S)$, which is a finite set since R is Noetherian. $\square$

Corollary 7.17. *If $R \longrightarrow S$ is integral, then $\operatorname{ht}(\mathfrak{q}) \leqslant \operatorname{ht}(\mathfrak{q} \cap R)$ for any $\mathfrak{q} \in \operatorname{Spec}(S)$. In particular, $\dim(S) \leqslant \dim(R)$.*

Proof. Given a chain of primes $\mathfrak{a}_0 \subsetneq \cdots \subsetneq \mathfrak{a}_n = \mathfrak{q}$ in $\operatorname{Spec}(S)$, we can contract to R, and by Theorem 7.15 we get a chain of distinct primes in $\operatorname{Spec}(R)$. $\square$

Theorem 7.18 (Going up). *If $R \longrightarrow S$ is integral, then for every $\mathfrak{p} \subsetneq \mathfrak{q}$ in $\operatorname{Spec}(R)$ and $P \in \operatorname{Spec}(S)$ with $P \cap R = \mathfrak{p}$, there is some $Q \in \operatorname{Spec}(S)$ with $P \subsetneq Q$ and $Q \cap R = \mathfrak{q}$.*

The picture looks something like this:

$$\begin{array}{ccccccc} P & & & & P & \subseteq & Q \\ \rotatebox{90}{$\subseteq$} & & & \overset{\exists Q}{\rightsquigarrow} & \rotatebox{90}{$\subseteq$} & & \rotatebox{90}{$\subseteq$} \\ \mathfrak{p} & \subseteq & \mathfrak{q} & & \mathfrak{p} & \subseteq & \mathfrak{q} \end{array}$$

Proof. Consider the map $R/\mathfrak{p} \longrightarrow S/\mathfrak{p}S \longrightarrow S/P$. This is integral, as we observed in Remark 7.13. It is also injective, so Lying Over, Theorem 7.10, applies. Thus, there is a prime $\mathfrak{a}$ of S/P that contracts to the prime $\mathfrak{q}/\mathfrak{p}$ in $\mathrm{Spec}(R/\mathfrak{p})$. We can write $\mathfrak{a} = Q/P$ for some $Q \in \mathrm{Spec}(S)$, and we must have that Q contracts to $\mathfrak{q}$. □

Corollary 7.19. *If $R \subseteq S$ is integral, then $\dim(R) = \dim(S)$.*

Proof. We have already shown that $\dim(S) \leqslant \dim(R)$ in Corollary 7.17, so we just need to show that $\dim(R) \leqslant \dim(S)$. Fix a chain of primes $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_n$ in $\mathrm{Spec}(R)$. By Lying Over, Theorem 7.10, there is a prime $\mathfrak{q}_0 \in \mathrm{Spec}(S)$ contracting to $\mathfrak{p}_0$. Then by Going up, Theorem 7.18, we have $\mathfrak{q}_0 \subsetneq \mathfrak{q}_1$ with $\mathfrak{q}_1 \cap R = \mathfrak{p}_1$. Continuing, we can build a chain of distinct primes in S of length n. So $\dim(R) \leqslant \dim(S)$, and equality follows. □

Recall that a domain is normal if it is integrally closed in its field of fractions. In a previous problem set, you showed that $\mathbb{Z}$ is normal; we now extend that result to any unique factorization domain.

Lemma 7.20. *A unique factorization domain is normal. In particular, a polynomial ring over a field is normal.*

Proof. Let R be a UFD, and $\frac{r}{s} \in \mathrm{frac}(R)$ be integral over R. We can assume that r and s have no common factor. Then we have some $a_i \in R$ such that

$$\frac{r^n}{s^n} + a_1 \frac{r^{n-1}}{s^{n-1}} + \cdots + a_n = 0 \qquad \Longrightarrow \qquad r^n = -(a_1 r^{n-1} s + \cdots + a_n s^n).$$

Any irreducible factor of s must then divide r^n, and hence divide r. If s is not a unit in R, then this contradicts that there is no common factor. Therefore, $r/s \in R$. □

Lemma 7.21. *Let R be a normal domain, x be an element integral over R in some larger domain. Let k be the fraction field of R, and $f(t) \in k[t]$ be the minimal polynomial of x over k.*

a) If x is integral over R, then $f(t) \in R[t] \subseteq k[t]$.

b) If x is integral over a prime $\mathfrak{p}$, then $f(t)$ has all of its nonleading coefficients in $\mathfrak{p}$.

Proof. Let x be integral over R. Fix an algebraic closure of k containing x, and let $x_1 = x, x_2, \ldots, x_u$ be the roots of f. Since $f(t)$ divides any polynomial with coefficients in k that x satisfies, it also divides a monic equation of integral dependance for x over R. Therefore, each x_i is a solution to such an equation of integral dependence, and thus must be integral over R.

Let $S = R[x_1, \ldots, x_u] \subseteq \overline{k}$. This is a module-finite extension of R, so all of its elements are integral over R. The leading coefficient of $f(t)$ is 1, and the remaining coefficients of $f(t)$

are polynomials in the x_i, hence they lie in S. On the other hand, R is normal, so $S \cap k = R$. We conclude that all the coefficients of f are in R, and $f \in R[t]$.

Now let x be integral over $\mathfrak{p}$. By the same argument as above, all of the x_i are integral over $\mathfrak{p}$. Since each $x_i \in \overline{\mathfrak{p}}^S$, any polynomial in the x_i lies in $\overline{\mathfrak{p}}^S$. So the nonleading coefficients of f lie in $\overline{\mathfrak{p}}^S \cap R = \mathfrak{p}$, by Theorem 7.10. $\square$

Theorem 7.22 (Going down)**.** *Suppose that R is a normal domain, S is a domain, and $R \subseteq S$ is integral. Then, for every $\mathfrak{p} \subsetneq \mathfrak{q}$ in* $\operatorname{Spec}(R)$ *and Q in* $\operatorname{Spec}(S)$ *with $Q \cap R = \mathfrak{q}$, there is some $P \in$* $\operatorname{Spec}(S)$ *with $P \subsetneq Q$ and $P \cap R = \mathfrak{p}$.*

The picture looks like

$$\begin{array}{ccc} & Q & & P \subseteq Q \\ & \cup\!\! | & \overset{\exists P}{\rightsquigarrow} & \cup\!\! | \quad \cup\!\! | \\ \mathfrak{p} \subseteq & \mathfrak{q} & & \mathfrak{p} \subseteq \mathfrak{q} \end{array}$$

Proof. As before, we can replace R and S by their localizations at the multiplicatively closed set $R \setminus \mathfrak{q}$ without loss of generality, since that extension is still integral. So now $\mathfrak{q}$ is the unique maximal ideal in R, and want to show that $\mathfrak{p}$ is the contraction of some prime ideal $P \subseteq Q$, so it suffices to find some prime ideal in S_Q. Se we can further compose with the localization of S at Q, and as before $R \longrightarrow S_Q$ is still an integrally closed extension. We have thus reduced to the case when $(R, \mathfrak{q})$ and (S, Q) are local. By Lemma 7.3, it suffices to show that $\mathfrak{p}S \cap R = \mathfrak{p}$.

Let $r \in \mathfrak{p}S \cap R$. Then $r = s_1a_1 + \cdots, +s_na_n$ for some $s_i \in S$ and $a_i \in \mathfrak{p}$, so $r \in R[s_1, \ldots, s_n]$.

Let $W = (S \smallsetminus Q)(R \smallsetminus \mathfrak{p})$ be the multiplicative set in S consisting of products of elements in $S \smallsetminus Q$ and $R \smallsetminus \mathfrak{p}$. Note that each of these sets contains 1, so each set is contained in W, the product of the two. We will show that $W \cap \mathfrak{p}S$ is empty. Once we do that, it will follow from Lemma 3.41 that there is a prime ideal P in S containing $\mathfrak{p}S$ such that $W \cap P$ is empty. Notice that such a prime is necessarily contained in Q, since $S \smallsetminus Q \subseteq W$. Moreover, $R \setminus \mathfrak{p} \subseteq W$, so $(Q \cap R) \cap (R \setminus \mathfrak{p})$ is empty, or equivalently, $Q \cap R \subseteq \mathfrak{p}$. We conclude that $Q \cap R = \mathfrak{p}$.

So our goal is to show that $W \cap \mathfrak{p}S$ is empty. We proceed by contradiction, and assume there is some $x \in \mathfrak{p}S \cap W$. We can write $x = rs$ for some $r \in R \smallsetminus \mathfrak{p}$ and $s \in S \smallsetminus Q$. Moreover, since $x \in \mathfrak{p}S$, x is integral over $\mathfrak{p}$, by Lemma 7.9.

Consider the minimal polynomial of x over $\operatorname{frac}(R)$, say

$$h(x) = x^n + a_1x^{n-1} + \cdots + a_n = 0.$$

By Lemma 7.21, each $a_i \in \mathfrak{q} \subseteq R$. Then substituting $x = rs$ in $\operatorname{frac}(R)$ and dividing by r^n yields

$$g(s) = s^n + \frac{a_1}{r}s^{n-1} + \cdots + \frac{a_n}{r^n} = 0.$$

We claim that this is the minimal polynomial of s. If s satisfied a monic polynomial of degree $d < n$, multiplying by r^d would give us a polynomial of degree d that x satisfies, which is impossible. So indeed, this is the minimal polynomial of s.

Since $s \in S$, and thus integral over R, Lemma 7.21 says that each $\frac{a_i}{r^i} =: v_i \in R$. Since $r \notin \mathfrak{p}$ and $r^iv_i = a_i \in \mathfrak{p}$, we must have $v_i \in \mathfrak{p}$. The equation $g(s) = 0$ then shows that

$s \in \sqrt{\mathfrak{p}S}$. Since $Q \in \operatorname{Spec}(S)$ contains $\mathfrak{q}S$ and hence $\mathfrak{p}S$, we have $s \in \sqrt{\mathfrak{p}S} \subseteq Q$. This is the desired contradiction. $\square$

Corollary 7.23. *If R is a normal domain, S is a domain, and $R \subseteq S$ is integral, then* $\operatorname{ht}(\mathfrak{q}) = \operatorname{ht}(\mathfrak{q} \cap R)$ *for any* $\mathfrak{q} \in \operatorname{Spec}(S)$.

Proof. We already know from Corollary 7.17 that $\operatorname{ht}(\mathfrak{q}) \leqslant \operatorname{ht}(\mathfrak{q} \cap R)$. Given a saturated chain up to $\mathfrak{q} \cap R$, we can apply Going Down, Theorem 7.22 to get a chain just as long that goes up to $\mathfrak{q}$. $\square$

7.2 Noether normalization and dimension of affine rings

Lemma 7.24 (Making a pure-power leading term)**.**

a) *Let A be a domain, and $f \in R = A[x_1, \dots, x_n]$ be a (not necessarily homogeneous) polynomial of degree at most N. The A-algebra automorphism of R given by $\phi(x_i) = x_i + x_n^{N^{n-i}}$ for $i < n$ and $\phi(x_n) = x_n$ maps f to a polynomial that, viewed as a polynomial in x_n with coefficients in $A[x_1, \dots, x_{n-1}]$, has leading term dx_n^a for some $d \in A$ and $a \in \mathbb{N}$.*

b) *Let k be an infinite field, and let $R = k[x_1, \dots, x_n]$ be standard graded, meaning* $\deg(x_i) = 1$*. Let $f \in R$ be a homogeneous polynomial of degree N. There is a degree-preserving k-algebra automorphism of R given by $\phi(x_i) = x_i + a_i x_n$ for $i < n$ and $\phi(x_n) = x_n$ that maps f to a polynomial that viewed as a polynomial in x_n with coefficients in $k[x_1, \dots, x_{n-1}]$, has leading term ax_n^N for some (nonzero) $a \in k$.*

Proof.

a) The map ϕ sends a monomial term $dx_1^{a_1} \cdots x_n^{a_n}$ to a polynomial with unique highest degree term $dx_n^{a_1 N^{n-1} + a_2 N^{n-2} + \cdots + a_{n-1} N + a_n}$. For each of the monomials $dx_1^{a_1} \cdots x_n^{a_n}$ in f with nonzero coefficient $d \neq 0$, we must have each $a_i \leqslant N$, so the map $(a_1, \dots, a_n) \mapsto a_1 N^{n-1} + a_2 N^{n-2} + \cdots + a_{n-1} N + a_n$ is injective when restricted to the set of exponent tuples of f. Therefore, none of the terms can cancel. We find that the leading term is of the promised form.

b) We just need to show that the x^N coefficient of $\phi(f)$ is nonzero for some choice of a_i. One can check that the coefficient of the x^N term is $f(-a_1, \dots, -a_{n-1}, 1)$. But $f(-a_1, \dots, -a_{n-1}, 1)$, when thought of as a polynomial in the a_i, is identically zero, then f must be the zero polynomial. $\square$

Theorem 7.25 (Noether Normalization)**.** *Let A be a domain, and R be a finitely generated A-algebra. There is some nonzero $a \in A$ and $x_1, \dots, x_t \in R$ algebraically independent over A such that R_a is module-finite over $A_a[x_1, \dots, x_t]$. In particular, if $A = k$ is a field, then R is module-finite over $k[x_1, \dots, x_t]$.*

Proof. We proceed by induction on the number of generators n of R over A. There is nothing to prove in the case when $n = 0$.

Now suppose that we know the result holds for A-algebras generated by at most $n - 1$ elements, and let $R = A[r_1, \dots, r_n]$. If $r_1, \dots, r_n$ are algebraically independent over A, we are done. If not, there is some $f(x_1, \dots, x_n) \in A[x_1, \dots, x_n]$ such that $f(r_1, \dots, r_n) = 0$. After possibly applying Lemma 7.24 to change our choice of algebra generators, we can assume that f has leading term ax_n^N for some a. Then f is monic in x_n after inverting a, so R_a is module-finite over $A_a[r_1, \dots, r_{n-1}]$. By hypothesis, $A_{ab}[r_1, \dots, r_{n-1}]$ is module-finite over $A_{ab}[x_1, \dots, x_s]$ for some $b \in A$ and $x_1, \dots, x_s$ that are algebraically independent over A. Since R_{ab} is module-finite over $A_{ab}[r_1, \dots, r_{n-1}]$, R_{ab} must also be module-finite over $A_{ab}[x_1, \dots, x_s]$, and we are done. □

Theorem 7.26 (Graded Noether Normalization). *Let k be an infinite field, and R be a finitely generated $\mathbb{N}$-graded k-algebra with $R_0 = k$ and $R = k[R_1]$. There are homogeneous elements $x_1, \dots, x_t \in R_1$ algebraically independent over k such that R is module-finite over $k[x_1, \dots, x_t]$.*

Proof. We repeat the proof of Theorem 7.25 but use Lemma 7.24 (2), the graded version. □

Remark 7.27. There also exist Noether normalizations for quotients of power series rings over fields: after a change of coordinates, one can rewrite any nonzero power series in $k[\![x_1, \dots, x_n]\!]$ as a series of the form $u(x_n^d + a_{d-1}x_n^{d-1} + \cdots + a_0)$ for a unit u and $a_0, \dots, a_{d-1} \in k[\![x_1, \dots, x_{n-1}]\!]$. This is called *Weierstrass preparation*. The proof of the Noether normalization theorem proceeds in essentially the same way. Thus, given $k[\![x_1, \dots, x_n]\!]/I$, we have some module-finite inclusion of another power series ring $k[\![z_1, \dots, z_d]\!] \subseteq k[\![x_1, \dots, x_n]\!]/I$.

Theorem 7.28. *Let R be a domain that is a finitely generated algebra over a field k, or a quotient of a power series ring over a field. Let $k[z_1, \dots, z_d]$ be any Noether normalization for R. For any maximal ideal $\mathfrak{m}$ of R, the length of any saturated chain of primes from 0 to $\mathfrak{m}$ is d. In particular, $\dim(R) = d$.*

Proof. We will show the proof in the case when R is a finitely generated domain over a field k; the power series case is similar, and left as an exercise. We prove by induction on d that for any finitely generated domain with a Noether normalization with d algebraically independent elements, any saturated chain of primes ending in a maximal ideal has length d.

When $d = 0$, R is a domain that is integral over a field, hence R is a field by Lemma 7.14. So suppose the statement holds for $d-1$, and let R be a finitely generated domain over some field k with Noether normalization $k[z_1, \dots, z_d]$. Consider a maximal ideal $\mathfrak{m}$ of R, and a saturated chain

$$0 \subsetneq \mathfrak{q}_1 \subsetneq \cdots \subsetneq \mathfrak{q}_k = \mathfrak{m}.$$

Consider the contraction of this chain to $A = k[z_1, \dots, z_d]$, which by Theorem 7.15 are distinct primes in R:

$$0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_k.$$

Our assumption that the original chain is saturated implies that $\mathfrak{q}_1$ has height 1. If $\mathfrak{p}_1$ had height 2 or more, then by Going Down, Theorem 7.22, so would $\mathfrak{q}_1$, so $\mathfrak{p}_1$ has height 1 as well. Since $k[z_1, \dots, z_d]$ is a UFD, $\mathfrak{p}_1 = (f)$ for some prime element f, by Example 6.3 d. After a change of variables, as in Lemma 7.24, we can assume that f is monic in z_d with coefficients in $k[z_1, \dots, z_{d-1}]$. So $k[z_1, \dots, z_{d-1}] \subseteq A/(f) \subseteq R/\mathfrak{q}_1$ are module-finite, and the induction hypothesis applies to $R/\mathfrak{q}_1$. Now

$$0 = \mathfrak{q}_1/\mathfrak{q}_1 \subsetneq \mathfrak{q}_2/\mathfrak{q}_1 \subsetneq \cdots \subsetneq \mathfrak{q}_k/\mathfrak{q}_1 = \mathfrak{m}/\mathfrak{q}_1$$

is a saturated chain in the affine domain $R/\mathfrak{q}_1$ going up to the maximal ideal $\mathfrak{m}/\mathfrak{q}_1$. The induction hypothesis then says that this chain has length $d-1$, so $k-1 = d-1$, and $k = d$. □

Corollary 7.29. *The dimension of the polynomial ring $k[x_1, \dots, x_d]$ is d.*

Proof. The polynomial ring $k[x_1, \dots, x_d]$ is a Noether normalization of itself, and Theorem 7.28 says that it must have dimension d. □

This matches our geometric intuition: $k[x_1, \dots, x_d]$ corresponds to $\mathbb{A}^d_k$, and we are used to thinking of $\mathbb{A}^d_k$ as a d-dimensional space. Moreover, if R is a finitely generated k-algebra, then R is a quotient of $k[x_1, \dots, x_d]$, where d is the number of generators of R as a k-algebra. Therefore, $\dim(R) \leqslant d$.

Corollary 7.30. *If R is a k-algebra, the dimension of R is less than or equal to the minimal size of an algebra generating set for R over k. If $R = k[f_1, \dots, f_d]$ and $\dim(R) = d$, then R is isomorphic to a polynomial ring over k, and the generators f_i are algebraically independent.*

Proof. The first statement is trivial unless R is finitely generated, in which case we can write $R = k[f_1, \dots, f_s] \cong k[x_1, \dots, x_s]/I$ for some ideal I, so

$$\dim(R) \leqslant \dim(k[x_1, \dots, x_s]) = d.$$

Suppose we chose s to be minimal. If $I \neq 0$, then $\dim(R) < s$, since the zero ideal is not contained in I. □

Corollary 7.31. *Let R be a finitely generated algebra or a quotient of a power series ring over a field.*

1) R is catenary.

If additionally R is a domain, then

2) R is equidimensional, and

2) $\operatorname{ht}(I) = \dim(R) - \dim(R/I)$ for all ideals I.

Proof.

1) Let $\mathfrak{p} \subseteq \mathfrak{q}$ be primes in R. We can quotient out by $\mathfrak{p}$, and assume that R is a domain and $\mathfrak{p} = 0$. Fix a saturated chain C from $\mathfrak{q}$ to a maximal ideal $\mathfrak{m}$. Given two saturated chains C', C'' from 0 to $\mathfrak{q}$, the concatenations $C'|C$ and $C''|C$ are saturated chains from 0 to $\mathfrak{m}$, so by Theorem 7.28 they must have the same length. It follows that C' and C'' have the same length.

2) Equidimensionality is immediate from Theorem 7.28.

3) We have

$$\mathrm{ht}(I) = \min\{\mathrm{ht}(\mathfrak{p}) \mid \mathfrak{p} \in \mathrm{Min}(I)\}$$

and

$$\dim(R/I) = \max\{\dim(R/\mathfrak{p}) \mid \mathfrak{p} \in \mathrm{Min}(I)\}.$$

Therefore, it suffices to show the equality for prime ideals, since if $\mathfrak{p} \in \mathrm{Min}(I)$ attains the minimal $\mathrm{ht}(\mathfrak{p})$, then it also attains the maximal $\dim(R/\mathfrak{p})$. Now, take a saturated chain of primes C from 0 to $\mathfrak{p}$, and a saturated chain C' from $\mathfrak{p}$ to a maximal ideal $\mathfrak{m}$. Since R is catenary, C has length $\mathrm{ht}(\mathfrak{p})$. Moreover, C' has length $\dim(R/\mathfrak{p})$ by Theorem 7.28, and $C|C'$ has length $\dim(R)$ by Theorem 7.28.

□

Example 7.32. Let's use our dimension theorems to give a few different proofs that $R = k[x, y, z]/(y^2 - xz)$ has dimension 2 for any field k.

1) $k[x, z]$ is a Noether normalization for R, so the dimension is 2.
2) We observe that $y^2 - xz$ is irreducible, e.g., by thinking of it as a polynomial in y and applying Eisenstein's criterion. Then $(y^2 - xz)$ is a prime of height one, so the dimension of R is $\dim(k[x, y, z]) - \mathrm{ht}((y^2 - xz)) = 3 - 1 = 2$.

Chapter 8

Hilbert functions

8.1 Hilbert functions of graded rings

We now introduce a useful combinatorial book keeping tool for the vector space dimensions of the graded components of a finitely generated k-algebra.

Definition 8.1. Let k be a field. If R is an $\mathbb{N}$-graded k-algebra, the **Hilbert function**[1] of R is the function $H_R \colon \mathbb{Z} \longrightarrow \mathbb{N} \cup \infty$ defined by

$$H_R(t) := \dim_k(R_t)$$

If M is a $\mathbb{Z}$-graded R-module, the Hilbert function of M is the function $H_R \colon \mathbb{Z} \longrightarrow \mathbb{N} \cup \infty$ defined by

$$H_M(t) := \dim_k(M_t).$$

We may write $H_R^k(t)$ or $H_M^k(t)$ if we want to emphasize what field k we are considering.

Sometimes it's useful to collect the values of the Hilbert function in the form of a power series.

Definition 8.2. If R is $\mathbb{Z}$-graded or $\mathbb{N}$-graded we define the **Hilbert series** of R or of a graded R-module M by $h_R(z) = \sum_{i \in \mathbb{Z}} H_R(i) z^i$ and $h_M(z) = \sum_{i \in \mathbb{Z}} H_M(i) z^i$.

Example 8.3. Consider the standard graded ring

$$R = k[x,y]/(x^2,y^3) = \underbrace{k}_{R_0} \bigoplus \underbrace{(k\,x \oplus k\,y)}_{R_1} \bigoplus \underbrace{(k\,xy \oplus k\,y^2)}_{R_2} \bigoplus \underbrace{k\,xy^2}_{R_3}.$$

Then $H_R(t) = \begin{cases} 1 & \text{if } t = 0 \\ 2 & \text{if } t = 1,2 \\ 1 & \text{if } t = 3 \\ 0 & \text{if } t \geqslant 4 \end{cases}$ and $h_R(z) = 1 + 2z + 2z^2 + z^3$.

Notice that in this example $H_R(t)$ is eventually the zero function, which we will take by convention to have degree -1 as a polynomial. Note also that R is a finite dimensional k-algebra, hence Artinian. So $\dim(R) = 0$.

[1] Some authors call the Hilbert series the Poicaré series, but in modern terminology that means something else.

The key example of a Hilbert function is what happens in the case of a polynomial ring.

Example 8.4. Let k be a field, and $R = k[x_1, \ldots, x_d]$ be a polynomial ring with the standard grading, meaning $\deg x_i = 1$ for each i. To compute the Hilbert function of R, we need to compute the size of a k-basis for $H_R(t)$ for each t. Such a basis is given by all the monomials in $x_1, \ldots, x_d$ of degree t:

$$R_t = \bigoplus_{a_1+\cdots+a_d=t} k \cdot x_1^{a_1} \cdots x_d^{a_d}.$$

We can easily count the number of monomials of degree t using elementary combinatorics, and we find that

$$H_R(t) = \binom{t+d-1}{d-1} = \binom{t+d-1}{t} \quad \text{for } t \geqslant 0.$$

We claim that the binomial function here can be expressed as a polynomial in t for $t \geqslant 0$. Consider

$$P_d(t) = \frac{(t+d-1)(t+d-2)\cdots(t+1)}{(d-1)!} \in \mathbb{Q}[t].$$

Observe that $P_n(t)$ has $-1, -2, \ldots, -(d-1)$ as roots. Then

$$H_R(t) = \begin{cases} P_d(t) & \text{if } t > -d \\ 0 & \text{if } t < 0. \end{cases}$$

Note that the two cases overlap for $-(d-1) \leqslant t \leqslant -1$.

Notice that in this example the Hilbert function is eventually (for $t \geqslant -d$) equal to a polynomial of degree $d-1$. Moreover, recall that $\dim(R) = d$.

To compute the Hilbert series, notice that the number of monomials of degree t is equal to the number of ordered tuples $(a_1, \ldots, a_d)$ with $a_1 + \cdots + a_d = t$. This is the coefficient of z^t in the product

$$(1 + z + z^2 + \cdots + z^{a_1} + \cdots)(1 + z + z^2 + \cdots + z^{a_2} + \cdots)\cdots(1 + z + z^2 + \cdots + z^{a_d} + \cdots)$$

hence

$$h_R(z) = (1 + z + z^2 + \cdots + z^i + \cdots)^d = \frac{1}{(1-z)^n}.$$

While the Hilbert function is a polynomial for any $n \in \mathbb{N}$ in this example, this is not always the case. Here's a cheap example:

Example 8.5. Let k be a field, and $R = k[x_1, \ldots, x_n]$ a polynomial ring with the standard grading $|x_i| = 1$ for each i as in the previous example and let d be an integer. Then

$$H_{R(-d)}(t) = \dim_k(R(-d)_t) = \dim_k(R(-d)_{t-d}) = H_R(t-d)$$

and $h_{R(-d)}(z) = z^d h_R(z)$. In particular, we see that $H_{R(-d)(t)} = P_n(t-d)$ for $t - d > -n$, can be expressed as a polynomial in t when $t - d > -n$, so for $t > d - n$. This Hilbert function is no longer a polynomial for all nonnegative integers, but it is a polynomial for high enough values of t.

To compute more sophisticated examples, we use short exact sequences. Unsurprisingly, Hilbert polynomials behave well with respect to short exact sequences.

Lemma 8.6. *Let R be a graded ring, and*

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0$$

be a degree-preserving short exact sequence of graded R-modules. Then $H_M = H_L + H_N$.

Proof. In each degree t we get short exact sequences of vector spaces

$$0 \longrightarrow L_t \xrightarrow{f} M_t \xrightarrow{g} N_t \longrightarrow 0 .$$

The claim follows from the Rank–Nullity Theorem from Linear Algebra:

$$\dim(M_t) = \dim(\operatorname{im} g) + \dim(\ker f) = \dim(\operatorname{im} g) + \dim(\operatorname{im} g) = \dim(N_t) + \dim(L_t).$$

$\square$

We can now compute a more sophisticated example.

Example 8.7. Let f be a homogeneous element of degree d in a $\mathbb{Z}$-graded ring R. We have the short exact sequence

$$0 \longrightarrow R(-d) \longrightarrow R \longrightarrow R/(f) \longrightarrow 0 .$$

By Lemma 8.6 and the definition of shift, this gives

$$H_R = H_{R(-d)} + H_{R/(f)} \implies H_{R/(f)}(t) = H_R(t) - H_{R(-d)}(t) = H_R(t) - H_R(t-d)$$

and

$$h_R = h_{R(-d)} + h_{R/(f)} \implies h_{R/(f)}(z) = h_R(z) - h_{R(-d)}(z) = h_R(z) - z^d h_R(z).$$

Then

$$H_{R/(f)}(t) = H_R(t) - H_R(t-d)$$

and

$$h_{R/(f)}(z) = (1 - z^d) h_R(z).$$

When $R = k[x_1, \dots, x_n]$, we saw in Example 8.4 that $H_R(t) = P_n(t)$ is a polynomial for $t > -n$. If $d < n$, then $t - d > -n$ for all $t \geqslant 0$, so

$$H_{R/(f)}(t) = P_n(t) - P_n(t-d) = \binom{t+n-1}{t} - \binom{t-d+n-1}{t}$$

is still given by a polynomial. When $d > n$, $H_{R/(f)}(t)$ still agrees with a polynomial *eventually*: for all $t \geqslant d - n$.

We can now show that Hilbert function is always eventually equal to a polynomial, as in Example 8.7.

Theorem 8.8. *Let k be a field, and R be a finitely graded k-algebra such that $R_0 = k$ and R is generated by elements of degree one. Let M be a finitely generated graded R-module.*[2] *There is a polynomial $P_M(t) \in \mathbb{Q}[t]$ and some $n \in \mathbb{N}$ such that $H_M(t) = P_M(t)$ for $t \geqslant n$. Moreover, $\deg(P_M) = \dim(M) - 1$, and we can write*

$$P_M(t) = \frac{e}{(\dim(M)-1)!} t^{\dim(M)-1} + \textit{ lower order terms}$$

for some positive integer e. Finally, if $\dim(M) = 0$ then $P_M = 0$.

Proof. We will use induction on the dimension of M.

If $\dim(M) = 0$, then M has finite length by Lemma 6.23. In particular, it must be finite dimensional as a k-vector space, so only finitely many graded pieces can be nonzero. So for $t \gg 0$, $H_M(t) = 0$, which is a polynomial of degree -1, by our convention.

Now suppose that the theorem holds for every ring R satisfying our hypotheses and for every R-module of dimension $n-1$. Assume M has dimension n, and take a homogeneous prime filtration of M, which we constructed in Theorem 5.28. Say this prime filtration is

$$M = M_m \supsetneq M_{m-1} \supsetneq M_{m-2} \supsetneq \cdots \supsetneq M_1 \supsetneq M_0 = 0$$

with $M_i/M_{i-1} \cong R/\mathfrak{p}_i(d_i)$ for some homogeneous primes $\mathfrak{p}_i$ and integers d_i. This breaks into short exact sequences

$$0 \longrightarrow M_i \longrightarrow M_{i+1} \longrightarrow M_{i+1}/M_i \longrightarrow 0\,.$$

Using Lemma 8.6 inductively on i, we get that $H_M(t) = H_{R/\mathfrak{p}_1(d_1)}(t) + \cdots + H_{R/\mathfrak{p}_m(d_m)}(t)$. Observe that $H_{R/\mathfrak{p}_i(d_i)}(t) = H_{R/\mathfrak{p}_i}(t + d_i)$ for each i. Since the associated primes of M are contained in $V(\operatorname{ann}_R(M))$, we have $\dim(R/\mathfrak{p}_i) \leqslant \dim(M)$. Moreover, there must be some $\mathfrak{p}_i$ for which equality occurs, since every associated prime of M occurs among the $\mathfrak{p}_i$, so in particular all the minimal primes of $\operatorname{ann}_R(M)$ are among the $\mathfrak{p}_i$. If we can show that each module of the form $R/\mathfrak{p}_i$ verifies the conclusion of the theorem, then we are done: all of the claims of polynomiality, degree, and positivity of leading term pass to $H_M(t)$ by the equality above, as the shifting does not change degree or the leading term, $\dim(M) = \max\{\dim(R/\mathfrak{p}_i)\}$, and the leading term satisfies the hypotheses again.

If $M = R/\mathfrak{p}_i$, then take a homogeneous Noether normalization A for this k-algebra M, and consider a homogeneous prime filtration for M as an A-module. Every factor is either a shift of A, or else has dimension less than $a := \dim(A) = \dim(M)$, since A is a domain. Applying the induction hypothesis and the formula

$$H_M(t) = H_{R/\mathfrak{p}_1(d_1)}(t) + \cdots + H_{R/\mathfrak{p}_m(d_m)}(t)$$

from above to this context, we find that $H_M(t)$ is a sum of shifts of the polynomial $P_a(t)$ from Example 8.4, plus polynomials of lower degree. Notice that

$$P_a(t) = \frac{(t+a-1)(t+a-2)\cdots(t+1)}{(a-1)!} = \frac{1}{(a-1)!} t^{a-1} + \text{ lower degree terms.}$$

Thus, the claims hold for M. $\square$

[2]Recall that the dimension of a module M is the dimension of $R/\operatorname{ann}_R(M)$.

Definition 8.9. The **Hilbert polynomial** of a graded module is the polynomial $P_M(t)$ that agrees with $H_M(t)$ for $t \gg 0$. The **multiplicity** of an R-module $M \neq 0$ is the positive integer $e(M)$ such that

$$P_M(t) = \frac{e(M)}{(\dim(M)-1)!} t^{\dim(M)-1} + \text{ lower order terms.}$$

Example 8.10. The multiplicity of a standard graded ring is $e(k[x_1, \ldots, x_n]) = 1$.

Proposition 8.11. *Let k be a field, and R be a finitely graded k-algebra such that $R_0 = k$ and R is generated by elements of degree one. Let*

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0$$

be a short exact sequence of graded R-modules. Then $P_M(t) = P_L(t) + P_N(t)$. If $\dim(L) = \dim(M) = \dim(N)$, then $e(M) = e(L) + e(N)$.

Proof. Both claims follow immediately from the fact that Hilbert functions are additive on short exact sequences, Lemma 8.6. $\square$

Example 8.12. If R is a polynomial ring, then $e(R) = 1$. If $R = S/fS$ for a polynomial ring S and a homogeneous element f of degree d, then $e(R) = d$.

Example 8.13. If $k[x_1, \ldots, x_n]$ is a standard $\mathbb{N}$-graded ring and f is a homogeneous element of degree d, then $R = k[x_1, \ldots, x_n]/(f)$ satisfies $e(R) = d$. We can compute this using Example 8.7.

We can use Theorem 8.8 something about the Hilbert series as well, thanks to the following lemma.

Lemma 8.14. *Let $\phi\colon \mathbb{N} \to \mathbb{N}$, and consider its generating function $g(z) = \sum_{i=0}^{\infty} \phi(i) z^i$. The following are equivalent:*

a) there is a polynomial $P \in \mathbb{Q}[t]$ of degree $d-1$ such that $\phi(n) = P(n)$ for $n \gg 0$

b) $g(z) = \dfrac{q(z)}{(1-z)^d}$ for some $q \in \mathbb{Z}[z]$ such that $q(1) \neq 0$ and $d \geqslant 0$.

Proof sketch. Use the "negative binomial" formula

$$\frac{1}{(1-z)^d} = \sum_{i=0}^{\infty} \binom{i+d-1}{d-1} z^i. \qquad \square$$

Corollary 8.15. *Let k be a field, and R be a finitely graded k-algebra such that $R_0 = k$ and R is generated by elements of degree one. Let M be a finitely generated graded R-module. Then the Hilbert series of M is of the form*

$$h_M(z) = \frac{q(z)}{(1-z)^d},$$

for some $q \in \mathbb{Z}[z]$, where $d = \dim(M)$.

Proof. Immediate from Theorem 8.8 and Lemma 8.14. □

We have now given many different characterizations for the dimension for a finitely generated graded k-algebra. Here's a summary:

Theorem 8.16 (The dimension theorem — graded version)**.** *Let k be a field, and R be a finitely generated graded k-algebra such that $R_0 = k$ and R is generated by elements of degree one, i.e. $R = R_0[R_1]$. The following numbers are equal:*

a) The Krull dimension of R.

b) The smallest d such that $\sqrt{(x_1, \dots, x_d)} = R_+$ for some homogeneous $x_1, \dots, x_d$.

c) $1 + \deg(P_R)$, where P_R is the Hilbert polynomial of R.

d) The order of pole of the Hilbert series of R at 1, that is, the number d such that $h_R(z) = \frac{q(z)}{(1-z)^d}$ and this fraction is in lowest terms, i.e. $q(1) \neq 0$.

Finally, we also want to consider the case when the ring is not necessarily generated in degree one. The key fact we will need is the following:

Exercise 21. Let k be a field, and R be a finitely generated positively graded k-algebra with $R_0 = k$. There is some $d \in \mathbb{N}$ such that the subring $R^{(d)} = \bigoplus_{i \geqslant 0} R_{id}$ is generated as a k-algebra by R_d.

The Hilbert function of a non-standard graded k-algebra is no longer eventually a polynomial. But it is eventually a *quasipolynomial*.

Definition 8.17. A function $f\colon \mathbb{Z} \longrightarrow \mathbb{R}$ is a **quasipolynomial** if there exists an integer b and polynomials $p_0, \dots, p_{b-1} \in \mathbb{R}[t]$ such that $f(n) = p_c(n)$ for $c \equiv n \bmod b$ for each $n \in \mathbb{Z}$.

So a quasipolynomial alternates between various polynomials.

Theorem 8.18. *Let k be a field, and R be a finitely graded k-algebra such that $R_0 = k$. Let M be a finitely generated graded R-module. Then there is a quasipolynomial with rational coefficients $P_M(t)$ such that $H_M(t) = P_M(t)$ for $t \gg 0$.*

Proof. Let d be such that $R^{(d)}$ is generated by R_d, which exists by Exercise 21. We can think of $R^{(d)}$ as a standard graded k-algebra, where we consider the elements of R_{id} to have degree i. Since R is finitely generated as a k-algebra, it is also a finitely generated $R^{(d)}$-algebra. Moreover, any homogeneous element $x \in R$ satisfies a monic equation of the form $t^d - x^d \in R^{(d)}[t]$, so R is integral over $R^{(d)}$. By Corollary 1.37, R is module-finite over $R^{(d)}$. So M is a finitely generated $R^{(d)}$-module. However, its grading over R is not consistent with the grading of $R^{(d)}$. We can decompose M as an $R^{(d)}$-module as

$$M = N_0 \oplus N_1 \oplus \cdots \oplus N_{d-1}, \quad \text{where } N_j = \bigoplus_{i \in \mathbb{N}} M_{j+id}.$$

Set $[N_j]_i := M_{j+id}$. This gives us a grading on each N_j that is compatible with $R^{(d)}$. Note also that each N_j is a submodule of a finitely generated module over a Noetherian ring, so is also finitely generated. Therefore, each N_j admits its own Hilbert polynomial. Taking each of these, we obtain a quasipolynomial that agrees with the Hilbert function for large values. □

One can then show (see [Mat89, Theorem 13.2]) that the Hilbert series of a finitely generated graded module over a non-standard graded k-algebra with $R_0 = k$ is of the form

$$\frac{q(t)}{(1-t)^{d_1}\cdots(1-t)^{d_n}},$$

where the integers d_i are the degrees of the algebra generators of R.

8.2 Associated graded rings and Hilbert functions for local rings

We next wish to give a version of the dimension theorem from the previous section in the local case. For this, we need a notion of Hilbert function that applies to local rings. We get this by associating a graded ring to each local ring.

Definition 8.19. The **associated graded ring** of an ideal I in a ring R is the ring

$$\mathrm{gr}_I(R) := \bigoplus_{n\geqslant 0} I^n/I^{n+1}$$

with n-th graded piece I^n/I^{n+1} and multiplication

$$(a+I^{n+1})(b+I^{m+1}) = ab + I^{m+n+1} \text{ for } a\in I^n,\ b\in I^m.$$

If $(R,\mathfrak{m})$ is local, then $\mathrm{gr}(R) := \mathrm{gr}_{\mathfrak{m}}(R)$ will be called the associated graded ring of R.

Note that the multiplication is well-defined.

Remark 8.20. If $a\in I^n$, $b\in I^m$, $u\in I^{n+1}$, $v\in I^{m+1}$, that is, $a+u\in a+I^{n+1}$ and $b+v\in b+I^{m+1}$ then

$$(a+u)(b+v) = ab = av + bu + uv \in ab + I^{m+n+1}.$$

So the multiplication on the associated graded ring is indeed well-defined.

Remark 8.21.

a) $[\mathrm{gr}_I(R)]_0 = R/I$, so if $(R,\mathfrak{m},k)$ is local then $[\mathrm{gr}(R)]_0 = R/\mathfrak{m} = k$.

b) Each graded piece $[\mathrm{gr}(R)]_n = I^n/I^{n+1}$ is an R-module annihilated by I, so it is an R/I-module. If $(R,\mathfrak{m},k)$ is local then $[\mathrm{gr}(R)]_n$ is a k-vector space.

c) Let R be Noetherian and $I = (f_1,\dots,f_n)$. Then $f_1+I^2,\dots,f_n+I^2 \in [\mathrm{gr}_I(R)]_1$ and $\mathrm{gr}_I(R) = (R/I)[f_1,\dots,f_n]$ is finitely generated as a $[\mathrm{gr}(R)]_0$-algebra by elements of degree one. If $(R,\mathfrak{m},k)$ is Noetherian and local, then by NAK Proposition 4.32 a basis for $[\mathrm{gr}(R)]_n$ corresponds to a minimal set of generators for $\mathfrak{m}^n$, and in particular $\dim_k[\mathrm{gr}(R)]_n = \mu(\mathfrak{m}^n) < \infty$.

Example 8.22. Take $R = k[x, y]$ and $I = (x, y)$. Then

$$\mathrm{gr}_I(R) = R/I \oplus I/I^2 \oplus I^2/I^3 \oplus \cdots = k \oplus (k\,x \oplus k\,y) \oplus (k\,x^2 \oplus k\,xy \oplus k\,y^2) \oplus \cdots$$

so we see that in fact $\mathrm{gr}_{(x,y)}(k[x, y]) = k[x, y]$. Similarly, for any graded k-algebra generated in degree 1, $\mathrm{gr}_{R_+}(R) = R$.

Example 8.23. Let $R = k[x, y]_{(x,y)}$.

a) Let $I = (x, y)$. The same computation as above applies to show

$$\mathrm{gr}_{(x,y)}\, k[x, y]_{(x,y)} = k[x, y].$$

b) Now take $I = (x^2, y^2)$. Then

$$\mathrm{gr}_I(R) = R/I \oplus (x^2R/I \oplus y^2R/I) \oplus (x^2R/I \oplus x^2y^2R/I \oplus y^2R/I) \oplus \cdots$$

so we get

$$\mathrm{gr}_I(R) = (R/I)[x^2, y^2]$$

with $\deg(x^2) = \deg(y^2) = 1$ and $\deg(r + I) = 0$ for all $r \in R$. In this case the R/I-algebra generators for $\mathrm{gr}_I(R)$ are algebraically independent.

c) Finally take $I = (x^2, xy)$. Then

$$\mathrm{gr}_I(R) = (R/I)[x^2, xy],$$

with $\deg(x^2) = \deg(xy) = 1$ and $\deg(r + I) = 0$ for all $r \in R$. However, in this case the algebra generators x^2, xy are not algebraically independent over R/I. For example, $\overline{y}x^2 - \overline{x}xy = 0$.

Definition 8.24. Let $(R, \mathfrak{m})$ be a local ring. The **Hilbert function** of R is

$$H_R(t) := H_{\mathrm{gr}(R)}(t).$$

and **Hilbert series** of R is

$$h_R(t) := h_{\mathrm{gr}(R)}(t).$$

Example 8.25. When $R = k[x, y]_{(x,y)}$,

$$H_R(t) = H_{k[x,y]}(t) = t + 1.$$

More generally, if $R = k[x_1, \ldots, x_n]_{(x_1,\ldots,x_n)}$ then

$$H_R(t) = H_{k[x_1,\ldots,x_n]}(t) = P_n(t)$$

as in Example 8.4.

To get a completely satisfactory analogue of the Hilbert function theory in this setting, we would like to understand the dimension of the associated graded ring. To understand this, we use the following related object.

Definition 8.26. Let R be a ring, and I an ideal. Recall that the **Rees algebra** of I is the $\mathbb{N}$-graded ring
$$R[It] = \bigoplus_{n \geqslant 1} I^n t^n = R \oplus It \oplus I^2t^2 \oplus \cdots \subseteq R[t],$$
The **extended Rees algebra** of I is the $\mathbb{Z}$-graded ring
$$R[It, t^{-1}] = \cdots \oplus Rt^{-2} \oplus Rt^{-1} \oplus R \oplus It \oplus I^2t^2 \oplus \cdots . \subseteq R[t, t^{-1}].$$
In both cases, the grading is given by setting $\deg t = 1$, and $\deg r = 0$ for all $r \in R$.

Note that $t \notin R[It, t^{-1}]$ since $1 \notin I$, so t^{-1} is not a unit, even though it looks like one.

Example 8.27. If $R = k[x, y]$ and $I = (x^2, y^2)$ then $R[It, t^{-1}] = R[x^2t, y^2t, t^{-1}]$. Think about t as being a constant which is allowed to vary in k. Then the extended Rees algebra of I can be viewed as a family of R-algebras, one for each value of t^{-1}. Let's explore some of the algebras in this family by plugging in values for t^{-1}:

- if $t^{-1} = 0$, which is ok to do since t^{-1} is not actually a unit, then we get
$$R[It, t^{-1}]|_{t^{-1}=0} = R[x^2t, y^2t] \cong \mathrm{gr}_I[t].$$
- If $t^{-1} = 1$ then we get $x^2t = x^2t \cdot 1 = x^2tt^{-1} = x^2 \in R$ and similarly $y^2t = y^2 \in R$ so
$$R[It, t^{-1}]|_{t^{-1}=1} = R[x^2_R, y^2_R] \cong R.$$

  In fact the same is true for every value $t^{-1} \in k^\times$.

The following lemma makes these observations rigorous.

Lemma 8.28. *There are isomorphisms*
$$R[It, t^{-1}]/(t^{-1}) \cong \mathrm{gr}_I(R) \qquad \textit{and} \qquad R[It, t^{-1}]/(t^{-1} - 1) \cong R.$$

Proof. For the first isomorphism, since t is homogeneous, we can use the graded structure. We have
$$t^{-1}R[It, t^{-1}] = \cdots \oplus Rt^{-2} \oplus Rt^{-1} \oplus I \oplus I^2t \oplus I^3t^2 \oplus \cdots,$$
so matching the graded pieces, we see that
$$\begin{aligned} R[It, t^{-1}]/(t^{-1}) &= \frac{\cdots \oplus Rt^{-2} \oplus Rt^{-1} \oplus R \oplus It \oplus I^2t^2 \oplus \cdots}{\cdots \oplus Rt^{-2} \oplus Rt^{-1} \oplus I \oplus I^2t \oplus I^3t^2 \oplus \cdots} \\ &\cong R/I \oplus I/I^2 \oplus I^2/I^3 \oplus \cdots \\ &= \mathrm{gr}_I(R). \end{aligned}$$

For the second isomorphism, we consider the map $R[It, t^{-1}] \longrightarrow R$ given by sending $t \mapsto 1$. This is surjective, and the kernel is the set of elements $a_mt^m + \cdots + a_nt^n$ such that $a_m + \cdots + a_n = 0$. We claim that this ideal is generated by $(t^{-1} - 1)$. We proceed by induction on $n - m$. The case $n - m = 0$ corresponds to there being at most one nonzero term, say at^m, in which case at^m is in the kernel if and only if $a = 0$. In $n - m = 1$, we have an element of the form $at^{n-1} - at^n$ for some a, which is of the form $(at^n)(t^{-1} - 1)$. For the inductive step, if $a_m + \cdots + a_n = 0$, write

$$a_m t^m + \cdots + a_n t^n = (a_m t^m + \cdots + (a_{n-1} + a_n)t^{n-1}) + (-a_n t^{n-1} + a_n t^n).$$

Observe that $-a_n t^{n-1} + a_n t^n \mapsto -a_n + a_n = 0$, and thus $a_m t^m + \cdots + (a_{n-1} + a_n)t^{n-1}$ must also be in the kernel. The induction hypothesis now applies to both $-a_n t^{n-1} + a_n t^n$ and $a_m t^m + \cdots + (a_{n-1} + a_n)t^{n-1}$, which must then be in $(t^{-1} - 1)$. Therefore, so is $a_m t^m + \cdots + a_n t^n$, and we are done. □

Lemma 8.29. *Let R be a Noetherian ring, and I an ideal in R. The minimal primes of $R[It, t^{-1}]$ are exactly the primes of the form $\mathfrak{p}R[t, t^{-1}] \cap R[It, t^{-1}]$ for $\mathfrak{p} \in \operatorname{Min}(R)$.*

Proof. Let $(0) = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_t$ be a minimal primary decomposition of (0) in R, and $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$. One can check (exercise!) that $\mathfrak{q}_i R[t]$ is primary with radical $\mathfrak{p}_i R[t]$. Then the same is true in $R[t, t^{-1}]$, by localizing at $\{1, t, t^2, \ldots\}$. Contracting to $R[It, t^{-1}]$, we get primary ideals that intersect to (0); none is contained in the intersection of the others, since this is the case after contracting to R, and likewise the radicals are distinct since they contract to different primes in R.

Therefore, setting $\mathfrak{q}_i' = \mathfrak{q}_i R[t, t^{-1}] \cap R[It, t^{-1}]$, $\mathfrak{q}_1' \cap \cdots \cap \mathfrak{q}_t'$ is a minimal primary decomposition of (0) in $R[It, t^{-1}]$, and thus the minimal primes in $R[It, t^{-1}]$ are $\mathfrak{p}_i R[t, t^{-1}] \cap R[It, t^{-1}]$ □

Theorem 8.30. *Let $(R, \mathfrak{m})$ be a Noetherian local ring, and $I \subseteq \mathfrak{m}$ an ideal. Then*

$$\dim(R) = \dim(R[It, t^{-1}]) - 1 = \dim(\operatorname{gr}_I(R)).$$

Proof. First, let's show $\dim(R) = \dim(R[It, t^{-1}]) - 1$. By Lemma 8.29, we can reduce to the case when R is a domain by localizing at each of the minimal primes of R. In particular, $R[It, t^{-1}]$ is also a domain.

By Lemma 8.28, $R[It, t^{-1}]/(t^{-1} - 1) \cong R$, so $\dim(R[It, t^{-1}]) \geqslant \dim(R)$. Also, since $(t^{-1} - 1)$ is principal, $\operatorname{ht}(t^{-1} - 1) \leqslant 1$, by Theorem 6.33. But $R[It, t^{-1}]$ is a domain, so $\operatorname{ht}(t^{-1} - 1) = 1$. By Corollary 7.31,

$$\dim R = \dim(R[It, t^{-1}]) - \operatorname{ht}(t^{-1} - 1) = \dim(R[It, t^{-1}]) - 1.$$

Now, we claim that

$$Q = \cdots \oplus Rt^{-2} \oplus Rt^{-1} \oplus \mathfrak{m} \oplus It \oplus I^2t^2 \oplus \cdots = (\mathfrak{m}, It, t^{-1})R[It, t^{-1}]$$

is a maximal ideal of height $\dim(R) + 1$ in $R[It, t^{-1}]$. The quotient ring is $R/\mathfrak{m}$, so it is clearly maximal. Given a chain $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_h = \mathfrak{m}$ of length $h = \dim(R)$, let $\mathfrak{q}_i = \mathfrak{p}_i R[t, t^{-1}] \cap R[It, t^{-1}]$. Since $\mathfrak{q}_i \cap R = \mathfrak{p}_i[t, t^{-1}] \cap R = \mathfrak{p}_i$, this is a proper chain of primes in $R[It, t^{-1}]$. We have

$$\mathfrak{q}_h = \cdots \oplus \mathfrak{m}t^{-2} \oplus \mathfrak{m}t^{-1} \oplus \mathfrak{m} \oplus It \oplus I^2t^2 \oplus \cdots = (\mathfrak{m}, It)R[It, t^{-1}] \subsetneq Q$$

so the height of Q is at least $\dim(R)+1$, and hence equal to $\dim(R)+1$ using the previous upper bound on the dimension.

For the last equality, since t^{-1} is a nonzerodivisor on $R[It, t^{-1}]$, we have

$$\dim(\mathrm{gr}_I(R)) \leqslant \dim(R[It, t^{-1}]) - 1.$$

For the other inequality, let $\overline{Q} = Q/(t^{-1})$. Then

$$\begin{aligned}\dim(\mathrm{gr}_I(R)) &\geqslant \dim(\mathrm{gr}_I(R)_{\overline{Q}}) = \dim(R[It, t^{-1}]_Q/(t^{-1})) \\ &\geqslant \dim(R[It, t^{-1}]_Q) - 1 \\ &= \mathrm{height}(Q) - 1 \\ &= \dim(R).\end{aligned}$$

□

Theorem 8.31. . *Let $(R, \mathfrak{m}, k)$ be a local ring. Then there is a polynomial $P_R(t) \in \mathbb{Q}[t]$ of degree equal to $\dim(R)-1$ such that $H_R(t) = P_R(t)$ for $t \gg 0$. Moreover, if $\dim(R) > 0$ then*

$$P + R(t) = \frac{e}{(\dim(R)-1)!} t^{\dim(R)-1} + \textit{ lower order terms.}$$

.

Proof. Because $H_R(t) = H_{\mathrm{gr}(R)}(t)$, we already know by Theorem 8.8 that $H_R(t)$ is eventually equal to a polynomial of degree $\dim(\mathrm{gr}(R))-1$ and that $(\dim(\mathrm{gr}(R))-1)!$ times the leading coefficient is positive. So the theorem follows as long as $\dim(R) = \dim(\mathrm{gr}(R)))$. □

Definition 8.32. The **Hilbert polynomial** of a local ring R is the polynomial $P_R(t)$ that agrees with $H_R(t)$ for $t \gg 0$. The **multiplicity** of R is the positive integer $e(R)$ such that

$$P_R(t) = \frac{e(R)}{(\dim(R)-1)!} t^{\dim(M)-1} + \text{ lower order terms.}$$

This gives an analogue of the dimension theorem in the local case:

Theorem 8.33 (The dimension theorem — local version)**.** *Let $(R, \mathfrak{m}, k)$ be a Noetherian ring. The following numbers are equal:*

a) the Krull dimension of R.

b) $1 + \deg(P_R)$, where P_R is the Hilbert polynomial of R (and $\mathrm{gr}(R)$).

c) The order of pole of the Hilbert series of R (really, of $\mathrm{gr}(R)$) at 1, that is, the number d such that

$$h_R(z) = \frac{q(z)}{(1-z)^d}$$

and this fraction is in lowest terms with $q(1) \neq 0$.

Part II

Homological Algebra

Chapter 9

What is homological algebra?

Homological algebra first appeared in the study of topological spaces. Roughly speaking, homology is a way of associating a sequence of abelian groups (or modules, or other more sophisticated algebraic objects) to another object, for example a topological space. The homology of a topological space encodes in topological information about the space in algebraic language — this is what algebraic topology is all about.

Despite its very concrete origins, modern homological algebra lives somewhere in the realm of abstract nonsense, and can be described as the study of abelian categories. An abelian category is, roughly speaking, modeled after **Ab**, the category of abelian groups and abelian group homomorphisms. It is this more general setting, rather than the classical topological setting, that we are interested in. However, it is very hard to understand general abelian categories without some good concrete examples we can get our hands on, so we should first start from a reasonable middle ground. While it might be very difficult to get a concrete handle on a general abelian category, it turns out that every reasonable abelian category embeds into R-**mod**, the category of R-modules and R-module homomorphisms. In this first chapter we introduce the main characters — complexes, homology, categories — and then we will spend most of our time studying R-**mod**, while keeping in mind that what happens in a general abelian category is simply an abstraction of what happens in R-**mod**.

9.1 Complexes and homology

Homological algebra is a branch of abstract algebra that studies the homology of chain complexes.

Definition 9.1. A **chain complex** of R-modules $(C_\bullet, \delta_\bullet)$, also referred to simply as a **complex**, is a sequence of R-modules C_i and R-module homomorphisms

$$\cdots \longrightarrow C_{n+1} \xrightarrow{\delta_{n+1}} C_n \xrightarrow{\delta_n} C_{n-1} \longrightarrow \cdots$$

such that $\delta_n \delta_{n+1} = 0$ for all n. We refer to C_n as the module in **homological degree** n. The maps δ_n are the **differentials** of our complex. We may sometimes omit the differentials δ_n and simply refer to the complex $C_\bullet$ or even C; we may also sometimes refer to $\delta_\bullet$ as *the* differential of $C_\bullet$.

In some contexts, it is important to make a distinction between chain complexes and co-chain complexes, where the arrows go the opposite way: a co-chain complex would look like

$$\cdots \longrightarrow C_{n-1} \xrightarrow{\delta_n} C_n \xrightarrow{\delta_{n+1}} C_{n+1} \longrightarrow \cdots.$$

We will not need to make such a distinction, so we will call both of these complexes and most often follow the convention in the definition above. We will say a complex is **bounded above** if $F_n = 0$ for all $n \gg 0$, and **bounded below** if $F_n = 0$ for all $n \ll 0$. A **bounded complex** is one that is both bounded above and below. If a complex is bounded, we may sometimes simply write it as a finite complex, say

$$C_n \xrightarrow{\delta_n} C_{n-1} \longrightarrow \cdots \longrightarrow C_m.$$

Remark 9.2. The condition that $\delta_n \delta_{n+1} = 0$ for all n implies that $\operatorname{im} \delta_{n+1} \subseteq \ker \delta_n$.

Definition 9.3. The complex $(C_\bullet, \delta_\bullet)$ is **exact** at n if $\operatorname{im} \delta_{n+1} = \ker \delta_n$. An **exact sequence** is a complex that is exact everywhere.

Historically, chain complexes first appeared in topology. To study a topological space, one constructs a particular chain complex that arises naturally from information from the space, and then calculates its homology, which ends up encoding important topological information in the form of a sequence of abelian groups.

Definition 9.4 (Homology)**.** The **homology** of the complex $(C_\bullet, \delta_\bullet)$ is the sequence of R-modules

$$\mathrm{H}_n(C_\bullet) := \frac{\ker \delta_n}{\operatorname{im} \delta_{n+1}}.$$

The n**th homology** of $(C_\bullet, \delta_\bullet)$ is $\mathrm{H}_n(C_\bullet)$. The submodules $Z_n(C_\bullet) := \ker \delta_n \subseteq C_n$ are called **cycles**, while the submodules $B_n(C_\bullet) := \operatorname{im} \delta_{n+1} \subseteq C_n$ are called **boundaries**.

The homology of a complex measures how far our complex is from being exact at each point. Again, we can talk about the **cohomology** of a cochain complex instead; we will for now not worry about the distinction.

Remark 9.5. Note that $(C_\bullet, \delta_\bullet)$ is exact at n if and only if $\mathrm{H}_n(C_\bullet) = 0$.

Example 9.6. Let $\mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/2\mathbb{Z}$ be the canonical projection map. Then

$$C = \underset{2}{\mathbb{Z}} \xrightarrow{4} \underset{1}{\mathbb{Z}} \xrightarrow{\pi} \underset{0}{\mathbb{Z}/2\mathbb{Z}}$$

is a complex, since the image of multiplication by 4 is $4\mathbb{Z}$, and that is certainly contained in

$\ker \pi = 2\mathbb{Z}$. The homology of C is

$$\begin{aligned}
\mathrm{H}_n(C) &= 0 && \text{for } n \geqslant 3 \\
\mathrm{H}_2(C) &= \frac{\ker(\mathbb{Z} \xrightarrow{4} \mathbb{Z})}{\operatorname{im}(0 \longrightarrow \mathbb{Z})} = \frac{0}{0} = 0 \\
\mathrm{H}_1(C) &= \frac{\ker(\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z})}{\operatorname{im}(\mathbb{Z} \xrightarrow{4} \mathbb{Z})} = \frac{2\mathbb{Z}}{4\mathbb{Z}} \cong \mathbb{Z}/2\mathbb{Z} \\
\mathrm{H}_0(C) &= \frac{\ker(\mathbb{Z}/2\mathbb{Z} \longrightarrow 0)}{\operatorname{im}(\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z})} = \frac{\mathbb{Z}/2\mathbb{Z}}{\mathbb{Z}/2\mathbb{Z}} = 0 \\
\mathrm{H}_n(C) &= 0 && \text{for } n < 0
\end{aligned}$$

Notice that our complex is exact at 2 and 0. The exactness at 2 says that the map $\mathbb{Z} \xrightarrow{4} \mathbb{Z}$ is injective, while exactness at 0 says that $\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z}$ is surjective.

Those last remarks were not a coincidence. We repeat Remark 1.11 but now in this more general context.

Remark 9.7. The sequence $0 \longrightarrow A \xrightarrow{f} B$ is exact if and only if f is injective. Similarly, $B \xrightarrow{f} C \longrightarrow 0$ is exact if and only if f is surjective. So

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is a short exact sequence if and only if

- f is injective
- g is surjective
- $\operatorname{im} f = \ker g$.

Requiring that f be injective is the same as asking that $\ker f = 0$, while g is surjective if and only if $\operatorname{coker} g = 0$.[1] When this is indeed a short exact sequence, we can identify A with its image $f(A)$, and $A = \ker g$. Moreover, since g is surjective, by the First Isomorphism Theorem we conclude that $C \cong B/A$, so we might abuse notation and identify C with B/A.

Remark 9.8. The complex $0 \longrightarrow M \xrightarrow{f} N \longrightarrow 0$ is exact if and only if f is an isomorphism.

Remark 9.9. The complex $0 \longrightarrow M \longrightarrow 0$ is exact if and only if $M = 0$.

Example 9.10. Let π be the canonical projection $\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z}$. The following is a short exact sequence:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0\,.$$

We will most often be interested in **complexes of R-modules**, where the abelian groups that show up are all modules over the same ring R.

[1] The **cokernel** of a homomorphism $f\colon M \longrightarrow N$ is the R-module $N/\operatorname{im} f$.

Example 9.11. Let $R = k[x]$ be a polynomial ring over the field k. Then following is a short exact sequence:

$$0 \longrightarrow R \xrightarrow{\cdot x} R \xrightarrow{\pi} R/(x) \longrightarrow 0\ .$$

The first map is multiplication by x, and the second map is the canonical projection.

Here are some more examples of complexes.

Example 9.12. Let $R = k[x]/(x^2)$. The following complex is exact:

$$\cdots \longrightarrow R \xrightarrow{\cdot x} R \xrightarrow{\cdot x} R \longrightarrow \cdots .$$

Indeed, the image and the kernel of multiplication by x are both (x).

Example 9.13. Let $R = k[x]/(x^3)$. Consider the following complex:

$$F_\bullet = \ \cdots \longrightarrow R \xrightarrow{\cdot x^2} R \xrightarrow{\cdot x^2} R \longrightarrow \cdots .$$

The image of multiplication by x^2 is (x^2), while the the kernel of multiplication by x^2 is $(x) \supseteq (x^2)$. For all n,

$$\mathrm{H}_n(F_\bullet) = (x)/(x^2) \cong R/(x).$$

9.2 Categories for the working homological algebraist

Most fields in modern mathematics follow the same basic recipe: there is a main type of object one wants to study — groups, rings, modules, topological spaces, etc — and a natural notion of arrows between these — group homomorphisms, ring homomorphisms, module homomorphisms, continuous maps, etc. The objects are often sets with some extra structure, and the arrows are often maps between the objects that preserve whatever that extra structure is. Category theory is born of this realization, by abstracting the basic notions that make math and studying them all at the same time. How many times have we felt a sense of déjà vu when learning about a new field of math? Category theory unifies all those ideas we have over and over in different contexts.

Category theory is an entire field of mathematics in its own right. As such, there is a lot to say about category theory, and unfortunately it doesn't all fit in the little time we have to cover it in this course. We include here some basic definitions and ideas from category theory we will need throughout the course, but you are strongly encouraged to learn more about category theory, for example from [ML98] or [Rie17].

First, we want to note that there is a long and fun story about why we used the word *c*ollection when describing the objects in a category. Not all collections are allowed to be sets, an issue that was first discovered by Russel with his famous Russel's Paradox. Russel exposed the fact that one has to be careful with how we formalize set theory. We follow the ZFC (Zermelo–Fraenkel with choice, short for the Zermelo–Fraenkel axioms plus the Axiom of Choice) axiomatization of set theory, and while we will not discuss the details of this formalization here, you are encouraged to read more on the subject.

Definition 9.14. A **category** $\mathscr{C}$ consists of three different pieces of data:

- a collection of **objects**, **ob** $(\mathscr{C})$,
- for each two objects, say A and B, a collection $\mathrm{Hom}_{\mathscr{C}}(A, B)$ of **arrows** or **morphisms** between A and B, and
- for each three objects A, B, and C, a composition

$$\begin{aligned} \mathrm{Hom}_{\mathscr{C}}(A,B) \times \mathrm{Hom}_{\mathscr{C}}(B,C) &\longrightarrow \mathrm{Hom}_{\mathscr{C}}(A,C)\ . \\ (f,g) &\longmapsto g \circ f \end{aligned}$$

We will often drop the $\circ$ and write simply fg for $f \circ g$.

These ingredients satisfy the following axioms:

1) The $\mathrm{Hom}_{\mathscr{C}}(A, B)$ are all disjoint. In particular, if f is an arrow in $\mathscr{C}$, we can talk about its **source** A and its **target** B as the objects such that $f \in \mathrm{Hom}_{\mathscr{C}}(A, B)$.
2) For each object A, there is an **identity arrow** $1_A \in \mathrm{Hom}_{\mathscr{C}}(A, A)$ such that $1_A \circ f = f$ and $g \circ 1_A = g$ for all $f \in \mathrm{Hom}_{\mathscr{C}}(B, A)$ and all $g \in \mathrm{Hom}_{\mathscr{C}}(A, B)$.
3) Composition is associative, meaning $f \circ (g \circ h) = (f \circ g) \circ h$.

Here are some categories you have likely encountered before:

Example 9.15.

a) The category **Set** with objects all sets and arrows all functions between sets.
b) The category **Grp** whose objects are the collection of all groups, and whose arrows are all the homomorphisms of groups. The identity arrows are the identity homomorphisms.
c) The category **Ab** whose objects are the collection of all abelian groups, and whose arrows are the homomorphisms of abelian groups. The identity arrows are the identity homomorphisms.
d) The category **Ring** of rings and ring homomorphisms.[2]
e) The category R-**mod** of modules over a fixed ring R and with R-module homomorphisms. Sometimes one writes R-**Mod** for this category, and reserve R-**mod** for the category of finitely generated R-modules with R-module homomorphisms.
f) The category **Top** of topological spaces and continuous functions.

While the objects and arrows might not actually be sets, sometimes they are.

Definition 9.16. A category $\mathscr{C}$ is **locally small** if for all objects A and B in $\mathscr{C}$, $\mathrm{Hom}_{\mathscr{C}}(A, B)$ is a set. A category $\mathscr{C}$ is **small** if it is locally small and the collection of all objects in $\mathscr{C}$ is a set.

Many important categories one would think about are at least locally small. For example, **Set** is locally small but not small. In a locally small category, we can now refer to its Hom-sets.

Not all categories consist of sets with extra structure and functions between them.

[2]Contrary to what you may expect, this is not nearly as important as the next one.

Example 9.17. Given a partially ordered set X, we can regard X as a category: the objects are the elements of X, and $\mathrm{Hom}_X(x, y)$ is either a singleton if $x \leqslant y$ or empty if $x \nleqslant y$.

This category is clearly locally small, since all nonempty Hom-sets are in fact singletons.

There are some special types of arrows we may want to consider.

Definition 9.18. Let $\mathscr{C}$ be any category.

- An arrow $f \in \mathrm{Hom}_{\mathscr{C}}(A, B)$ is an **isomorphism** if there exists $g \in \mathrm{Hom}_{\mathscr{C}}(B, A)$ such that $gf = 1_A$ and $fg = 1_B$. Unsurprisingly, such an arrow g is called the **inverse** of f.
- An arrow $f \in \mathrm{Hom}(B, C)$ is **monic**, a **monomorphism**, or a **mono** if for all arrows

$$\begin{tikzcd} A \arrow[r, shift left, "g_1"] \arrow[r, shift right, "g_2"'] & B \arrow[r, "f"] & C \end{tikzcd}$$

  if $fg_1 = fg_2$ then $g_1 = g_2$.
- Similarly, an arrow $f \in \mathrm{Hom}(A, B)$ is an **epi** or an **epimorphism** if for all arrows

$$\begin{tikzcd} A \arrow[r, "f"] & B \arrow[r, shift left, "g_1"] \arrow[r, shift right, "g_2"'] & C \end{tikzcd}$$

  if $g_1 f = g_2 f$ then $g_1 = g_2$.

We follow a familiar pattern and define the related concepts one can guess should be defined.

Definition 9.19. A **subcategory** $\mathscr{C}$ of a category $\mathscr{D}$ consists of a subcollection of the objects of $\mathscr{D}$ and a subcollection of the morphisms of $\mathscr{D}$ such that the following hold:

- For every object C in $\mathscr{C}$, the arrow $1_C \in \mathrm{Hom}_{\mathscr{D}}(C, C)$ is an arrow in $\mathscr{C}$.
- For every arrow in $\mathscr{C}$, its source and target in $\mathscr{D}$ are objects in $\mathscr{C}$.
- For every pair of arrows f and g in $\mathscr{C}$ such that fg is an arrow that makes sense in $\mathscr{D}$, fg is an arrow in $\mathscr{C}$.

In particular, $\mathscr{C}$ is a category in its own right.

Example 9.20. The category of finitely generated R-modules with R-module homomorphisms is a subcategory of R-**mod**.

Of course that once we have categories, we want maps between categories as well.

Definition 9.21. Let $\mathscr{C}$ and $\mathscr{D}$ be categories. A **covariant functor** $F \colon \mathscr{C} \longrightarrow \mathscr{D}$ is a mapping that assigns to each object A in $\mathscr{C}$ an object $F(A)$ in $\mathscr{D}$, and to each arrow $f \in \mathrm{Hom}_{\mathscr{C}}(A, B)$ an arrow $F(f) \in \mathrm{Hom}_{\mathscr{D}}(F(A), F(B))$, such that

- F preserves the composition of maps, meaning $F(fg) = F(f)F(g)$ for all arrows f and g in $\mathscr{C}$, and
- F preserves the identity arrows, meaning $F(1_A) = 1_{F(A)}$ for all objects A in $\mathscr{C}$.

A **contravariant functor** $F\colon \mathscr{C} \longrightarrow \mathscr{D}$ is a mapping that assigns to each object A in $\mathscr{C}$ an object $F(A)$ in $\mathscr{D}$, and to each arrow $f \in \operatorname{Hom}_{\mathscr{C}}(A, B)$ an arrow $F(f) \in \operatorname{Hom}_{\mathscr{D}}(F(B), F(A))$, such that

- F preserves the composition of maps, meaning

$$F(fg) = F(g)F(f)$$

  for all composable arrows f and g in $\mathscr{C}$, and
- F preserves the identity arrows, meaning $F(1_A) = 1_{F(A)}$ for all objects A in $\mathscr{C}$.

So a contravariant functor is a functor that flips all the arrows. We can also describe a contravariant functor as a covariant functor from $\mathscr{C}$ to the opposite category of $\mathscr{D}$, $\mathscr{D}^{\mathrm{op}}$, which is a category built out of $\mathscr{D}$ by flipping all the arrows.

Definition 9.22. Let $\mathscr{C}$ be a category. The **opposite category** of $\mathscr{C}$, denoted $\mathscr{C}^{\mathrm{op}}$, is a category whose objects are the objects of $\mathscr{C}$, and such that each arrow $f \in \operatorname{Hom}_{\mathscr{C}^{\mathrm{op}}}(A, B)$ is the same as some arrow in $\operatorname{Hom}_{\mathscr{C}^{\mathrm{op}}}(A, B)$. The composition of two morphisms fg in $\mathscr{C}^{\mathrm{op}}$ is defined as the composition gf in $\mathscr{C}$.

Many objects and concepts one might want to describe are obtained from existing ones by flipping the arrows. Opposite categories give us the formal framework to talk about such things. We will often want to refer to **dual** notions, which will essentially mean considering the same notion in a category $\mathscr{C}$ and in the opposite category $\mathscr{C}^{\mathrm{op}}$; in practice, this means we should flip all the arrows involved. We will see examples of this later on.

In particular, we can now make our previous remark more precise.

Remark 9.23. A contravariant functor $F\colon \mathscr{C} \longrightarrow \mathscr{D}$ can be thought of as a covariant functor $\mathscr{C}^{\mathrm{op}} \longrightarrow \mathscr{D}$, or also as a covariant functor $\mathscr{C} \longrightarrow \mathscr{D}^{\mathrm{op}}$. If using one of these conventions, one needs to be careful, however, when composing functors, so that the respective sources and targets match up correctly. While we haven't specially discussed how one composes functors, it should be clear that applying a functor $F : \mathscr{C} \longrightarrow \mathscr{D}$ and $G : \mathscr{D} \longrightarrow \mathscr{E}$ is the same as applying a functor $\mathscr{C} \longrightarrow \mathscr{D}$, which we can write as GF.

For example, if $F : \mathscr{C} \longrightarrow \mathscr{D}$ and $G : \mathscr{D} \longrightarrow \mathscr{E}$ are both contravariant functors, the composition $GF\colon \mathscr{C} \longrightarrow \mathscr{E}$ is a covariant functor, since

$$\begin{tikzcd}
A \arrow[d, "f"'] & \rightsquigarrow & F(A) & \rightsquigarrow & GF(A) \arrow[d, "GF(f)"'] \\
B & & F(B) \arrow[u, "F(f)"'] & & GF(B)
\end{tikzcd}$$

So we could think of F as a covariant functor $\mathscr{C} \longrightarrow \mathscr{D}^{\mathrm{op}}$ and G as a covariant functor $\mathscr{D}^{\mathrm{op}} \longrightarrow \mathscr{E}$. Similarly, if $F : \mathscr{C} \longrightarrow \mathscr{D}$ is a covariant functor and $G : \mathscr{D} \longrightarrow \mathscr{E}$ is a contravariant functor, $GF\colon \mathscr{C} \longrightarrow \mathscr{E}$ is a contravariant functor. In this case, we can think of G as a covariant functor $\mathscr{D} \longrightarrow \mathscr{E}^{\mathrm{op}}$, so that GF is now a covariant functor $\mathscr{C} \longrightarrow \mathscr{E}^{\mathrm{op}}$.

But back to functors, let's see some examples.

Example 9.24. Here are some examples of functors you may have encountered before.

a) In many categories one may think about, the objects are sets with some extra structure, and the arrows are functions between those sets that preserved that extra structure. The **forgetful functor** from such a category to **Set** is the functor that, just as the name says, *forgets* that extra structure, and sees only the underlying sets and functions of sets. For example, the forgetful functor **Gr** $\longrightarrow$ **Set** sends each group to its underlying set, and each group homomorphism to the corresponding function of sets.

b) The identity functor on any category $\mathscr{C}$ does what the name suggests: it sends each object to itself and each arrow to itself.

c) Localization is a functor. Let R be a ring and W be a multiplicatively closed set in R. There is localization at W induces a a functor R-mod $\longrightarrow W^{-1}R$-mod that sends each R-module M to $W^{-1}M$, and each R-module homomorphism $\alpha\colon M \longrightarrow N$ to the R-module homomorphism $W^{-1}\alpha : W^{-1}M \longrightarrow W^{-1}N$.

Remark 9.25. Any functor sends isomorphisms to isomorphisms, since it preserves compositions and identities.

If we think about functors as functions between categories, it's natural to consider what would be the appropriate versions of the notions of injective or surjective.

Definition 9.26. A covariant functor $F : \mathscr{C} \longrightarrow \mathscr{D}$ between locally small categories is

- **faithful** if all the functions of sets
$$\begin{aligned} \mathrm{Hom}_{\mathscr{C}}(A,B) &\longrightarrow \mathrm{Hom}_{\mathscr{D}}(F(A),F(B)) \\ f &\longmapsto F(f) \end{aligned}$$
are injective.
- **full** if all the functions of sets
$$\begin{aligned} \mathrm{Hom}_{\mathscr{C}}(A,B) &\longrightarrow \mathrm{Hom}_{\mathscr{D}}(F(A),F(B)) \\ f &\longmapsto F(f) \end{aligned}$$
are surjective.
- **fully faithful** if it is full and faithful.
- an **embedding** if it is fully faithful and injective on objects.

Example 9.27. The forgetful functor R**-mod** $\longrightarrow$ **Set** is faithful since any two maps of R-modules with the same source and target coincide if and only if they are the same function of sets. This functor is not full, since there not every functions between the underlying sets of two R-modules is an R-module homomorphism.

Remark 9.28. A fully faithful functor is not necessarily injective on objects, but it is injective on objects up to isomorphism.

Definition 9.29. A subcategory $\mathscr{C}$ of $\mathscr{D}$ is a **full subcategory** if $\mathscr{C}$ includes *all* of the arrows in $\mathscr{D}$ between any two objects in $\mathscr{C}$. In other words, a subcategory is full if the inclusion functor $F\colon \mathscr{C} \longrightarrow \mathscr{D}$ is full.

Example 9.30.

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

b) The category whose objects are all sets and with arrows all bijections is a subcategory of **Set** that is not full.

And finally, because mathematicians do not lack a sense of humor, things do not stop there. We can then consider mappings between functors, because why not.[3]

Definition 9.31. Let F and G be covariant functors $\mathscr{C} \longrightarrow \mathscr{D}$. A **natural transformation** between F and G is a mapping that to each object A in $\mathscr{C}$ assigns an arrow $\eta_A \in \mathrm{Hom}_{\mathscr{D}}(F(A), G(A))$ such that for all $f \in \mathrm{Hom}_{\mathscr{C}}(A, B)$, the diagram

$$\begin{tikzcd} F(A) \arrow[r, "\eta_A"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\ F(B) \arrow[r, "\eta_B"'] & G(B) \end{tikzcd}$$

commutes. A **natural isomorphism** is a natural transformation η where each η_A is an isomorphism. We sometimes write $\begin{tikzcd} \mathscr{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow \eta"] & \mathscr{D} \end{tikzcd}$ or simply $\eta \colon F \implies G$.

Saying that the diagram commutes means that the resulting compositions are all the same whichever way we go, meaning $\tau_B F(f) = G(f)\tau_A$. We could have started our story by saying that homological algebra is the study of commutative diagrams, and we wouldn't exactly be far off from reality.

Often, when studying a particular topic, we sometimes say a certain map is *natural* to mean that there is actually a natural transformation behind it.

Example 9.32. The abelianization of a group G is the abelian group $G^{\mathrm{ab}} = G/[G, G]$, where $[G, G]$ denotes the commutator subgroup of G, the subgroup generated by all commutators of elements in G, given by

$$[G, G] := \left\langle ghg^{-1}h^{-1} \mid g, h \in G \right\rangle.$$

It turns out that $[G, G]$ is a normal subgroup of G, and so G^{ab} is simply a quotient of G. In particular, the abelianization comes equipped with a natural projection map $\pi_G \colon G \longrightarrow G^{\mathrm{ab}}$, the usual quotient map from G to a normal subgroup. Here we mean natural in two different ways: both that this is common sense map to consider, and that this is in fact coming from a natural transformation. What's happening behind the scenes is that abelianization is a functor ab : **Grp** $\longrightarrow$ **Grp**. On objects, the abelianizations functor is defined as $G \longrightarrow G^{\mathrm{ab}}$. Given an arrow, meaning a group homomorphism $G \xrightarrow{f} H$, one can check that $[G, G]$ is contained in the kernel of $\pi_H f$, so $\pi_H f$ factors through G^{ab}, and there exists a

[3] And also because they are useful.

group homomorphism f^{ab} making the following diagram commute:

$$\begin{tikzcd} G \arrow[r, "\pi_G"] \arrow[d, "f"'] & G^{\mathrm{ab}} \arrow[d, dashed, "f^{\mathrm{ab}}"] \\ H \arrow[r, "\pi_H"] & H^{\mathrm{ab}} \end{tikzcd}.$$

So the abelianization functor takes the arrow f to f^{ab}. The commutativity of the diagram above says that $\pi_{\_}$ is a natural transformation between the identity functor on **Grp** and the abelianization functor, which we can write more compactly as $\begin{tikzcd} \mathbf{Grp} \arrow[r, bend left, "\mathrm{id}"] \arrow[r, bend right, "\mathrm{ab}"'] \arrow[r, phantom, "\Downarrow\pi"] & \mathbf{Grp} \end{tikzcd}$.

Definition 9.33. Let $F, G : \mathscr{C} \longrightarrow \mathscr{D}$ be two functors between the categories $\mathscr{C}$ and $\mathscr{D}$. We write

$$\mathrm{Nat}(F, G) = \{\text{natural transformations } F \longrightarrow G\}.$$

Given two categories $\mathscr{C}$ and $\mathscr{D}$, one can build a **functor category**[4] with objects all covariant functors $\mathscr{C} \longrightarrow \mathscr{D}$, and arrows the corresponding natural transformations. This category is denoted $\mathscr{D}^{\mathscr{C}}$. Sometimes one writes $\mathrm{Hom}(F, G)$ for $\mathrm{Nat}(F, G)$, but we will avoid that, as it might make things even harder to follow.

For the functor category to truly be a category, though, we need to know how to compose natural transformations.

Remark 9.34. Consider natural transformations

$$\begin{tikzcd} \mathscr{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow\varphi"] & \mathscr{D} \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} \mathscr{C} \arrow[r, bend left, "G"] \arrow[r, bend right, "H"'] \arrow[r, phantom, "\Downarrow\eta"] & \mathscr{D} \end{tikzcd}.$$

We can compose them for form a new natural transformation

$$\begin{tikzcd} \mathscr{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "H"'] \arrow[r, phantom, "\Downarrow\eta\varphi"] & \mathscr{D} \end{tikzcd}$$

We should think of this composition as happening *vertically*. For each object C in $\mathscr{C}$, $\eta\varphi$ sends C to the arrow $F(A) \xrightarrow{\varphi_A} G(A) \xrightarrow{\eta_A} H(A)$. This makes the diagram

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, "\varphi_A"'] & F(B) \arrow[d, "\varphi_B"'] \\ G(A) \arrow[r, "G(f)"] \arrow[d, "\eta_A"'] & G(B) \arrow[d, "\eta_B"'] \\ H(A) \arrow[r, "H(f)"] & H(B) \end{tikzcd}$$

commute.

[4]Yes, the madness is neverending.

Even though this is only a short introduction to category theory, we would be remiss not to mention the Yoneda Lemma, arguably the most important statement in category theory. For that, we will need some of the most important functors we will discuss in this class.

Definition 9.35. Let $\mathscr{C}$ be a locally small category. An object A in $\mathscr{A}$ induces two Hom functors:

- The covariant functor $\mathrm{Hom}_{\mathscr{C}}(A, -) : \mathscr{C} \longrightarrow \mathbf{Set}$ sends each object B to the set $\mathrm{Hom}_{\mathscr{C}}(A, B)$, and each $f \in \mathrm{Hom}_{\mathscr{C}}(B, C)$ to the function $\mathrm{Hom}_{\mathscr{C}}(f, B) =: f_*$

$$\begin{aligned} \mathrm{Hom}_{\mathscr{C}}(A, B) &\longrightarrow \mathrm{Hom}_{\mathscr{C}}(A, C)\ . \\ g &\longmapsto fg \end{aligned}$$

- The contravariant functor $\mathrm{Hom}_{\mathscr{C}}(-, B) : \mathscr{C} \longrightarrow \mathbf{Set}$ sends each object A to the set $\mathrm{Hom}_{\mathscr{C}}(A, B)$, and each $f \in \mathrm{Hom}_{\mathscr{C}}(A, C)$ to the function $\mathrm{Hom}_{\mathscr{C}}(A, f) =: f^*$

$$\begin{aligned} \mathrm{Hom}_{\mathscr{C}}(A, B) &\longrightarrow \mathrm{Hom}_{\mathscr{C}}(C, B)\ . \\ g &\longmapsto gf \end{aligned}$$

Theorem 9.36 (Yoneda Lemma)**.** *Let $\mathscr{C}$ be a locally small category, and fix an object A in $\mathscr{C}$. Let $F\colon \mathscr{C} \longrightarrow \mathbf{Set}$ be a covariant functor. Then there is a bijection*

$$\mathit{Nat}(\mathrm{Hom}_{\mathscr{C}}(A, -), F) \xrightarrow{\ \gamma\ } F(A)\ .$$

Moreover, this correspondence is natural in both A and F.

Proof. Let φ be a natural transformation in $\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A, -), F)$. The proof of Yoneda's Lemma is essentially the following diagram:

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{C}}(A, A) \arrow[rrr, "\mathrm{Hom}_{\mathscr{C}}(A,f)"] \arrow[ddd, "\varphi_A"'] & & & \mathrm{Hom}_{\mathscr{C}}(A, X) \arrow[ddd, "\varphi_X"] \\
& 1_A \arrow[r, mapsto] \arrow[d, mapsto] & f \arrow[d, mapsto] & \\
& u \arrow[r, mapsto] & (F(f))u = \varphi_X(f) & \\
F(A) \arrow[rrr, "F(f)"'] & & & F(X)
\end{tikzcd}$$

Our bijection will be defined by $\gamma(\varphi) := \varphi_A(1_A)$. We should first check that this makes sense: arrows in **Set** are just functions between sets, and so φ_A is a function of sets $\mathrm{Hom}_{\mathscr{C}}(A, A) \longrightarrow F(A)$. Also, $\mathrm{Hom}_{\mathscr{C}}(A, A)$ is a set that contains at least the element 1_A, and $\varphi_A(1_A)$ is some element in the set $F(A)$.

Given any fixed $f \in \mathrm{Hom}_{\mathscr{C}}(A, X)$, the fact that φ is a natural transformation translates into the outer commutative diagram. In particular, the maps of sets $F(f)\varphi_A$ and $\varphi_X \mathrm{Hom}_{\mathscr{C}}(A, f)$ coincide, and must in particular take 1_A to the same element in $F(X)$. This is the commutativity of the inner diagram, with $u := \varphi_A(1_A)$.

The commutativity of the diagram above says that φ is completely determined by $\varphi_A(1_A)$, since for any other object X in $\mathscr{C}$ and any arrow $f \in \mathrm{Hom}_{\mathscr{C}}(A, X)$, we necessarily have $\varphi_X(f) = F(f)\varphi_A(1_A)$. In particular, our map $\gamma(\varphi) = \varphi_A(1_A)$ is injective. Moreover, note that each choice of $u \in F(A)$ gives rise to a different natural transformation φ by setting $\varphi_X(f) = F(f)u$. So our map γ is indeed a bijection.

We now have two naturality statements to prove. Naturality in the functor means that given a natural isomorphism $\eta\colon F \longrightarrow G$, the diagram

$$\begin{tikzcd}
\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A, -), F) \arrow[r, "\gamma_F"] \arrow[d, "\eta_*"'] & F(A) \arrow[d, "\eta_A"] \\
\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A, -), G) \arrow[r, "\gamma_G"'] & G(A)
\end{tikzcd}$$

commutes. Given a natural transformation φ between $\mathrm{Hom}_{\mathscr{C}}(A, -)$ and F,

$$\begin{aligned}
\eta_A \circ \gamma_F(\varphi) &= \eta_A(\varphi_A(1_A)) && \text{by definition of } \gamma \\
&= (\eta \circ \varphi)_A(1_A) && \text{by definition of composition of natural transformations} \\
&= \gamma_G(\eta \circ \varphi) && \text{by definition of } \gamma \\
&= \gamma_G \circ \eta_*(\varphi) && \text{by definition of } \eta_*
\end{aligned}$$

so commutativity does hold. Naturality on the object means that given an arrow $f\colon A \longrightarrow B$, the diagram

$$\begin{tikzcd}
\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A, -), F) \arrow[r, "\gamma"] \arrow[d, "(f^*)^*"'] & F(A) \arrow[d, "F(f)"] \\
\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(B, -), F) \arrow[r, "\gamma"'] & F(B)
\end{tikzcd}$$

commutes. Given a natural transformation φ between $\mathrm{Hom}_{\mathscr{C}}(A, -)$ and F,

$$F(f) \circ \gamma_A(\varphi) = F(f)(\varphi_A(1_A)),$$

while

$$\gamma_B \circ (f^*)^*(\varphi) = \gamma_B(\varphi \circ f^*) = (\varphi \circ f^*)_B(1_B).$$

Now notice that

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{C}}(B, B) \arrow[r, "f^*"] & \mathrm{Hom}_{\mathscr{C}}(A, B) \arrow[r, "\varphi_B"] & F(B) \; . \\
1_B \arrow[r, mapsto] & f \arrow[r, mapsto] & \varphi_B(f)
\end{tikzcd}$$

Let's look back at the big commutative diagram we started our proof with. It says, in particular, that $\varphi_B(f) = F(f)(\varphi_A(1_A))$. So again commutativity does hold, and we are done. $\square$

One can naturally (pun intended) define the notion of functor category of contravariant functors, and then prove the corresponding Yoneda Lemma, which will instead use the contravariant Hom functor.

Exercise 22 (Contravariant version of the Yoneda Lemma)**.** Let $\mathscr{C}$ be a locally small category, and fix an object B in $\mathscr{C}$. Let $F\colon \mathscr{C} \longrightarrow \mathbf{Set}$ be a contravariant functor. Then there is a bijection

$$\mathrm{Nat}\ (\mathrm{Hom}_{\mathscr{C}}(-,B),F) \xrightarrow{\gamma} F(B)\ .$$

In a way, the Yoneda Lemma says that to give a natural transformation between the functors $\mathrm{Hom}_{\mathscr{C}}(A,-)$ and F is choosing an element in $F(A)$.

Remark 9.37. Notice that the Yoneda Lemma says in particular that the collection of all natural transformations from $\mathrm{Hom}_{\mathscr{C}}(A,-)$ to F is a set. This wasn't clear a priori, since the collection of objects in $\mathscr{C}$ is not necessarily a set.

Remark 9.38. If we apply the Yoneda Lemma 9.36 to the case when F itself is also a Hom functor, say $F = \mathrm{Hom}_{\mathscr{C}}(B,-)$, the Yoneda Lemma says that there is a bijection between $\mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A,-),\mathrm{Hom}_{\mathscr{C}}(B,-))$ and $\mathrm{Hom}_{\mathscr{C}}(B,A)$. In particular, each arrow in $\mathscr{C}$ determines a natural transformation between Hom functors.

One of the consequences of the Yoneda Lemma is the Yoneda Embedding, which roughly says that every locally small category can be embedded into the category of contravariant functors from $\mathscr{C}$ to $\mathbf{Set}$. In particular, the Yoneda embedding says that natural transformations between representable functors correspond to arrows between the representing objects.

Theorem 9.39 (Yoneda Embedding)**.** *Let $\mathscr{C}$ be a locally small category. The covariant functor*

$$\begin{tikzcd}
\mathscr{C} \arrow[rr] & & \boldsymbol{Set}^{\mathscr{C}^{op}} \\
A \arrow[d, "f"'] & \longmapsto & \mathrm{Hom}_{\mathscr{C}}(-,A) \arrow[d, "f_*"] \\
B & & \mathrm{Hom}_{\mathscr{C}}(-,B)
\end{tikzcd}$$

*from $\mathscr{C}$ to the category of contravariant functors $\mathscr{C} \longrightarrow$ **Set** is an embedding. Moreover, the contravariant functor*

$$\begin{tikzcd}
\mathscr{C} \arrow[rr] & & \boldsymbol{Set}^{\mathscr{C}} \\
A \arrow[d, "f"'] & \longmapsto & \mathrm{Hom}_{\mathscr{C}}(A,-) \\
B & & \mathrm{Hom}_{\mathscr{C}}(B,-) \arrow[u, "f^*"']
\end{tikzcd}$$

*from the category $\mathscr{C}$ to the category of covariant functors $\mathscr{C} \longrightarrow$ **Set** is also an embedding.*

Proof. First, note that our functors are injective on objects because the Hom-sets in our category are all disjoint. We need to check that given objects A and B in $\mathscr{C}$, we have bijections

$$\mathrm{Hom}_{\mathscr{C}}(A,B) \cong \mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(-,A),\mathrm{Hom}_{\mathscr{C}}(-,B))$$

and

$$\mathrm{Hom}_{\mathscr{C}^{\mathrm{op}}}(A,B) \cong \mathrm{Nat}(\mathrm{Hom}_{\mathscr{C}}(A,-),\mathrm{Hom}_{\mathscr{C}}(B,-)).$$

We will do the details for the first one, and leave the second as an exercise.

First, let us take a sanity check and confirm that indeed our proposed functors take arrows $f\colon A \longrightarrow B$ in $\mathscr{C}$ to natural transformations between $\mathrm{Hom}_{\mathscr{C}}(-,A)$ and $\mathrm{Hom}_{\mathscr{C}}(-,B)$. This is essentially the content of Remark 9.38, but let's carefully check the details. The Yoneda Lemma 9.36 applied here tells us that each natural transformation φ between $\mathrm{Hom}_{\mathscr{C}}(-,A)$ and $F = \mathrm{Hom}_{\mathscr{C}}(-,B)$ corresponds to an element $u \in \mathrm{Hom}_{\mathscr{C}}(A,B)$, which we obtain by taking $u := \varphi_A(1_A)$. As we discussed in the proof of the Yoneda Lemma 9.36, we can recover φ from u by taking the natural transformation φ that for each object X in $\mathscr{C}$ has $\varphi_X : \mathrm{Hom}_{\mathscr{C}}(X,A) \longrightarrow \mathrm{Hom}_{\mathscr{C}}(X,B)$ given by $\varphi_X(f) = \mathrm{Hom}_{\mathscr{C}}(f,B)(u) = f_*(u)$.

We can see that different arrows f give rise to different natural transformations by applying the resulting natural transformation $f_;*$ to the identity arrow 1_A, which takes it to f. Moreover, the Yoneda Lemma 9.36 tells us that every natural transformation φ between $\mathrm{Hom}_{\mathscr{C}}(-,A)$ and $\mathrm{Hom}_{\mathscr{C}}(-,B)$ is the image of some u, as described above. $\square$

The functors that are naturally isomorphic to some Hom functor are important.

Definition 9.40. A covariant functor $F : \mathscr{C} \longrightarrow \mathbf{Set}$ is **representable** if there exists an object A in $\mathscr{C}$ such that F is naturally isomorphic to $\mathrm{Hom}_{\mathscr{C}}(A,-)$. A contravariant functor $F : \mathscr{C} \longrightarrow \mathbf{Set}$ is **representable** if there exists an object B in $\mathscr{C}$ such that F is naturally isomorphic to $\mathrm{Hom}_{\mathscr{C}}(-,B)$.

Example 9.41. We claim that the identity functor $\mathbf{Set} \longrightarrow \mathbf{Set}$ is representable. Let $\mathbf{1}$ be a singleton set. Given any set X, there is a bijection between elements $x \in X$ and functions $\mathbf{1} \longrightarrow X$ sending the one element in $\mathbf{1}$ to each x. Moreover, given any other set Y, and a function $f\colon X \longrightarrow Y$, our bijections make the following diagram commute:

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{Set}}(\mathbf{1},X) \arrow[r, "\cong"] \arrow[d, "f_*"'] & X \arrow[d, "f"] \\
\mathrm{Hom}_{\mathbf{Set}}(\mathbf{1},Y) \arrow[r, "\cong"] & Y.
\end{tikzcd}$$

This data gives a natural isomorphism between the identity functor and $\mathrm{Hom}_{\mathbf{Set}}(\mathbf{1},-)$.

A representable functor encodes a universal property of the object that represents it. For example, in Example 9.41, mapping out of the singleton set is the same as choosing an element x in a set X. We have all seen constructions that are at first a bit messy but that end up satisfying some nice universal property that makes everything work out. At the end of the day, a universal property allows us to ignore the messy details and focus on the universal property, which usually says everything we need to know about the construction. And as you may have already realized, universal properties are *everywhere*. Here is a formal definition.

Definition 9.42. Let $\mathscr{C}$ be a locally small category. A **universal property** of an object C in $\mathscr{C}$ consists of a representable functor $F : \mathscr{C} \longrightarrow \mathbf{Set}$ together with a **universal element**

$X \in F(C)$ such that F is naturally isomorphic to either $\mathrm{Hom}_{\mathscr{C}}(C, -)$ (if F is covariant) or $\mathrm{Hom}_{\mathscr{C}}(-, C)$ (if F is contravariant), via the natural isomorphism that corresponds to X via the bijection in the Yoneda Lemma 9.36.

We can rephrase this in terms of universal arrows.

Definition 9.43. Let $F : \mathscr{C} \longrightarrow \mathscr{D}$ be covariant functor. Given an object $D \in \mathscr{D}$, a **universal arrow from** D **to** F is a unique pair (U, u) where U is an object in $\mathscr{C}$ and a unique arrow $u \in \mathrm{Hom}_{\mathscr{D}}(D, F(U))$ with the following **universal property**: for any arrow $f \in \mathrm{Hom}_{\mathscr{D}}(D, F(Y))$, there exists a unique arrow $h \in \mathrm{Hom}_{\mathscr{C}}(U, Y)$ such that the following diagram commutes:

$$\begin{tikzcd} U \arrow[d, dashed, "h"'] \\ Y \end{tikzcd} \qquad \begin{tikzcd} D \arrow[r, "u"] \arrow[dr, "f"'] & F(U) \arrow[d, dashed, "F(h)"] \\ & F(Y) \end{tikzcd}$$

There is a dual to this definition. A **universal arrow from** F **to** D is a unique pair (U, u), where C is an object in $\mathscr{C}$ and $u \in \mathrm{Hom}_{\mathscr{D}}(F(U), D)$ that satisfy the following **universal property**: for any arrow $f \in \mathrm{Hom}_{\mathscr{D}}(F(Y), D)$, there exists a unique $h \in \mathrm{Hom}_{\mathscr{C}}(Y, U)$ such that the following diagram commutes:

$$\begin{tikzcd} U \\ Y \arrow[u, dashed, "h"] \end{tikzcd} \qquad \begin{tikzcd} D & F(U) \arrow[l, "u"'] \\ & F(Y) \arrow[ul, "f"] \arrow[u, dashed, "F(h)"'] \end{tikzcd}$$

Remark 9.44. Let $F : \mathscr{C} \longrightarrow \mathscr{D}$ be a covariant functor, and fix an object U in $\mathscr{C}$, an object DD in $\mathscr{D}$, and an arrow $u \in \mathrm{Hom}_{\mathscr{D}}(D, F(U))$. Notice that $\mathrm{Hom}_{\mathscr{D}}(D, F(-))$ determines a covariant functor $\mathscr{C} \longrightarrow$ **Set**. By the Yoneda Lemma 9.36, the following is a recipe for a natural transformation between $\mathrm{Hom}_{\mathscr{C}}(U, -)$ and $\mathrm{Hom}_{\mathscr{D}}(D, F(-))$: for each object Y in $\mathscr{C}$ and each arrow $h \in \mathrm{Hom}_{\mathscr{C}}(U, Y)$, set $\varphi_X(h) := \mathrm{Hom}_{\mathscr{D}}(D, F(h))(u)$. Notice that

$$\begin{tikzcd} \mathrm{Hom}_{\mathscr{D}}(D, F(U)) \arrow[r, "\mathrm{Hom}_{\mathscr{D}}(D, F(h))"] & \mathrm{Hom}_{\mathscr{D}}(D, F(Y)) , \\ f \arrow[r, mapsto] & F(h) \circ u \end{tikzcd}$$

so $\varphi_X(h)(f) = F(h) \circ u$.

We get the following commutative diagram:

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{C}}(U,U) \arrow[rrr, "\mathrm{Hom}_{\mathscr{C}}(U,h)"] \arrow[ddd, "\varphi_U"'] & & & \mathrm{Hom}_{\mathscr{C}}(U,Y) \arrow[ddd, "\varphi_Y"] \\
& 1_U \arrow[r, mapsto] \arrow[d, mapsto] & h \arrow[d, mapsto] & \\
& u \arrow[r, mapsto] & F(h)\circ u =: \varphi_Y(h) & \\
\mathrm{Hom}_{\mathscr{D}}(D,F(U)) \arrow[rrr, "\mathrm{Hom}_{\mathscr{D}}(D,F(h))"'] & & & \mathrm{Hom}_{\mathscr{D}}(D,F(Y))
\end{tikzcd}$$

Given an arrow $f \in \mathrm{Hom}_{\mathscr{D}}(D, F(Y))$, $\varphi_Y(h) = f$ for some $h \in \mathrm{Hom}_{\mathscr{C}}(U, Y)$ if and only if $F(h) \circ u = f$.

On the one hand, φ is a natural isomorphism if and only if for every object Y in $\mathscr{C}$ and every $f \in \mathrm{Hom}_{\mathscr{D}}(D, F(Y))$ there exists a unique $h \in \mathrm{Hom}_{\mathscr{C}}(U, Y)$ such that $F(h) \circ u = f$. On the other hand, that is exactly the condition required for (U, u) to be a universal arrow from D to F. So we have shown that the following are equivalent:

- (U, u) is a universal arrow from D to F.
- U represents the functor $\mathrm{Hom}_{\mathscr{D}}(D, F(-)) \colon \mathscr{C} \longrightarrow \mathbf{Set}$, via $u \in \mathrm{Hom}_{\mathscr{D}}(D, F(U))$.

Similarly, one can prove the dual statement:

- (U, u) is a universal arrow from F to D.
- U represents the functor $\mathrm{Hom}_{\mathscr{D}}(F(-), D) \colon \mathscr{C} \longrightarrow \mathbf{Set}$, via $u \in \mathrm{Hom}_{\mathscr{D}}(F(U), D)$.

Products and coproducts are a great example of constructions with universal properties.

Definition 9.45. Let $\mathscr{C}$ be a locally small category, and consider a family of objects $\{A_i\}_{i\in I}$ in $\mathscr{C}$. The **product** of the A_i is an object in $\mathscr{C}$, denoted by $\prod_i A_i$ or $A_1 \times \cdots \times A_n$ if I is finite, together with arrows $\pi_j \in \mathrm{Hom}_{\mathscr{C}}(\prod_i A_i, A_j)$ for each j, satisfying the following universal property: given any object B in $\mathscr{C}$ and arrows $f_i \colon B \longrightarrow A_i$ for each i,

$$\begin{tikzcd}
& B \arrow[dl, "f_i"'] \arrow[d, dashed, "f"] \arrow[dr, "f_j"] & \\
A_i & \prod_i A_i \arrow[l, "\pi_i"] \arrow[r, "\pi_j"'] & A_j
\end{tikzcd}$$

The **coproduct** of the A_i is an object in $\mathscr{C}$, denoted by $\coprod_i A_i$ or in some contexts $\bigoplus_i A_i$, together with arrows $\iota_j \in \mathrm{Hom}_{\mathscr{C}}(A_j, \coprod_i A_i,)$ for each j, satisfying the following universal property: given any object B in $\mathscr{C}$ and arrows $f_i \colon A_i \longrightarrow B$ for each i, the following diagram commutes:

$$\begin{tikzcd}
& B & \\
A_i \arrow[ur, "f_i"] \arrow[r, "\iota_i"'] & \coprod_i A_i \arrow[u, dashed, "f"] & A_j \arrow[ul, "f_j"'] \arrow[l, "\iota_j"]
\end{tikzcd}$$

Let's phrase the universal property of products as a universal property in this formal sense, at least in the case of the product of two object C_1 and C_2 in $\mathscr{C}$. To do that, we need to consider the **product category** $\mathscr{C}\times\mathscr{C}$ with objects given by pairs (C_1, C_2) of objects in $\mathscr{C}$ and arrows in $(C_1, C_2) \longrightarrow (C_3, C_4)$ given by pairs of arrows (f_1, f_2) with $f_1 \in \operatorname{Hom}_{\mathscr{C}}(C_1, C_3)$ and $f_2 \in \operatorname{Hom}_{\mathscr{C}}(C_2, C_4)$. The diagonal functor $\Delta\colon \mathscr{C} \longrightarrow \mathscr{C}\times\mathscr{C}$ is exactly what it sounds like: $\Delta(C) = (C, C)$ for every object C in $\mathscr{C}$ and $\Delta(f) = (f, f)$ for every arrow f in $\mathscr{C}$.

Given objects X and Y in $\mathscr{C}$, consider the projection arrows $\pi_1 : X \times Y \longrightarrow X$ and $\pi_2\colon X\times Y \longrightarrow Y$. We claim that the object $X\times Y$ together with the arrow (π_1, π_2) in $\mathscr{C}\times\mathscr{C}$ form a universal arrow from Δ to (X, Y) in $\mathscr{C}\times\mathscr{C}$. Why? This means that given any arrow $(f_1, f_2) \in \operatorname{Hom}_{\mathscr{C}\times\mathscr{C}}((X_1, X_2), \Delta(Y))$, there exists a unique $h \in \mathscr{C}(X_1 \times X_2, Y)$ such that

$$\begin{tikzcd} X_1 \times X_2 \\ Y \arrow[u, dashed, "h"] \end{tikzcd} \qquad \begin{tikzcd} (X_1, X_2) & \Delta(X_1 \times X_2) \arrow[l, "(\pi_1,\pi_2)"'] \\ & \Delta(Y) \arrow[ul, "(f_1,f_2)"] \arrow[u, dashed, "\Delta(h)"'] \end{tikzcd}$$

commutes. This is indeed the universal property of products we just described less formally above: given $f_1\colon Y \longrightarrow X_1$ and $f_2\colon Y \longrightarrow X_2$, there is a unique $h\colon Y \longrightarrow X_1 \times X_2$ such that

$$\begin{tikzcd} X_1 \times X_2 \\ Y \arrow[u, dashed, "h"] \end{tikzcd} \qquad \begin{tikzcd} (X_1, X_2) & (X_1 \times X_2, X_1 \times X_2) \arrow[l, "(\pi_1,\pi_2)"'] \\ & (Y, Y) \arrow[ul, "(f_1,f_2)"] \arrow[u, dashed, "\Delta(h)"'] \end{tikzcd}$$

Equivalently, following the recipe we described in Remark 9.44, the universal property of the product is encoded in the representable functor $\operatorname{Hom}_{\mathscr{C}\times\mathscr{C}}(\Delta(-), (X_1, X_2))$, which is represented by $X_1 \times X_2$ via (π_1, π_2). So more precisely, that says that there is a natural isomorphism

$$\operatorname{Hom}_{\mathscr{C}}(-, X_1 \times X_2) \cong \operatorname{Hom}_{\mathscr{C}\times\mathscr{C}}(\Delta(-), (X_1, X_2)),$$

and that is precisely the natural transformation that the Yoneda bijection we constructed in Theorem 9.36 takes to $(\pi_1, \pi_2) \in \operatorname{Hom}_{\mathscr{C}}(\Delta(X_1 \times X_2), (X_1, X_2))$. If we follow that bijection, our natural isomorphism φ sends an object Y in $\mathscr{C}$ to the arrow

$$\begin{aligned} \operatorname{Hom}_{\mathscr{C}}(Y, X_1 \times X_2) &\xrightarrow{\quad\varphi_Y\quad} \operatorname{Hom}_{\mathscr{C}\times\mathscr{C}}(\Delta(Y), (X_1, X_2)) \\ f &\longmapsto \left(\Delta(Y) \xrightarrow{(f,f)} \Delta(X_1 \times X_2) \xrightarrow{(\pi_1,\pi_2)} (X_1, X_2)\right). \end{aligned}$$

Since φ_Y is a bijection, every arrow $(f_1, f_2) \in \operatorname{Hom}_{\mathscr{C}\times\mathscr{C}}(\Delta(Y), (X_1, X_2))$ is $\varphi_Y(f)$ for some $f \in \operatorname{Hom}_{\mathscr{C}}(Y, X_1 \times X_2)$. Ultimately, this means that there exists f such that $f_1 = \pi_1 f$ and $f_2 = \pi_2 f$. And suprise surprise: we just rediscovered the universal property of the product.

Universal properties are closely related to adjoint functors.

Definition 9.46. Let $\mathscr{C}$ and $\mathscr{D}$ be locally small categories. Two covariant functors

$$\mathscr{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{D}$$

form an **adjoint** pair (F, G) if given any objects $C \in \mathscr{C}$ and $D \in \mathscr{D}$, there is a bijection between the Hom-sets

$$\mathrm{Hom}_{\mathscr{D}}(F(C), D) \xrightarrow{\cong} \mathrm{Hom}_{\mathscr{C}}(C, G(D))$$

which is natural on both objects, meaning that for all $f \in \mathrm{Hom}_{\mathscr{C}}(C_1, C_2)$ and $g \in \mathrm{Hom}_{\mathscr{D}}(D_1, D_2)$, the diagrams

$$\begin{tikzcd} \mathrm{Hom}_{\mathscr{D}}(F(C_2), D) \arrow[r, "\cong"] \arrow[d, "F(f)^*"'] & \mathrm{Hom}_{\mathscr{C}}(C_2, G(D)) \arrow[d, "f^*"] \\ \mathrm{Hom}_{\mathscr{D}}(F(C_1), D) \arrow[r, "\cong"] & \mathrm{Hom}_{\mathscr{C}}(C_1, G(D)) \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} \mathrm{Hom}_{\mathscr{D}}(F(C), D_1) \arrow[r, "\cong"] \arrow[d, "g_*"'] & \mathrm{Hom}_{\mathscr{C}}(C, G(D_1)) \arrow[d, "G(g)_*"] \\ \mathrm{Hom}_{\mathscr{D}}(F(C), D_2) \arrow[r, "\cong"] & \mathrm{Hom}_{\mathscr{C}}(C, G(D_2)) \end{tikzcd}$$

commute for all $C \in \mathscr{C}$ and all $D \in \mathscr{D}$. We say that F is the **left adjoint** of G, or that F has a **right adjoint**, and that G is the **right adjoint** of F, or that G has a **left adjoint**.

We can think of adjoint functors as solutions to optimization problems. A particular adjoint functor gives the most efficient functorial solution to some problem.

Example 9.47. Given a set I, what is the most efficient way to assign an R-module to I in a functorial way? The solution to this problem is the construction of free modules, the functor $\mathbf{Free}\colon \mathbf{Set} \longrightarrow R\text{-}\mathbf{mod}$ that sends each set I to the free R-module R^I on I. The free functor is precisely a left adjoint to the forgetful functor $R\text{-}\mathbf{mod} \longrightarrow \mathbf{Set}$.

As Mac Lane said [ML98], "the slogan is *adjoint functors arise everywhere*".

Remark 9.48. We can rephrase the condition that $G\colon \mathscr{D} \longrightarrow \mathscr{C}$ has a left adjoint functor $F : \mathscr{C} \longrightarrow \mathscr{D}$ as follows: for every object C in $\mathscr{C}$, there is a universal arrow from C to G, and for every object D in $\mathscr{D}$ there exists a universal arrow from F to D. To see that, let $\eta_D \in \mathrm{Hom}_{\mathscr{D}}(F(G(D)), D)$ be the image of the identity on $\mathrm{Hom}_{\mathscr{D}}(G(D), G(D))$ via the bijection

$$\begin{tikzcd} \mathrm{Hom}_{\mathscr{C}}(G(D), G(D)) \arrow[r, "\cong"] & \mathrm{Hom}_{\mathscr{D}}(F(G(D)), D) \\ \mathrm{id}_{G(D)} \arrow[r, mapsto] & \eta_D \end{tikzcd}$$

given by the definition of adjoint functors, and let $\varepsilon_C \in \mathrm{Hom}_{\mathscr{C}}(C, GF(C))$ be be the image of the identity on $\mathrm{Hom}_{\mathscr{C}}(F(C), F(C))$ via the bijection

$$\begin{tikzcd} \mathrm{Hom}_{\mathscr{D}}(F(C), F(C)) \arrow[r, "\cong"] & \mathrm{Hom}_{\mathscr{D}}(C, GF(C)) \\ \mathrm{id}_{F(C)} \arrow[r, mapsto] & \varepsilon_C \end{tikzcd}.$$

We claim that $(F(C), \varepsilon_C)$ is a universal arrow from C to G. That would mean that given arrow $f \in \mathrm{Hom}_{\mathscr{C}}(C, G(Y))$, there must exist a unique arrow $h \in \mathrm{Hom}_{\mathscr{D}}(F(C), Y)$ such that the following diagram commutes:

$$\begin{tikzcd} F(C) \arrow[d, dashed, "h"'] \\ Y \end{tikzcd} \qquad \begin{tikzcd} D \arrow[r, "\varepsilon_C"] \arrow[dr, "f"'] & G(F(C)) \arrow[d, dashed, "G(h)"] \\ & G(Y). \end{tikzcd}$$

This says that $G(h)_*(\varepsilon_C) = G(h) \circ \varepsilon_C = f$, which means that

$$\begin{tikzcd}
1_{F(C)} \arrow[rrr, mapsto] \arrow[ddd, mapsto] & & & \varepsilon_C \arrow[ddd, mapsto] \\
& \mathrm{Hom}_{\mathscr{D}}(F(C), F(C)) \arrow[r, "\cong"] \arrow[d, "h_*"'] & \mathrm{Hom}_{\mathscr{C}}(C, GF(C)) \arrow[d, "G(h)_*"] & \\
& \mathrm{Hom}_{\mathscr{D}}(F(C), Y) \arrow[r, "\cong"'] & \mathrm{Hom}_{\mathscr{C}}(C, G(Y)) & \\
h \arrow[rrr, mapsto] & & & f
\end{tikzcd}$$

On the one hand, such an h does exist: just take $h \in \mathrm{Hom}_{\mathscr{C}}(F(C), Y)$ that is sent to f via the bijection between $\mathrm{Hom}_{\mathscr{D}}(F(C), Y)$ and $\mathrm{Hom}_{\mathscr{C}}(C, G(Y))$. Since this map is a bijection, such an h is unique.

Similarly, we claim that $(G(D), \eta_D)$ is a universal arrow from F to D. That would mean that for any arrow $f \in \mathrm{Hom}_{\mathscr{D}}(F(Y), D)$, there exists a unique $h \in \mathrm{Hom}_{\mathscr{C}}(Y, G(D))$ such that the following diagram commutes:

$$\begin{tikzcd}
G(D) & & D & F(G(D)) \arrow[l, "\eta_D"'] \\
Y \arrow[u, dashed, "h"] & & & F(Y) \arrow[u, dashed, "F(h)"'] \arrow[ul, "f"]
\end{tikzcd}$$

This means that $(F(h))^*(\eta_D) = \eta_D \circ F(h) = f$, which means that

$$\begin{tikzcd}
1_{G(D)} \arrow[rrr, mapsto] \arrow[ddd, mapsto] & & & \eta_D \arrow[ddd, mapsto] \\
& \mathrm{Hom}_{\mathscr{C}}(G(D), G(D)) \arrow[r, "\cong"] \arrow[d, "h^*"'] & \mathrm{Hom}_{\mathscr{D}}(D, FG(D)) \arrow[d, "F(h)^*"] & \\
& \mathrm{Hom}_{\mathscr{C}}(G(D), Y) \arrow[r, "\cong"'] & \mathrm{Hom}_{\mathscr{D}}(D, F(Y)) & \\
h \arrow[rrr, mapsto] & & & f
\end{tikzcd}$$

Again, such an h exists and it is unique because it must correspond to f via the bijection between $\mathrm{Hom}_{\mathscr{D}}(D, F(Y))$ and $\mathrm{Hom}_{\mathscr{C}}(G(D), Y)$.

We can talk about *the* left or right adjoint to a given functor.

Exercise 23. Left and right adjoints are unique up to natural isomorphism. More precisely, given an adjoint pair of functors (F, G), show that if G' is also a right adjoint to F, then G' and G are naturally isomorphic. Similarly, one can show that if F' is also a left adjoint to G, then F and F' are naturally isomorphic.

Example 9.49. Fix a ring R. The forgetful functor $F\colon R\textbf{-mod} \longrightarrow \textbf{Set}$ has a left adjoint. That left adjoint is the functor $G\colon \textbf{Set} \longrightarrow R\textbf{-mod}$ that takes a set S to the free R-module $\oplus_S R$ on S, meaning the R-module whose elements are finite formal linear combinations

$r_1 s_1 + \cdots + r_n s_n$ of elements $s_i \in S$ with coefficients $r_i \in R$. This is the same free module we described much earlier. Each function of sets defines an R-module map by sending each basis element to its image via the given function.

Even without any category theory, one often describes the free R-module on a set S by the following universal property: given a function f from a set S to an R-module M, there exists a unique R-module homomorphism from the free module $\oplus_S R$ to M that agrees with f on the basis elements. And indeed, one can check that this is the universal property we formally obtain from the fact that the free R-module functor is left adjoint to the forgetful functor from R-**mod**.

This type of *free* construction is quite common, and often gives rise to adjunctions. We can think about the free functor from **Set** to R-**mod** as the most efficient way of defining an R-module from a given set. It's efficient because it comes with a nice universal property.

We close this short detour into the wonderful world of category theory to point out that if we wanted to sound really obscure, we could have defined chain complexes in this categorical language.

Remark 9.50. First, we view $\mathbb{Z}$ as a partially ordered set under $\geqslant$. As in Example 9.15 9.17, $\mathbb{Z}$ now gives us a category whose objects are the integers, and where we have an arrow in $\operatorname{Hom}_{\mathbb{Z}}(n, m)$ if $n \geqslant m$. If we ignore the identity maps $\operatorname{Hom}_{\mathbb{Z}}(n, n)$ and composite maps, we can represent this category in the following diagram:

$$\cdots \longrightarrow n+1 \longrightarrow n \longrightarrow n-1 \longrightarrow \cdots.$$

From this perspective, a chain complex is a functor $F\colon \mathbb{Z} \longrightarrow \mathbf{Ab}$: for each $n \in \mathbb{Z}$, we get an R-module F_n, and we also get an R-module homomorphisms $F_{n+1} \longrightarrow F_n$ for each n. Indeed, this can all be represented as a sequence

$$\cdots \longrightarrow F_{n+1} \longrightarrow F_n \longrightarrow F_{n-1} \longrightarrow \cdots.$$

For our functor to truly be a complex, though, we must require that all compositions $F_{n+1} \longrightarrow F_n \longrightarrow F_{n-1}$ be 0. A map of complexes, also known as a chain map, is a natural transformation between two such functors.

9.3 Maps of complexes

Unsurprisingly, we can form a category of complexes, but to do that we need the right definition of maps between complexes. We also take this section as a chance to set up some definitions we will need later.

Definition 9.51. Let $(F_\bullet, \delta_\bullet^F)$ and $(G_\bullet, \delta_\bullet^G)$ be complexes. A **map of complexes** or a **chain map**, which we write as $h\colon (F_\bullet, \delta_\bullet^F) \longrightarrow (G_\bullet, \delta_\bullet^G)$ or simply $h\colon F_\bullet \longrightarrow G_\bullet$, is a sequence of homomorphisms of R-modules $h_n\colon F_n \longrightarrow G_n$ such that the following diagram commutes:

$$\begin{tikzcd}
\cdots \arrow[r] & F_{n+1} \arrow[r] \arrow[d, "h_{n+1}"'] & F_n \arrow[r] \arrow[d, "h_n"'] & F_{n-1} \arrow[r] \arrow[d, "h_{n-1}"'] & \cdots \\
\cdots \arrow[r] & G_{n+1} \arrow[r] & G_n \arrow[r] & G_{n-1} \arrow[r] & \cdots
\end{tikzcd}$$

This means that $h_n \delta^F_{n+1} = \delta^G_{n+1} h_{n+1}$ for all n.

Example 9.52. The zero and the identity maps of complexes $(F_\bullet, \delta_\bullet) \longrightarrow (F_\bullet, \delta_\bullet)$ are exactly what they sound like: the zero map $0_{F_\bullet}$ is 0 in every homological degree, and the identity map $1_{F_\bullet}$ is the identity in every homological degree.

This is the notion of morphism we would want to form a category of chain complexes.

Definition 9.53. Let R be a ring. The **category of chain complexes** of R-modules, denoted $\mathrm{Ch}(R-\mathbf{mod})$ or simply $\mathrm{Ch}(R)$, is the category with objects all chain complexes of R-modules and arrows all complex maps between them. When $R = \mathbb{Z}$, we write $\mathrm{Ch}(\mathbf{Ab})$ for $\mathrm{Ch}(\mathbb{Z})$, the category of chain complexes of abelian groups.

Exercise 24. Show that the isomorphisms in the category $\mathrm{Ch}(R)$ are precisely the maps of complexes

$$\begin{tikzcd} \cdots \arrow[r] & F_{n+1} \arrow[r] \arrow[d, "h_{n+1}"'] & F_n \arrow[r] \arrow[d, "h_n"'] & F_{n-1} \arrow[r] \arrow[d, "h_{n-1}"'] & \cdots \\ \cdots \arrow[r] & G_{n+1} \arrow[r] & G_n \arrow[r] & G_{n-1} \arrow[r] & \cdots \end{tikzcd}$$

such that h_n is an isomorphism for all n.

This is a good notion of map of complexes: it induces homomorphisms in homology.

Lemma 9.54. *Let $h\colon (F_\bullet, \delta^F_\bullet) \longrightarrow (G_\bullet, \delta^G_\bullet)$ be a map of complexes. For all n, h_n restricts to homomorphisms $B_n(h) : B_n(F_\bullet) \longrightarrow B_n(G_\bullet)$ and $Z_n(h) : Z_n(F_\bullet) \longrightarrow Z_n(G_\bullet)$. As a consequence, h induces homomorphisms on homology $\mathrm{H}_n(h) : \mathrm{H}_n(F_\bullet) \longrightarrow \mathrm{H}_n(G_\bullet)$.*

Proof. Since $h_n \delta^F_{n+1} = \delta^G_{n+1} h_{n+1}$, any element $a \in B_n(F_\bullet)$, say $a = \delta^F_{n+1}(b)$, is taken to

$$h_n(a) = h_n \delta^F_{n+1}(b) = \delta^G_{n+1} h_{n+1}(b) \in \operatorname{im} \delta^G_{n+1} = B_n(G_\bullet).$$

Similarly, if $a \in Z_n(F_\bullet) = \ker \delta^F_n$, then

$$\delta_n h_n(a) = h_{n-1} \delta^F_n(a) = 0,$$

so $h_n(a) \in \ker \delta^G_n = Z_n(G_\bullet)$. Finally, the restriction of h_n to $Z_n(F_\bullet) \longrightarrow Z_n(G_\bullet)$ sends $B_n(F_\bullet)$ into $B_n(G_\bullet)$, and thus it induces a well-defined homomorphism on the quotients $H_n(F_\bullet) \longrightarrow \mathrm{H}_n(G_\bullet)$. □

In particular, this says that taking nth homology is a functor $\mathrm{H}_n\colon \mathrm{Ch}(R) \longrightarrow R\text{-}\mathbf{mod}$, which takes each map of complexes $h\colon F_\bullet, \longrightarrow G_\bullet$ to the R-module homomorphism $\mathrm{H}_n(h) : \mathrm{H}_n(F_\bullet) \longrightarrow \mathrm{H}_n(G_\bullet)$.

Definition 9.55. A map of chain complexes h is a **quasi-isomorphism** if it induces an isomorphism in homology, meaning $\mathrm{H}_n(h)$ is an isomorphism of R-modules for all n.

Remark 9.56. Note that saying that if f is a quasi-isomorphism between F and G says more than just that $\mathrm{H}_n(F) \cong \mathrm{H}_n(G)$ for al n: it says that there are such isomorphisms that are all induced by f.

Exercise 25. Let π denote the projection map from $\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. The chain map

$$\begin{tikzcd}
\cdots \arrow[r] & 0 \arrow[r] \arrow[d, "0"'] & \mathbb{Z} \arrow[r, "2"] \arrow[d, "0"'] & \mathbb{Z} \arrow[r] \arrow[d, "\pi"] & 0 \arrow[r] \arrow[d, "0"'] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 0 \arrow[r] & \cdots
\end{tikzcd}$$

is a quasi-isomorphism.

Definition 9.57. Let $f, g \colon F \longrightarrow G$ be maps complexes. A **homotopy**, sometimes referred to as a **chain homotopy**, between f and g is a sequence of maps $h_n \colon F_n \longrightarrow G_{n+1}$

$$\begin{tikzcd}
\cdots \arrow[r, "\delta_{n+2}"] & F_{n+1} \arrow[r, "\delta_{n+1}"] \arrow[d, shift right, "g_{n+1}"'] \arrow[d, shift left, "f_{n+1}"] & F_n \arrow[r, "\delta_n"] \arrow[dl, "h_n"] \arrow[d, shift right, "g_n"'] \arrow[d, shift left, "f_n"] & F_{n-1} \arrow[r, "\delta_{n-1}"] \arrow[dl, "h_{n-1}"] \arrow[d, shift right, "g_{n-1}"'] \arrow[d, shift left, "f_{n-1}"] & \cdots \\
\cdots \arrow[r, "\delta_{n+2}"'] & G_{n+1} \arrow[r, "\delta_{n+1}"'] & G_n \arrow[r, "\delta_n"'] & G_{n-1} \arrow[r, "\delta_{n-1}"'] & \cdots
\end{tikzcd}$$

such that

$$\delta_{n+1} h_n + h_{n-1} \delta_n = f_n - g_n$$

for all n. If there exists a homotopy between f and g, we say that f and g are **homotopic**. If f is homotopic to the zero map, we say it is **null-homotopic**. If $f \colon (F_\bullet, \delta_\bullet^F) \longrightarrow (G_\bullet, \delta_\bullet^G)$ and $g \colon (G_\bullet, \delta_\bullet^G) \longrightarrow (F_\bullet, \delta_\bullet^F)$ are maps of complexes such that fg is homotopic to the identity map on $(G_\bullet, \delta_\bullet^G)$ and gf is homotopic to the identity chain map on $(F_\bullet, \delta_\bullet^F)$, we say that f and g are **homotopy equivalences** and $(F_\bullet, \delta_\bullet^F)$ and $(G_\bullet, \delta_\bullet^G)$ are **homotopy equivalent**.

Exercise 26. Homotopy is an equivalence relation.

This is an interesting relation because homotopic maps induce the same map on homology.

Lemma 9.58. *Homotopic maps of complexes induce the same map on homology.*

Proof. Let $f, g : (F_\bullet, \delta_\bullet^F) \longrightarrow (G_\bullet, \delta_\bullet^G)$ be homotopic maps of complexes, and let h be a homotopy between f and g. We claim that the map of complexes $f - g$ (defined in the obvious way) sends cycles to boundaries. If $a \in Z_n(F_\bullet)$, then

$$(f - g)_n(a) = \delta_{n+1} h_n + h_{n-1} \underbrace{\delta_n(a)}_{0} = \delta_{n+1}(h_n(a)) \in B_n(G_\bullet).$$

The map on homology induced by $f - g$ must then be the 0 map, so f and g induce the same map on homology. $\square$

Corollary 9.59. *Homotopy equivalences are quasi-isomorphisms.*

Proof. If $f \colon (F_\bullet, \delta_\bullet^F) \longrightarrow (G_\bullet, \delta_\bullet^G)$ and $g \colon (G_\bullet, \delta_\bullet^G) \longrightarrow (F_\bullet, \delta_\bullet^F)$ are such that fg is homotopic to $1_{G_\bullet}$ and gf is homotopic to $1_{F_\bullet}$, then by Lemma 9.58 fg induces the identity map on homology. Then for each n, $\mathrm{H}_n(f)\,\mathrm{H}_n(g) = \mathrm{H}_n(fg)$ is an isomorphism, and thus $\mathrm{H}_n(f)$ and $\mathrm{H}_n(g)$ must both be isomorphisms. $\square$

The converse is false.

Exercise 27. Let π denote the projection map from $\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. The chain map

$$\begin{tikzcd}
\cdots \arrow[r] & 0 \arrow[r] \arrow[d, "0"'] & \mathbb{Z} \arrow[r, "2"] \arrow[d, "0"'] & \mathbb{Z} \arrow[r] \arrow[d, "\pi"] & 0 \arrow[r] \arrow[d, "0"'] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 0 \arrow[r] & \cdots
\end{tikzcd}$$

is a quasi-isomorphism but not a homotopy equivalence.

Remark 9.60. In fact, the relation "there is a quasi-isomorphism from F to G" is not symmetric: in the example in Exercise 27, there is no quasi-isomorphism going in the opposite direction of the one given.

Now that we know about maps between complexes, it's time to point out that we can also talk about complexes and exact sequences of complexes. While we will later formalize this a little better when we discover that $\mathrm{Ch}(R)$ is an abelian category, let's for now give quick definitions that we can use.

Definition 9.61. Given complexes B and C, B is a **subcomplex** of C if B_n is a submodule of C_n for all n, and the inclusion maps $\iota_n\colon B_n \subseteq C_n$ define a map of complexes $\iota : B \longrightarrow C$. Given a subcomplex B of C, the **quotient** of C by B is the complex C/B that has C_n/B_n in homological degree n, with differential induced by the differential on C_n.

Exercise 28. If B is a subcomplex of C, then the differential d on C satisfies $d_n(B_n) \subseteq B_{n-1}$. Therefore, d_n induces a map of R-modules $C_n/B_n \longrightarrow C_{n-1}/B_{n-1}$ for all n, so that our definition of the differential on C/B actually makes sense.

We can also talk about kernels and cokernels of maps of complexes.

Definition 9.62. Given any map of complexes $f : B_\bullet \longrightarrow C_\bullet$, the **kernel** of f is the subcomplex $\ker f$ of $B_\bullet$ that we can assemble from the the kernels $\ker f_n$. More precisely, $\ker f$ is the complex

$$\cdots \longrightarrow \ker f_{n+1} \longrightarrow \ker f_n \longrightarrow \ker f_{n-1} \longrightarrow \cdots$$

where the differentials are simply the corresponding restrictions of the differentials on $B_\bullet$. Similarly, the **image** of f is the subcomplex of $C_\bullet$

$$\cdots \longrightarrow \operatorname{im} f_{n+1} \longrightarrow \operatorname{im} f_n \longrightarrow \operatorname{im} f_{n-1} \longrightarrow \cdots$$

where the differentials are given by restriction of the corresponding differentials in $C_\bullet$. The **cokernel** of f is the quotient complex $C_\bullet / \operatorname{im} f$.

Again, there are some details to check.

Exercise 29. Show that the kernel, image, and cokernel of a complex map are indeed complexes.

Definition 9.63. A **complex** in $\mathrm{Ch}(R)$ is a sequence of complexes of R-modules C^n and chain maps $d_n\colon C^n \longrightarrow C^{n-1}$ between them

$$\cdots \longrightarrow C^{n+1} \xrightarrow{d_{n+1}} C^n \xrightarrow{d_n} C^{n-1} \longrightarrow \cdots$$

such that $d_n d_{n+1} = 0$ for all n.

Given a complex C in $\mathrm{Ch}(R)$, we can talk about cycles and boundaries, which are a sequence of subcomplexes of the complexes in C, and thus its homology. Such a complex is exact if $\operatorname{im} d_{n+1} = \ker d_n$ for all n.

Definition 9.64. A **short exact sequence** of complexes is an exact complex in $\mathrm{Ch}(R)$ of the form

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0.$$

Equivalently, a short exact sequence of complexes is a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & A_{i+1} \arrow[r, "f_{i+1}"] \arrow[d, "\delta_{i+1}"'] & B_{i+1} \arrow[r, "g_{i+1}"] \arrow[d, "\delta_{i+1}"'] & C_{i+1} \arrow[r] \arrow[d, "\delta_{i+1}"'] & 0 \\
0 \arrow[r] & A_i \arrow[r, "f_i"] \arrow[d, "\delta_i"'] & B_i \arrow[r, "g_i"] \arrow[d, "\delta_i"'] & C_i \arrow[r] \arrow[d, "\delta_i"'] & \cdots \\
\cdots \arrow[r] & A_{i-1} \arrow[r, "f_{i-1}"] & B_{i-1} \arrow[r, "g_{i-1}"] & C_{i-1} \arrow[r] & \cdots
\end{tikzcd}$$

where the rows are exact and the columns are complexes.

9.4 Long exact sequences

A long exact sequence is just what it sounds like: an exact sequence that is, well, long. Long exact sequences arise naturally in various ways, and are often induced by some short exact sequence. The first long exact sequence one encounters is the long exact sequence on homology. All other long exact sequences are, in some way, a special case of this one. The main tool we need to build it is the Snake Lemma.

Theorem 9.65 (Snake Lemma)**.** *Consider the commutative diagram of R-modules*

$$\begin{tikzcd}
 & A' \arrow[r, "i'"] \arrow[d, "f"'] & B' \arrow[r, "p'"] \arrow[d, "g"] & C' \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & A \arrow[r, "i"'] & B \arrow[r, "p"'] & C &
\end{tikzcd}.$$

If the rows of the diagram are exact, then there exists an exact sequence

$$\ker f \longrightarrow \ker g \longrightarrow \ker h \xrightarrow{\partial} \operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h$$

Given $c' \in C'$, pick $b' \in B'$ such that $p'(b') = c'$, and $a \in A$ such that $i(a) = g(b')$. Then

$$\partial(c') = a + \operatorname{im} f \in \operatorname{coker} f.$$

The picture to keep in mind (and which explains the name of the lemma) is the following:

$$\begin{tikzcd}
\ker f \arrow[r] \arrow[d] & \ker g \arrow[r] \arrow[d] & \ker h \arrow[d] \arrow[dddll, rounded corners, "\partial"] \\
A' \arrow[r] \arrow[d, "f"'] & B' \arrow[r] \arrow[d, "g"'] & C' \arrow[d, "h"'] \\
A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[d] \\
\operatorname{coker} f \arrow[r] & \operatorname{coker} g \arrow[r] & \operatorname{coker} h
\end{tikzcd}$$

Definition 9.66. The map ∂ in the Snake Lemma is the **connecting homomorphism**.

Proof. If $a' \in \ker f$, then $b' := i'(a')$ must satisfy $g(b') = if(a') = 0$, by commutativity, so $b' \in \ker g$. Similarly, the image of $b' \in \ker g$ by p' is in the kernel of h. So the maps

$$\ker f \longrightarrow \ker g \longrightarrow \ker h$$

are restrictions of the maps $A' \xrightarrow{i'} B' \xrightarrow{p'} C'$, so $i'(\ker f) \subseteq \ker(B' \xrightarrow{p'} C')$. If $b' \in \ker g$ is such that $p'(b') = 0$, then there exists $a' \in A'$ such that $i'(a') = b'$; we only need to check that $a' \in \ker f$. An indeed, by commutativity we have

$$if(a') = gi'(a') = g(b') = 0,$$

and since i is injective, we must have $f(a') = 0$.

Similarly, if $a \in \operatorname{im} f$, the commutativity of the diagram guarantees that $i(a) \in \operatorname{im} g$, and if $b \in \operatorname{im} g$, then $p(b) \in \operatorname{im} h$. So the maps $A \xrightarrow{i} B \xrightarrow{p} C$ restrict to maps

$$\operatorname{im} f \longrightarrow \operatorname{im} g \longrightarrow \operatorname{im} h \ ,$$

which then induce maps

$$\operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h \ .$$

Again, we automatically get $i(\operatorname{coker} f) \subseteq \ker(\operatorname{coker} g \longrightarrow \operatorname{coker} h)$, so we only need to check equality. If $b \in B$ is such that $p(b) = 0$ in $\operatorname{coker} h$, meaning $p(b) \in \operatorname{im} h$, let $c' \in C$ be such that $h(c') = p(b)$. Since p is surjective, there exists $b' \in B'$ such that $p'(b') = c'$, and by commutativity,

$$pg(b') = hp'(b') = h(c') = p(b).$$

Then $b - g(b') \in \ker p = \operatorname{im} i$. Since $b = b - g(b')$ in $\operatorname{coker} g$, this shows that the class of b in $\operatorname{coker} g$ is in $i(\operatorname{coker} f)$. So we have shown exactness at $\ker g$ and $\operatorname{coker} g$.

So everything we need to prove concerns the connecting homomoprhism ∂. Our definition of ∂ can be visualized as follows:

$$
\begin{tikzcd}
 & & c' \in \ker h \arrow[d] & \\
A' \arrow[r, "i'"] \arrow[d, "f"'] & b' \in B' \arrow[r, dashed, mapsto, "p'"] \arrow[d, mapsto, "g"] & c' \in C' \arrow[d, "h"] & \\
a \in A \arrow[r, mapsto, "i"'] \arrow[d, mapsto] & g(b') \in B \arrow[r, "p"'] & 0 \in C \arrow[r, Rightarrow] & g(b') \in \ker p = \operatorname{im} i \\
a + \operatorname{im} f \in \operatorname{coker} f & & &
\end{tikzcd}
$$

We need to show the following:

1) ∂ is well-defined.
2) $p'(\ker g) = \ker \partial$.
3) $\operatorname{im} \partial = \ker(\, \operatorname{coker} f \xrightarrow{i} \operatorname{coker} g \,)$.

The last two points together say that the sequence

$$\ker g \longrightarrow \ker h \xrightarrow{\partial} \operatorname{coker} f \longrightarrow \operatorname{coker} g$$

is exact.

To show that ∂ is well-defined, let's fix some $c' \in \ker h \subseteq C'$. Since p' is surjective, $c' \in \operatorname{im} p'$. Consider $b_1', b_2' \in B'$ such that $p'(b_1') = p'(b_2') = c'$. Then $p'(b_1' - b_2') = 0$. We will show that our definition of $\delta(0)$ is independent of the choice of $b' \in \ker g$, which implies that our definition of $\delta(c')$ is independent of our choice of b_1' or b_2'. Given $b' \in \ker p' = \operatorname{im} i'$, there exists $a' \in A'$ such that $i'(a') = b'$. Notice that $a := f(a') \in A$ is such that

$$i(a) = if(a') = gi'(b')$$

so $\delta(0)$ is defined as $a + \operatorname{im} f \in \operatorname{coker} f$. Since $a = f(a') \in \operatorname{im} f$, we conclude that $\delta(0) = 0$ for any choice of b'. This shows that δ is well-defined, and 1) holds.

If $b' \in \ker g$, then $g(b') = 0$ and the only $a \in A$ such that $i(a) = g(b') = 0$ is $a = 0$. Therefore, $\delta(p'(b')) = 0$, so $p'(\ker g) \subseteq \ker \delta$. On the other hand, let $c' \in \ker h$ be such that $\partial(c') = 0$. That means that for any $b' \in B'$ such that $p'(b') = c'$ we must have $g(b') = i(a)$ for some $a \in \operatorname{im} f$. Let $a' \in A'$ be such that $f(a') = a$. Then

$$gi'(a') = if(a') = i(a) = g(b')$$

so $b' - i'(a') \in \ker g$. Since $p'i' = 0$, $p'(b' - i'(a')) = p'(b') = c'$, so $c' \in \operatorname{im} p'$. We conclude that $\ker \partial = p'(\ker g)$, and this shows 2).

Let $a \in A$. The statement $i(a + \operatorname{im} f) = 0$ lifts to B as $i(a) \in \operatorname{im} g$, so we can choose $b' \in B'$ such that $g(b') = i(a)$. Then $\partial(p'(b')) = a + \operatorname{im} f$, and moreover $p'(b')$ is in $\ker h$, since by commutativity we have $hp'(b') = pg(b') = pi(b') = 0$. This shows that $\ker(\, \operatorname{coker} f \xrightarrow{i} \operatorname{coker} g \,) \subseteq \operatorname{im} \partial$. Finally, if c', b', and a are as in the diagram above, $i(a + \operatorname{im} f) = g(b') + \operatorname{im} g = 0$, so $\operatorname{im} \partial \subseteq \ker(\, \operatorname{coker} f \xrightarrow{i} \operatorname{coker} g \,)$. This shows 3). $\square$

This proof is what we call a *diagram chase*, for reasons that may be obvious by now: we followed the diagram in the natural way, and everything worked out in the end.

Now that we have the Snake Lemma, we can construct the long exact sequence in homology:

Theorem 9.67 (Long exact sequence in homology)**.** *Given a short exact sequence in* $\mathrm{Ch}(R)$

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0,$$

there are connecting homomorphisms $\partial : \mathrm{H}_n(C) \longrightarrow \mathrm{H}_{n-1}(A)$ *such that*

$$\cdots \longrightarrow \mathrm{H}_{n+1}(C) \xrightarrow{\partial} \mathrm{H}_n(A) \xrightarrow{f} \mathrm{H}_n(B) \xrightarrow{g} \mathrm{H}_n(C) \xrightarrow{\partial} \mathrm{H}_{n-1}(A) \longrightarrow \cdots$$

is an exact sequence.

Proof. For each n, we have short exact sequences

$$0 \longrightarrow A_n \longrightarrow B_n \longrightarrow C_n \longrightarrow 0.$$

The condition that f and g are maps of complexes implies, by Lemma 9.54, that f and g both take boundaries to boundaries, so that we get exact sequences

$$A_n/\operatorname{im} d^A_{n+1} \longrightarrow B_n/\operatorname{im} d^B_{n+1} \longrightarrow C_n/\operatorname{im} d^C_{n+1} \longrightarrow 0\ .$$

Again by Lemma 9.54, the condition that f and g are maps of complexes also implies that f and g take cycles to cycles, so we get exact sequences

$$0 \longrightarrow Z_n(A) \longrightarrow Z_n(B) \longrightarrow Z_n(C)\ .$$

Let F be one of A, B, of C. The boundary maps on F induce maps $F_n \longrightarrow Z_{n-1}(F)$ that send $\operatorname{im} d_{n+1}$ to 0, so we get induced maps $F_n/\operatorname{im} d_{n+1} \longrightarrow Z_{n-1}(F)$. Putting all this together, we have a commutative diagram with exact rows

$$\begin{tikzcd}
 & A_n/\operatorname{im} d^A_{n+1} \arrow[r] \arrow[d, "d^A_n"'] & B_n/\operatorname{im} d^B_{n+1} \arrow[r] \arrow[d, "d^B_n"'] & C_n/\operatorname{im} d^C_{n+1} \arrow[r] \arrow[d, "d^C_n"'] & 0\ . \\
0 \arrow[r] & Z_n(A) \arrow[r] & Z_n(B) \arrow[r] & Z_n(C) &
\end{tikzcd}$$

For each $F = A, B, C$, the kernel of $F_n/\operatorname{im} d^F_{n+1} \xrightarrow{d^F_n} Z_{n-1}(F)$ is $\mathrm{H}_n(F)$, and its cokernel is $Z_{n-1}(F)/\operatorname{im} d^F_n = \mathrm{H}_{n-1}(F)$. The Snake Lemma now gives us exact sequences

$$\mathrm{H}_n(A) \longrightarrow \mathrm{H}_n(B) \longrightarrow \mathrm{H}_n(C) \xrightarrow{\partial} \mathrm{H}_{n-1}(A) \longrightarrow \mathrm{H}_{n-1}(B) \longrightarrow \mathrm{H}_{n-1}(C)$$

Finally, we glue all these together to obtain the long exact sequence in homology. $\square$

Remark 9.68. It's helpful to carefully consider how we compute the connecting homomorphisms in the long exact sequence, which we can easily put together from the proof of the Snake Lemma. Suppose that $c \in \ker d^C_{n+1}$. When we view c as an element in C_{n+1}, we can find $b \in B_{n+1}$ such that $g_{n+1}(b) = c$, since g_{n+1} is surjective by assumption. Since $d^B_{n+1}(b) \in \ker g_n$, we can find $a \in A_n$ with $f_n(a) = d^B_{n+1}(b)$. Finally, $\partial(c) = a + \operatorname{im} d^A_{n+1}$.

We will soon see that long exact sequences appear everywhere, and that they are very helpful. Before we see more examples, we want to highlight a connection between long and short exact sequences.

Remark 9.69. Suppose that

$$\cdots \longrightarrow C_{n+1} \xrightarrow{f_{n+1}} C_n \xrightarrow{f_n} \cdots$$

is a long exact sequence. This long exact sequence breaks into the short exact sequences

$$0 \longrightarrow \ker f_n \xrightarrow{i} C_n \xrightarrow{\pi} \operatorname{coker} f_{n+1} \longrightarrow 0\,.$$

The first map i is simply the inclusion of the submodule $\ker f_n$ into C_n, while the second map π is the canonical projection onto the quotient. While it is clear that i is injective and π is surjective, exactness at the middle is less obvious. This follows from the exactness of the original complex, which gives $\operatorname{im} i = \ker f_n = \operatorname{im} f_{n+1} = \ker \pi$.

The long exact sequence in homology is natural.

Theorem 9.70 (Naturality of the long exact sequence in homology)**.** *Any commutative diagram in* $\mathrm{Ch}(R)$

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "i"] \arrow[d, "f"'] & B \arrow[r, "p"] \arrow[d, "g"'] & C \arrow[r] \arrow[d, "h"'] & 0 \\
0 \arrow[r] & A' \arrow[r, "i'"'] & B' \arrow[r, "p'"'] & C' \arrow[r] & 0
\end{tikzcd}$$

with exact rows induces a commutative diagram where the rows are long exact sequences

$$\begin{tikzcd}
\cdots \arrow[r] & \mathrm{H}_{n+1}(C) \arrow[r, "\partial"] \arrow[d, "h"'] & \mathrm{H}_n(A) \arrow[r, "i"] \arrow[d, "f"'] & \mathrm{H}_n(B) \arrow[r, "p"] \arrow[d, "g"'] & \mathrm{H}_n(C) \arrow[r, "\partial"] \arrow[d, "h"'] & \mathrm{H}_{n-1}(A) \arrow[r] \arrow[d, "f"'] & \cdots \\
\cdots \arrow[r] & \mathrm{H}_{n+1}(C') \arrow[r, "\partial'"'] & \mathrm{H}_n(A') \arrow[r, "i'"'] & \mathrm{H}_n(B) \arrow[r, "p'"'] & \mathrm{H}_n(C) \arrow[r, "\partial'"'] & \mathrm{H}_{n-1}(A) \arrow[r] & \cdots
\end{tikzcd}$$

Proof. The rows of the resulting diagram are the long exact sequences in homology induced by each row of the original diagram, as in Theorem 9.67. So the content of the theorem is that the maps induced in homology by f, g, and h make the diagram commute. The commutativity of

$$\begin{tikzcd}
\mathrm{H}_n(A) \arrow[r, "i"] \arrow[d, "f"'] & \mathrm{H}_n(B) \arrow[r, "p"] \arrow[d, "g"'] & \mathrm{H}_n(C) \arrow[d, "h"'] \\
\mathrm{H}_n(A') \arrow[r, "i'"'] & \mathrm{H}_n(B) \arrow[r, "p'"'] & \mathrm{H}_n(C)
\end{tikzcd}$$

follows from the fact that H_n is a functor, so we only need to check commutativity of the square

$$\begin{tikzcd}
\mathrm{H}_n(C) \arrow[r, "\partial"] \arrow[d, "h"'] & \mathrm{H}_{n-1}(A) \arrow[d, "f"] \\
\mathrm{H}_n(C) \arrow[r, "\partial'"'] & \mathrm{H}_{n-1}(A)
\end{tikzcd}$$

that involves the connecting homomorphisms ∂ and ∂'. Consider the following commutative diagram:

$$\begin{tikzcd}
& 0 \arrow[rr] & & A_n \arrow[rr, "i"] \arrow[dl, "d"'] \arrow[dd, dashed] & & B_n \arrow[rr, "p"] \arrow[dl, "d"'] \arrow[dd, dashed] & & C_n \arrow[rr] \arrow[dl, "d"'] \arrow[dd, dashed] & & 0 \\
0 \arrow[rr] & & A_{n-1} \arrow[rr] \arrow[dd] & & B_{n-1} \arrow[rr] \arrow[dd] & & C_{n-1} \arrow[rr] \arrow[dd] & & 0 & \\
& 0 \arrow[rr] & & A'_n \arrow[rr] \arrow[dl, "d"'] & & B'_n \arrow[rr] \arrow[dl, "d"'] & & C'_n \arrow[rr] \arrow[dl, "d"'] & & 0 \\
0 \arrow[rr] & & A'_{n-1} \arrow[rr, "i'"'] & & B'_{n-1} \arrow[rr, "p'"'] & & C'_{n-1} \arrow[rr] & & 0 &
\end{tikzcd}$$

Given $c \in \ker(d_n \colon C_n \longrightarrow C_{n-1})$, we need to check that $f_{n-1}(\partial(c)) = \partial' h_n(c)$ in $\mathrm{H}_{n-1}(A)$. To compute $\partial(c)$, we find a lift $b \in B_n$ such that $p_n(b) = c$, and $a \in A_{n-1}$ with $i_{n-1}(a) = d_n(b)$, and set $\partial(c) = a + \operatorname{im} d_n \in \mathrm{H}_{n-1}(A)$. So $f_{n-1}\partial(c) = f_{n-1}(a) + \operatorname{im} d_n$. On the other hand, to compute $\partial' h_n(c)$, we start by finding $b' \in B'_n$ such that $p'_n(b') = h_n(c)$. By commutativity of the top square

$$\begin{tikzcd}
B_n \arrow[r, "p_n"] \arrow[d, "g_n"'] & C_n \arrow[d, "h_n"] \\
B'_n \arrow[r, "p'_n"'] & C'_n
\end{tikzcd}$$

we can choose $b' = g_n(b)$, since

$$p'_n(b') = p'_n g_n(b) = h_n p_n(b) = h_n(c).$$

Next we take $a' \in A'_{n-1}$ such that $i'_{n-1}(a') = d_n(b')$, and set $\partial'(h(c)) = a' + \operatorname{im} d_n \in \mathrm{H}_{n-1}(A')$. By commutativity of the middle square

$$\begin{tikzcd}
B_n \arrow[r, "d_n"] \arrow[d, "g_n"'] & B_{n-1} \arrow[d, "g_{n-1}"] \\
B'_n \arrow[r, "d_n"'] & B'_{n-1}
\end{tikzcd}$$

we have

$$d_n(b') = d_n g_n(b) = g_{n-1} d_n(b).$$

By our choice of a, we have

$$d_n(b') = g_{n-1}d_n(b) = g_{n-1}i_{n-1}(a),$$

and by commutativity of the front left square

$$\begin{tikzcd}
A_{n-1} \arrow[r, "i_{n-1}"] \arrow[d, "f_{n-1}"'] & B_{n-1} \arrow[d, "g_{n-1}"] \\
A'_{n-1} \arrow[r, "i'_{n-1}"'] & B'_{n-1}
\end{tikzcd}$$

we have

$$i'_{n-1}f_{n-1}(a) = g_{n-1}i_{n-1}(a) = d_n(b').$$

So we can take $a' = f_{n-1}(a)$. Finally, this means $\partial'(h_n(c)) = f_{n-1}(a) + \operatorname{im} d_{n-1}$, as we wanted to prove. $\square$

Remark 9.71. Let

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

be a short exact sequence in $\mathrm{Ch}(R)$. We can think of Theorem 9.70 as saying that the induced maps on homology $i_*\colon \mathrm{H}_n(A) \longrightarrow \mathrm{H}_n(B)$ and $p_*\colon \mathrm{H}_n(B) \longrightarrow \mathrm{H}_n(C)$ and the connecting homomorphism $\partial\colon \mathrm{H}_n(C) \longrightarrow \mathrm{H}_{n-1}(A)$ are all natural transformations. More precisely, consider the category **SES** of short exact sequences of R-modules, which is a full subcategory of $\mathrm{Ch}(R)$. Homology gives us functors $\mathbf{SES} \longrightarrow R\text{-}\mathbf{mod}$ that given a short exact sequence

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

return the R-modules $\mathrm{H}_n(A)$, $\mathrm{H}_n(B)$, or $\mathrm{H}_n(C)$). A map between two short exact sequences then induces R-module homomorphisms between the corresponding homologies. With this framework, Theorem 9.70 says that $i_*\colon \mathrm{H}_n(A) \longrightarrow \mathrm{H}_n(B)$, and $p_*\colon \mathrm{H}_n(B) \longrightarrow \mathrm{H}_n(C)$ and the connecting homomorphism $\partial : \mathrm{H}_n(C) \longrightarrow \mathrm{H}_{n-1}(A)$ are all natural transformations between the corresponding homology functors.

Chapter 10

R-mod

Before we study abelian categories in general, we want to understand our best prototype for what an abelian category looks like: the category R**-mod** of R-modules and R-module homomorphisms.

10.1 Hom

From now on, let's fix a ring R. Our goal is to get to know the category R**-mod**, which as we are about to discover is a very nice category. To make the notation less heavy, we write $\operatorname{Hom}_R(M,N)$ instead of $\operatorname{Hom}_{R\textbf{-mod}}(M,N)$ for the Hom-set between M and N in R**-mod**. The arrows in $\operatorname{Hom}_R(M,N)$ are all the R-module homomorphisms from M to N. This is a locally small category, meaning that the Hom-sets are actual sets, but more even is true: the Hom-sets are actually R-modules.

Given $f,g \in \operatorname{Hom}_R(M,N)$, $f+g$ is the R-module homomorphism defined by

$$(f+g)(m) := f(m) + g(m).$$

Given $r \in R$ and $f \in \operatorname{Hom}_R(M,N)$, $r \cdot f$ is the R-module homomorphism defined by

$$(r \cdot f)(m) := r \cdot f(m).$$

Exercise 30. Let M and N be R-modules. Then $\operatorname{Hom}_R(M,N)$ is an R-module.

Some Hom-sets can easily be identified with other well-understood modules.

Exercise 31. Let M be an R-module, and I an ideal in R. Then:

a) $\operatorname{Hom}_R(R,M) \cong M$.

b) $\operatorname{Hom}_R(R^n,M) \cong M^n$ for any $n \geqslant 1$.

c) $Hom_R(R/I,M) \cong (0 :_M I) := \{m \in M \mid Im = 0\}$.

Since R**-mod** is a locally small category, we saw in Definition 9.35 that there are two Hom-functors from R**-mod** to **Set**, the covariant functor $\operatorname{Hom}_R(M,-)\colon R\textbf{-mod} \longrightarrow \textbf{Set}$ and the contravariant functor $\operatorname{Hom}_R(-,N)\colon R\textbf{-mod} \longrightarrow \textbf{Set}$. In light of Exercise 30, we can upgrade these functors to land in R**-mod**, not just in **Set**. And they are indeed functors to R**-mod**, since they preserve identities and compositions.

Definition 10.1. Let R and S be rings. A functor $T\colon R\text{-}\mathbf{mod} \longrightarrow S\text{-}\mathbf{mod}$ is an **additive functor** if

$$T(f+g) = T(f) + T(g)$$

for all $f, g \in \mathrm{Hom}_R(M,N)$.

So an additive functor is one that restricts to a homomorphism of abelian groups for all Hom-sets.

Exercise 32. Show that $\mathrm{Hom}_R(M,-)$ and $\mathrm{Hom}_R(-,N)$ are both additive functors.

Lemma 10.2. *Let $T\colon R\text{-}\boldsymbol{mod} \longrightarrow S\text{-}\boldsymbol{mod}$ be an additive functor.*

a) If 0 is the 0 map between any two R-modules M and N, then $T(0) = 0$.

b) If 0 is the 0 R-module, $T(0) = 0$ is the zero S-module.

Proof.

a) As a function defined on each fixed $\mathrm{Hom}_R(M,N)$, T is a group homomorphism, so it must send 0 to 0.

b) An R-module M is the zero module if and only if the zero and identity maps on M coincide. Let N be the image of the zero R-module via T. On the one hand, any functor must send identity maps to identity maps, so the identity map on the zero module must be sent to the identity on N. On the other hand, we have shown that the zero map must be sent to the zero map on N, so the zero and identity maps on N must coincide, so $N = 0$. □

We want to study some of the properties of the Hom functor.[1] For starters, it behaves well with direct sums and products.

First, a note on direct sums versus (direct) products. When we have finitely many N_i, $\prod_i N_i = \bigoplus_i N_i = N_i^n$. The only difference is when we take infinitely many direct sums or products: the direct sum $\bigoplus_i N_i$ is the submodule of $\prod_i N_i$ whose elements, when written as tuples of elements in each N_i, have only finitely many nonzero entries. These two constructions are dual to each other, and have the following universal properties:

- Mapping *out* of a direct sum is completely determined by mapping out of each factor. More precisely, given homomorphisms $f_i : M_i \longrightarrow N$ for each i, then there exists a unique homomorphism $f\colon \bigoplus\limits_i M_i \longrightarrow N$ such that the following diagram commutes:

$$\begin{tikzcd} & N & \\ M_i \arrow[ur, "f_i"] \arrow[r, "\iota_i"'] & \bigoplus_i M_i \arrow[u, dashed, "f"] & M_j . \arrow[ul, "f_j"'] \arrow[l, "\iota_j"] \end{tikzcd}$$

[1]Indeed there are two of them, but they are so similar that we will sometimes refer to *the* Hom functor when talking about properties that are common to both of them.

- Mapping *in* to a product is completely determined by mapping in to each of the factors. More precisely, given homomorphisms $f_i : M \longrightarrow N_i$ for each i, there exists a unique homomorphism $f : M \longrightarrow \prod_i N_i$ making the following diagram commute:

$$\begin{tikzcd} & N \arrow[dl, "f_i"'] \arrow[d, dashed, "f"] \arrow[dr, "f_j"] & \\ M_i & \prod_i M_i \arrow[l, "\pi_i"] \arrow[r, "\pi_j"'] & M_j. \end{tikzcd}$$

In categorical language, this says that the product of R-modules is the product in R**-mod**, while the direct sum of modules is the coproduct in R**-mod**.

Given this, it should not be surprising that Hom behaves well when we map into products or out of corpoducts.

Theorem 10.3. *Let M, N, M_i, and N_i be R-modules. There are isomorphisms*

$$\operatorname{Hom}_R(M, \prod_i N_i) \cong \prod_i \operatorname{Hom}_R(M, N_i) \text{ and } \operatorname{Hom}_R(\bigoplus_i M_i, N) \cong \bigoplus_i \operatorname{Hom}_R(M_i, N).$$

Proof. For each i, let $\pi_i : \prod_j N_j \longrightarrow N_i$ be the canonical projection map. Consider the map

$$\begin{aligned} \operatorname{Hom}_R(M, \textstyle\prod_i N_i) &\xrightarrow{\alpha} \textstyle\prod_i \operatorname{Hom}_R(M, N_i) . \\ f &\longmapsto (\pi_i f) \end{aligned}$$

We claim this map is the desired isomorphism. First, take $(f_i)_i \in \prod_i \operatorname{Hom}_R(M, N_i)$. Define a map

$$\begin{aligned} M &\xrightarrow{\psi} \textstyle\prod_i N_i \quad . \\ m &\longmapsto (f_i(m)) \end{aligned}$$

This makes the diagram

$$\begin{tikzcd} & N_i & \\ \prod_j N_j \arrow[ur, "\pi_i"] & & M \arrow[ul, "f_i"'] \arrow[ll, dashed, "\psi"] \end{tikzcd}$$

commute, so that $\alpha(\psi) = (\pi_i \psi)_i = (f_i)$. This shows that α us surjective. Now suppose $f \; in \operatorname{Hom}_R(M, \prod_i N_i)$ is such that $\alpha(f) = 0$. For each $m \in M$, let $f(m) = (n_i)_i$, so $\pi_i f(m) = n_i$. By assumption, $(\pi_i f(m)) = 0$, which means that $\pi_i \alpha = 0$ for all i, and thus $n_i = 0$ for all i. So $f = 0$. We conclude that α is an isomorphism.

Now consider the map

$$\begin{aligned} \operatorname{Hom}_R(\bigoplus_i M_i, N) &\xrightarrow{\beta} \bigoplus_i \operatorname{Hom}_R(M_i, N) \\ f &\longmapsto (f \iota_i) \end{aligned}$$

where $\iota_j \colon M_j \longrightarrow \bigoplus_i M_i$ is the inclusion of the jth factor. Given $(f_i)_i \in \bigoplus_i \operatorname{Hom}_R(M_i, N)$, let

$$\begin{aligned} \bigoplus_i M_i &\xrightarrow{\quad\psi\quad} N \\ (m_i) &\longmapsto \sum_i f_i(m_i) \end{aligned}$$

Then $\beta(\psi) = (\psi\iota_i)_i$, so for each i and each $m_i \in M_i$, $\psi\iota_i(m_i) = f_i(m_i)$, and $\beta(\psi) = (f_i)_i$. This shows that β is surjective. Now assume $\beta(f) = 0$, which implies that $f\iota_i$ is the zero map for each i. Consider any $(m_i)_i \in \bigoplus_i M_i$. For each i, $f\iota_i(m_i) = 0$. On the other hand, $(m_i)_i = \sum_i \iota_i(m_i)$, so $f((m_i)_i) = \sum_i \iota_i(m_i) = 0$. We conclude that $f = 0$, and β is injective. $\square$

In particular,

$$\operatorname{Hom}_R(A \oplus B, C) \cong \operatorname{Hom}_R(A, C) \oplus \operatorname{Hom}_R(B, C)$$

and

$$\operatorname{Hom}_R(A, B \oplus C) \cong \operatorname{Hom}_R(A, B) \oplus \operatorname{Hom}_R(A, C).$$

Another important property of Hom is how it interacts with exact sequences.

Definition 10.4. A covariant additive functor $T \colon R\textbf{-mod} \longrightarrow S\textbf{-mod}$ is **left exact** if it takes every exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$$

of R-modules to the exact sequence

$$0 \longrightarrow T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C)$$

of S-modules, and **right exact** if it takes every exact sequence of R-modules

$$A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

to the exact sequence of S-modules

$$T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C) \longrightarrow 0\,.$$

Finally, T is an **exact functor** if it preserves short exact sequences, meaning every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is taken to the short exact sequence

$$0 \longrightarrow T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C) \longrightarrow 0\,.$$

A contravariant additive functor $T\colon R$**-mod** $\longrightarrow S$**-mod** is **left exact** if it takes every exact sequence

$$A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

of R-modules to the exact sequence

$$0 \longrightarrow T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A)$$

of S-modules, and **right exact** if it takes every exact sequence of R-modules

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$$

to the exact sequence of S-modules

$$T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A) \longrightarrow 0\,.$$

Finally, T is an **exact functor** if it preserves short exact sequences, meaning every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is taken to the short exact sequence

$$0 \longrightarrow T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A) \longrightarrow 0\,.$$

Exercise 33. The definitions above all stay unchanged if for each condition we start with a short exact sequence. For example, a covariant additive functor T is left exact if for every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

of R-modules,

$$0 \longrightarrow T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C)$$

is exact.

Remark 10.5. Left exact covariant functors take kernels to kernels, while right exact covariant functors take cokernels to cokernels. Similarly, left exact contravariant functors take kernels to cokernels, and right exact contravariant functors take cokernels to kernels.

Exactness is preserved by natural isomorphisms

Remark 10.6. Suppose that $F, G\colon R$**-mod** $\longrightarrow S$**-mod** are naturally isomorphic additive functors. We claim that F is exact if and only if G is exact. Let's prove it in the case when F and G are covariant. Given any short exact sequence

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

applying each of our functors yields complexes of R-modules which may or may not be exact. Our natural isomorphism gives us an isomorphism of complexes (displayed vertically)

$$\begin{tikzcd} 0 \arrow[r] & F(A) \arrow[r] \arrow[d] & F(B) \arrow[r] \arrow[d] & F(C) \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & G(A) \arrow[r] & G(B) \arrow[r] & G(C) \arrow[r] & 0. \end{tikzcd}$$

Isomorphisms of complexes induce isomorphisms in homology, so the top sequence is exact if and only if the bottom sequence is exact.

The same argument shows that F is left (respectively, right) exact if and only if G is left (respectively, right) exact.

Hom is left exact.

Theorem 10.7. *Let M be an R-module.*

a) *The covariant functor $\operatorname{Hom}_R(M, -)$ is left exact: for every exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$$

of R-modules, the sequence

$$0 \longrightarrow \operatorname{Hom}_R(M, A) \xrightarrow{\operatorname{Hom}_R(M,f)} \operatorname{Hom}_R(M, B) \xrightarrow{\operatorname{Hom}_R(M,g)} \operatorname{Hom}_R(M, C)$$

is exact.

b) *The contravariant functor $\operatorname{Hom}_R(-, M)$ is left exact: for every exact sequence*

$$A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

of R-modules, the sequence

$$0 \longrightarrow \operatorname{Hom}_R(C, M) \xrightarrow{\operatorname{Hom}_R(g,M)} \operatorname{Hom}_R(B, M) \xrightarrow{\operatorname{Hom}_R(f,M)} \operatorname{Hom}_R(A, M)$$

is exact.

Proof. To make the notation less heavy, we will write $f_* := \operatorname{Hom}_R(M, f)$, $g_* := \operatorname{Hom}_R(M, g)$, $f^* := \operatorname{Hom}_R(f, M)$, and $g^* := \operatorname{Hom}_R(g, M)$.

a) We have three things to show:

- f_* is injective.
  Suppose that $h \in \operatorname{Hom}_R(M, A)$ is such that $f_*(h) = 0$. By definition, this means that $fh = 0$. But f is injective, so

  $$fh(a) = 0 \implies h(a) = 0.$$

  We conclude that $h = 0$, and f_* is injective.

- $\operatorname{im} f_* \subseteq \ker g_*$.
  Let $h \in \operatorname{Hom}_R(M, A)$. Then $g_* f_*(h) = gfh$, but $gf = 0$ by assumption, so $g_* f_*(h) = gfh = 0$.
- $\ker g_* \subseteq \operatorname{im} f_*$.
  Let $h \in \operatorname{Hom}_R(M, B)$ be in the kernel of g_*. Then $gh = g_*(h) = 0$, so for each $m \in M$, $gh(m) = 0$. Then $h(m) \in \ker g = \operatorname{im} f$, so there exists $a \in A$ such that $f(a) = h(m)$. Since f is injective, this element a is unique for each $m \in M$. So setting $k(m) := a$ gives us a well-defined function $k \colon M \longrightarrow A$. We claim that k is in fact an R-module homomorphism. To see that, notice that if $k(m_1) = a_1$ and $k(m_2) = a_2$, then
  $$f(a_1 + a_2) = f(a_1) + f(a_2) = h(m_1) + h(m_2) = h(m_1 + m_2),$$
  so that $k(m_1 + m_2) = a_1 + a_2 = k(m_1) + k(m_2)$. Similarly, given any $r \in R$,
  $$f(ra_1) = rf(a_1) = rh(m_1) = h(rm_1),$$
  so $k(rm_1) = ra_1 = rk(m_1)$. Finally, this element $k \in \operatorname{Hom}_R(M, A)$ satisfies
  $$f_*(k)(m) = f(k(m)) = h(m)$$
  for all $m \in M$, so $f_*(k) = h$ and $h \in \operatorname{im} f_*$.

b) Again, we have three things to show:

- g^* is injective.
  If $g^*(h) = 0$ for some $h \in \operatorname{Hom}_R(C, M)$, then $hg = g^*(h) = 0$. Consider any $c \in C$. Since g is surjective, there exists $b \in B$ such that $g(b) = c$. Then $h(c) = hg(b) = 0$, so h must be the zero map.
- $\operatorname{im} g^* \subseteq \ker f^*$.
  Let $h \in \operatorname{Hom}_R(B, M)$ be in $\operatorname{im} g^*$, so that there exists $k \in \operatorname{Hom}_R(C, M)$ such that $kg = g^*(k) = h$. Then
  $$f^*(h) = hf = k\underbrace{gf}_{0} = 0,$$
  so $h \in \ker f^*$.
- $\ker f^* \subseteq \operatorname{im} g^*$.
  Let $h \in \operatorname{Hom}_R(B, M)$ be in $\ker f^*$, so that $hf = 0$. Given any $c \in C$, there exists $b \in B$ such that $g(b) = c$, since g is surjective. Let $k \colon C \longrightarrow M$ be the function defined by $k(c) := h(b)$ for some b with $g(b) = c$. This function is well-defined, since whenever $g(b') = g(b) = c$, $b - b' \in \ker g = \operatorname{im} f$, say $b - b' = f(a)$, and thus $h(b - b') = h(f(a)) = 0$. Moreover, we claim that k is indeed a homomorphism of R-modules. If $c_1, c_2 \in C$, and $g(b_1) = c_1$, $g(b_2) = c_2$, then $g(b_1 + b_2) = c_1 + c_2$, so
  $$k(c_1 + c_2) = h(b_1 + b_2) = h(b_1) + h(b_2) = k(b_1) + k(b_2).$$
  Finally, this element $k \in \operatorname{Hom}_R(C, M)$ is such that $g^*(k)$ satisfies
  $$(g_*(k))(b) = k(g(b)) = h(b)$$
  for all $b \in B$, so $g^*(k) = h$, and $h \in \operatorname{im} g^*$. □

So $\mathrm{Hom}_R(M, -)$ preserves kernels, and $\mathrm{Hom}_R(-, N)$ sends kernels to cokernels. However, Hom is *not* right exact in general.

Example 10.8. Consider the short exact sequence of abelian groups

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Q} \longrightarrow \mathbb{Q}/\mathbb{Z} \longrightarrow 0,$$

where the first map is the inclusion of $\mathbb{Z}$ into $\mathbb{Q}$, and the second map is the canonical projection. The elements in the abelian group $\mathbb{Q}/\mathbb{Z}$ are cosets of the form $\frac{p}{q} + \mathbb{Z}$, where $\frac{p}{q} \in \mathbb{Q}$, and whenever $\frac{p}{q} \in \mathbb{Z}$, $\frac{p}{q} + \mathbb{Z} = 0$. While Theorem 10.7 says that

$$0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}/\mathbb{Z})$$

is exact, we claim that this cannot be extended to a short exact sequence, since the map $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}/\mathbb{Z})$ is not surjective. On the one hand, there are no non-trivial homomorphisms from $\mathbb{Z}/2$ to either $\mathbb{Z}$ nor $\mathbb{Q}$, since there are no elements in $\mathbb{Z}$ nor $\mathbb{Q}$ of order 2. This shows that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}) \cong\cong 0$. On the other hand, $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}/\mathbb{Z})$ is nonzero, since $\frac{1}{2} + \mathbb{Z}$ is an element of order 2 in $\mathbb{Q}/\mathbb{Z}$, so the map sending 1 in $\mathbb{Z}/2$ to $\frac{1}{2} + \mathbb{Z}$ in $\mathbb{Z}/\mathbb{Q}$ is nonzero. So after applying $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, -)$, we get the exact sequence

$$0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, \mathbb{Q}/\mathbb{Z}) \; .$$

So this shows that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2, -)$ is not an exact functor, only left exact.

Similarly, we can show that $\mathrm{Hom}_{\mathbb{Z}}(-, \mathbb{Z})$ is not exact by applying it to the same original short exact sequence. This time, Theorem 10.7 says that

$$0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Q}/\mathbb{Z}, \mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}) \; .$$

is exact. We claim that the last map is not surjective. By Exercise 31, $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}$. On the other hand, we claim that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Z}) = 0$. Indeed, if $f \colon \mathbb{Q} \longrightarrow \mathbb{Z}$ is a homomorphism of abelian groups, then for all $n \geqslant 1$ we have $f(1) = nf(\frac{1}{n})$. This says that $f(1)$ is an integer that is divisible by every integer, which is impossible unless $f(1) = 0$. We conclude that $f = 0$.

10.2 Tensor products

Definition 10.9. Let M, N, and L be R-modules. A function $f \colon M \times N \longrightarrow L$ is R-**bilinear** if for all $m, m' \in M$, all $n, n' \in N$, and all $r \in R$ we have

- $f(m + m', n) = f(m, n) + f(m', n)$
- $f(m, n + n') = f(m, n) + f(m, n')$
- $f(rm, n) = f(m, rn) = rf(m, n)$.

Example 10.10. The product on R is a bilinear function $R \times R \longrightarrow R$.

Definition 10.11. Let M and N be R-modules. The **tensor product** of M and N is an R-module $M \otimes_R N$ together with an R-bilinear map $\tau\colon M \times N \longrightarrow M \otimes_R N$ with the following universal property: for every R-module A and every R-bilinear map $f\colon M \times N \longrightarrow A$ there exists a unique R-module homomorphism $\tilde{f}\colon M \otimes_R N \longrightarrow A$ such that the following diagram commutes:

$$\begin{tikzcd} M \otimes_R N \arrow[dr, dashed, "\tilde{f}"] & \\ M \times N \arrow[u, "\tau"] \arrow[r, "f"'] & A \end{tikzcd}$$

Remark 10.12. We can express this universal property in the framework of Definition 9.42. Consider the functor $\mathrm{Bilin}(M \times N, -)\colon R\text{-}\mathbf{mod} \longrightarrow \mathbf{Set}$ that sends an R-module A to the set of R-bilinear maps $M \times N \longrightarrow A$, and a map of R-modules $f\, A \longrightarrow B$ to the function of sets induced by post-composition of functions. The universal property of the tensor product is encoded in the representable functor $\mathrm{Bilin}(M \times N, -) : R\text{-}\mathbf{mod} \longrightarrow \mathbf{Set}$ together with the bilinear map $\tau \in \mathrm{Bilin}(M \times N, M \otimes_R N)$. Indeed, this says that τ induces a natural isomorphism between $\mathrm{Hom}_R(M \otimes_R N, -)$ and $\mathrm{Bilin}(M \times N, -)$ by sending each R-module A to the bijection

$$\begin{aligned} \mathrm{Hom}_R(M \otimes_R N, A) &\longrightarrow \mathrm{Bilin}(M \times N, A) \\ f &\longmapsto \mathrm{Bilin}(M \times N, f)\tau = f_*(\tau) = f\tau. \end{aligned}$$

The fact that this is a bijection says that for every R-bilinear map g there exists a unique R-module homomorphism f such that

$$\begin{tikzcd} M \otimes_R N \arrow[dr, dashed, "f"] & \\ M \times N \arrow[u, "\tau"] \arrow[r, "g"'] & A \end{tikzcd}$$

commutes. So this is indeed the universal property we described before.

Tensor products exist.

Theorem 10.13. *Given any two R-modules M and N, their tensor product exists.*

Proof. Let F be the free R-module on the set $M \times N$, meaning that F has a basis element for each (m,n) with $m \in M$ and $n \in N$. In what follows, we identify (m,n) with the corresponding basis element for F. Let S be the submodule of F generated by

$$S = \left(\left\{ \begin{array}{c|c} (m, n+n') - (m,n) - (m,n') & m, m' \in M \\ (m+m', n) - (m,n) - (m',n) & n, n' \in N \\ (rm, n) - r(m,n), (m, rn) - r(m,n) & r \in R \end{array} \right\} \right).$$

Let $M \otimes_R N := F/S$, and let $m \otimes n$ denote the class of (m,n) in the quotient. We claim that this module $M \otimes_R N$ is a tensor product for M and N, together with the map

$$\begin{aligned} M \times N &\xrightarrow{\tau} M \otimes N \\ (m,n) &\longmapsto m \otimes n \end{aligned}$$

Notice τ is the restriction of the quotient map $F \longrightarrow F/S$ to the basis elements of F. Moreover, by construction of $M \otimes_R N$, the following identities hold:

$$\begin{aligned} m \otimes (n+n') &= m \otimes n + m \otimes n' \\ (m+m') \otimes n &= m \otimes n + m \otimes n' \\ (rm) \otimes n &= m \otimes (rn) \end{aligned}$$

Together, these make τ an R-bilinear map. The map $M \times N \longrightarrow F$ that sends each pair (m, n) to the corresponding basis element is R-bilinear by construction. Moreover, there is a natural quotient map $F \longrightarrow M \otimes_R N$, which, and these maps make the diagram

$$\begin{tikzcd} M \times N \arrow[rr, "\tau"] \arrow[dr, "i"'] & & M \otimes_R N \\ & F \arrow[ur, two heads] & \end{tikzcd}$$

commute.

Now suppose that A is any other R-module, and consider any R-bilinear map $M \times N \xrightarrow{f} A$. Since F is the free R-module on $M \times N$, f induces a homomorphism of R-modules $\varphi \colon F \longrightarrow A$ such that $fi = \varphi$, meaning $f(m, n) = \varphi(m, n)$ for all $m \in M$ and all $n \in N$. Finally, the fact that f is bilinear implies that $S \subseteq \ker \varphi$. Therefore, φ induces an R-module homomorphism on $F/S = M \otimes_R N$. All this fits in the following commutative diagram:

$$\begin{tikzcd} M \times N \arrow[rr, "\tau"] \arrow[dr, "i"] \arrow[ddr, "f"'] & & M \otimes_R N \arrow[ddl, dashed, "\tilde{f}"] \\ & F \arrow[ur, two heads] \arrow[d, dashed, "\varphi"'] & \\ & A & \end{tikzcd}$$

Finally, this map $\tilde{f}$ we constructed agrees satisfies $\tilde{f}(n \otimes n) = f(m, n)$, and since $M \otimes_R N$ is generated by such elements, $\tilde{f}$ is completely determined by the images of $m \otimes n$, and thus unique. $\square$

The construction in Theorem 10.13 gives us generators $m \otimes n$ for $M \otimes_R N$. These are usually called **simple tensors**. So any element in $M \otimes_R N$ is of the form

$$\sum_{i=1}^{k} m_i \otimes n_i.$$

Such expressions are *n*ot unique. For a cheap example, consider the relations we used to construct $M \otimes_R N$ from the free R-module on $M \times N$, which gives us nontrivial ways to write the 0 element in $M \otimes_R N$:

$$\begin{aligned} 0 &= m \otimes (n+n') - m \otimes n - m \otimes n' \\ 0 &= (m+m') \otimes n - m \otimes n - m \otimes n' \\ 0 &= (rm) \otimes n - m \otimes (rn). \end{aligned}$$

This makes things unexpectedly tricky. For example, a particular tensor product might unexpectedly be the zero module. Also, whenever we try to define some R-module homomorphism from $M \otimes_R N$ into some other R-module, we must carefully check that our map is well-defined, which is in principle not an easy task. Therefore, the easiest way to define some R-module homomorphism from $M \otimes_R N$ is to give some R-bilinear map from $M \times N$ into our desired R-module.

Before we get in too deep with such details, we point out that we can talk about *the* tensor product of two R-modules.

Lemma 10.14. *The tensor product of M and N is unique up to unique isomorphism. More precisely, if $M \times N \xrightarrow{\tau_1} T_1$ and $M \times N \xrightarrow{\tau_2} T_2$ are two tensor products, then there exists a unique isomorphism $T_1 \xrightarrow{i} T_2$ such that*

$$\begin{tikzcd}
& T_1 \arrow[dd, dashed, "i"] \\
M \times N \arrow[ur, "\tau_1"] \arrow[dr, "\tau_2"'] & \\
& T_2
\end{tikzcd}$$

Proof. First, note that the universal property of the tensor product implies that there exists a unique φ such that

$$\begin{tikzcd}
T_i \arrow[dr, dashed, "\varphi"] & \\
M \times N \arrow[u, "\tau_i"] \arrow[r, "\tau_i"'] & T_i
\end{tikzcd}$$

commutes. Since the identity map $T_i \longrightarrow T_i$ *is* such a map, it must be the *only* such map.

Similarly, there are unique maps $\varphi_1 \colon T_1 \longrightarrow T_2$ and $\varphi_2 \colon T_2 \longrightarrow T_1$ such that

$$\begin{tikzcd}
T_1 \arrow[dr, dashed, "\varphi_1"] & \\
M \times N \arrow[u, "\tau_1"] \arrow[r, "\tau_2"'] & T_2
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
T_2 \arrow[dr, dashed, "\varphi_2"] & \\
M \times N \arrow[u, "\tau_2"] \arrow[r, "\tau_1"'] & T_1
\end{tikzcd}$$

both commute. Stacking these up, we get commutative diagrams

$$\begin{tikzcd}
T_1 \arrow[dr, dashed, "\varphi_1"] & & \\
& T_2 \arrow[dr, dashed, "\varphi_2"] & \\
M \times N \arrow[uu, "\tau_1"] \arrow[ur, "\tau_2"] \arrow[rr, "\tau_1"'] & & T_1
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
T_2 \arrow[dr, dashed, "\varphi_2"] & & \\
& T_1 \arrow[dr, dashed, "\varphi_1"] & \\
M \times N \arrow[uu, "\tau_2"] \arrow[ur, "\tau_1"] \arrow[rr, "\tau_2"'] & & T_2
\end{tikzcd}$$

so that $\varphi_2\varphi_1$ must be the identity on T_1 and $\varphi_1\varphi_2$ must be the identity on T_2. In particular, T_1 and T_2 are isomorphic, and the isomorphisms φ_1 and φ_2 are unique. $\square$

From now on we talk about *the* tensor product $M \otimes_R N$ of M and N, which is generated by the simple tensors $m \otimes n$. It's also important to remember (though we're all bound to forget once or twice) that *not* all elements in $M \otimes_R N$ are simple tensors, and that even though M and N are nonzero, $M \otimes_R N$ could very well be zero.

So let's study some properties of and get some practice with tensor products.

Remark 10.15. Two R-module maps $M \otimes_R N \longrightarrow L$ coincide if and only if they agree on simple tensors, since these are generators for $M \otimes_R N$.

Lemma 10.16. *Let $f\colon A \longrightarrow C$, $g\colon B \longrightarrow D$ be R-module homomorphisms. There exists a unique homomorphism of R-modules $f \otimes f\colon A \otimes_R B \longrightarrow C \otimes_R D$ such that*

$$(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$$

for all $a \in A$ and $b \in B$.

Proof sketch. The function

$$\begin{aligned} A \times B &\longrightarrow C \otimes_R D \\ (a,b) &\longmapsto f(a) \otimes g(b) \end{aligned}$$

is R-bilinear, so the universal property of tensor products gives the desired R-module homomorphism, which is unique. $\square$

Lemma 10.17. *Given R-module maps $A_1 \xrightarrow{f_1} A_2 \xrightarrow{f_2} A_3$ and $B_1 \xrightarrow{b_1} B_2 \xrightarrow{g_2} B_3$, the composition of $f_1 \otimes g_1$ satisfies $f_2 \otimes g_2$*

$$(f_2 \otimes g_2) \circ (f_1 \otimes g_1) = (f_2 f_1) \otimes (g_2 g_1).$$

Proof. It's sufficient to check that these maps agree on simple tensors, and indeed they both take $a \otimes b$ to $(f_2 f_1(a)) \otimes (g_2 g_1(b))$. $\square$

We are particularly interested in tensor products because of the tensor functor.

Theorem 10.18. *Let M be an R-module. There is an additive covariant functor*

$$M \otimes_R -\colon R\text{-}\boldsymbol{mod} \longrightarrow R\text{-}\boldsymbol{mod}$$

that takes each R-module N to $M \otimes_R N$, and each R-module homomorphism $f\colon A \longrightarrow B$ to the R-module homomorphism $1_M \otimes f\colon M \otimes_R A \longrightarrow M \otimes_R B$.

Similarly, there is an additive covariant functor

$$- \otimes_R N\colon R\text{-}\boldsymbol{mod} \longrightarrow R\text{-}\boldsymbol{mod}$$

that takes each R-module M to $M \otimes RN$, and each R-module homomorphism $f\colon A \longrightarrow B$ to the R-module homomorphism $f \otimes 1_N\colon A \otimes_R N \longrightarrow B \otimes_R M$.

Proof. We do the case of $T := M \otimes_R -$, and leave $- \otimes_R N$ as an exercise.

First, note that T preserves identities, meaning $T(1_N) = 1_{T(N)}$, since the identity map on $M \otimes_R N$ agrees with $T(1_N) = 1_M \otimes 1_N$ on simple tensors. Moreover, T preserves compositions, since by Lemma 10.17 we have

$$T(f)T(g) = (1 \otimes f)(1 \otimes g) = 1 \otimes (fg) = T(fg).$$

Thefore, T is a functor. To check that it is an additive functor, we need to prove that $T(f+g) = T(f) + T(g)$ for all $f, g \in \operatorname{Hom}_R(A,B)$. Again, the maps $T(f+g) = 1 \otimes (f+g)$ and $T(f) + T(g) = 1 \otimes f + 1 \otimes g$ agree on simple tensors, and so they are equal. □

Lemma 10.19. *If f and g are isomorphisms of R-modules, then $f \otimes g$ is an isomorphism.*

Proof. On the one hand, $1 \otimes_R g$ is the image of an isomorphism by a functor, and thus an isomorphism. Similarly, $f \otimes_R 1$ must be an isomorphism. Finally, $f \otimes g = (f \otimes 1)(1 \otimes g)$, by Lemma 10.17, which is a composition of isomorphisms, and thus an isomorphism. □

We can now prove some useful properties of tensor products.

Lemma 10.20 (Commutativity of tensor products)**.** *Let M and N be R-modules. Then $M \otimes_R N \cong N \otimes_R M$, and this isomorphism is natural.*

Proof sketch. The map $M \times N \longrightarrow N \otimes_R M$ given by $(m,n) \mapsto n \otimes m$ is R-bilinear. The universal property of the tensor product $M \otimes_R N$ gives us an R-module homomorphism φ such that the diagram

$$\begin{tikzcd}
& M \otimes_R N \arrow[dr, dashed, "\varphi"] & \\
M \times N \arrow[ur] \arrow[rr] & & N \otimes_R M \\
(m,n) \arrow[rr, mapsto] & & n \otimes m
\end{tikzcd}.$$

commutes. Similarly, we get a map ψ and a commutative diagram

$$\begin{tikzcd}
& N \otimes_R M \arrow[dr, dashed, "\psi"] & \\
N \times M \arrow[ur] \arrow[rr] & & M \otimes_R N \\
(m,n) \arrow[rr, mapsto] & & n \otimes m
\end{tikzcd}.$$

Then $\varphi\psi$ agrees with the identity on $N \otimes_R M$ on simple tensors, so it is the identity. Similarly, $\psi\varphi$ is the identity on $M \otimes_R N$, and these are the desired isomorphisms.

The statement about naturality is more precisely the following: for every R-module maps $f : M_1 \longrightarrow M_2$ and $g : N_1 \longrightarrow N_2$, our isomorphism maps $M_1 \otimes_R N_1 \cong N_1 \otimes_R M_1$ and $M_2 \otimes_R N_2 \cong N_2 \otimes_R M_2$ make the diagram

$$\begin{tikzcd}
M_1 \otimes_R N_1 \arrow[r, "\cong"] \arrow[d, "f \otimes g"'] & N_1 \otimes_R M_1 \arrow[d, "g \otimes f"] \\
M_2 \otimes_R N_2 \arrow[r, "\cong"] & N_2 \otimes_R M_2
\end{tikzcd}$$

commute. To check this, it's sufficient to check commutativity on simple tensors, and indeed

$$\begin{tikzcd}
m \otimes n \arrow[rrr, mapsto] \arrow[ddd, mapsto] & & & n \otimes m \arrow[ddd, mapsto] \\
& M_1 \otimes_R N_1 \arrow[r, "\cong"] \arrow[d, "f\otimes g"'] & N_1 \otimes_R M_1 \arrow[d, "g\otimes f"] & \\
& M_2 \otimes_R N_2 \arrow[r, "\cong"] & N_2 \otimes_R M_2 & \\
f(m) \otimes g(n) \arrow[rrr, mapsto] & & & g(n) \otimes f(m).
\end{tikzcd}$$

Lemma 10.21 (Associativity of tensors)**.** *Let A, B, and C be R-modules. Then*

$$(A \otimes_R B) \otimes_R C \cong A \otimes_R (B \otimes_R C).$$

Proof. Fix $c \in C$. The map

$$\begin{aligned} A \times B &\longrightarrow A \otimes_R (B \otimes_R C) \\ (a, b) &\longmapsto a \otimes (b \otimes c) \end{aligned}$$

is R-bilinear, so it induces a homomorphism of R-modules $\varphi_c \colon A \otimes_R B \longrightarrow A \otimes_R (B \otimes_R C)$. Then

$$\begin{aligned} (A \otimes_R B) \times C &\longrightarrow A \otimes_R (B \otimes_R C) \\ (a \otimes b, c) &\longmapsto a \otimes (b \otimes c) \end{aligned}$$

is also R-bilinear, and it induces a homomorphism of R-modules that sends $(a \otimes b) \otimes c$ to $a \otimes (b \otimes c)$. Similarly, we can define a homomorphism of R-modules $A \otimes_R (B \otimes_R C) \longrightarrow (A \otimes_R B) \otimes_R C$ that sends $a \otimes (b \otimes c)$ to $(a \otimes b) \otimes c$. The composition of these two homomorphisms of R-modules in either order is the identity on simple tensors, and thus they are both isomorphisms. $\square$

Lemma 10.22. *There is a natural isomorphism between $R \otimes_R -$ and the identity functor on R-**mod**. In particular, $R \otimes_R M \cong M$ for every R-module M.*

Proof. The R-bilinear map

$$\begin{aligned} R \times M &\longrightarrow M \\ (r, m) &\longmapsto rm \end{aligned}$$

induces a homomorphism of R-modules $R \otimes_R M \xrightarrow{\varphi_M} M$. By definition, φ_M is surjective. Moreover, the map

$$\begin{aligned} M &\xrightarrow{f_M} R \otimes_R M \\ m &\longmapsto 1 \otimes m \end{aligned}$$

is a homomorphism of R-modules, since

$$f_M(a + b) = 1 \otimes (a + b) = 1 \otimes a + 1 \otimes b \text{ and } f_M(ra) = 1 \otimes (ra) = r(1 \otimes a) = r f_M(a).$$

For every $m \in M$, $\varphi_M f_M(m) = \varphi_M(1 \otimes m) = 1m = m$, and for every simple tensor, $f_M \varphi_M(r \otimes m) = f_M(rm) = 1 \otimes (rm) = r \otimes m$. This shows that φ_M is an isomorphism.

Finally, given any $f \in \mathrm{Hom}_R(M, N)$, since f is R-linear we conclude that the diagram

$$\begin{tikzcd}
r \otimes m \arrow[rrr, mapsto] \arrow[ddd, mapsto] & & & rm \arrow[ddd, mapsto] \\
& R \otimes_R M \arrow[r, "\varphi_M"] \arrow[d, "1\otimes f"'] & M \arrow[d, "f"] & \\
& R \otimes N \arrow[r, "\varphi_N"'] & N & \\
r \otimes f(m) \arrow[rrr, mapsto] & & & rf(m) = f(rm)
\end{tikzcd}$$

commutes, so our isomorphism is natural. $\square$

Example 10.23. We claim that $\mathbb{Z}/(2) \otimes_{\mathbb{Z}} \mathbb{Q} = 0$, despite the fact that both of these $\mathbb{Z}$-modules are nonzero. To see that, simply note that given any $a \in \mathbb{Z}/2$ and any $p \in \mathbb{Q}$,

$$a \otimes p = a \otimes \frac{2p}{2} = (2a) \otimes \frac{p}{2} = 0 \otimes \frac{p}{2} = 0.$$

Since $\mathbb{Z}/(2) \otimes_{\mathbb{Z}} \mathbb{Q}$ is generated by simple tensors, which are all 0, we conclude that $\mathbb{Z}/(2) \otimes_{\mathbb{Z}} \mathbb{Q} = 0$.

Example 10.24. Consider the abelian group $\mathbb{Q}/\mathbb{Z}$. Again, this is very much nonzero, and yet we claim that $\mathbb{Q}/\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Q}/\mathbb{Z} = 0$. For any simple tensor,

$$\begin{aligned}
\left(\frac{p}{q} + \mathbb{Z}\right) \otimes \left(\frac{a}{b} + \mathbb{Z}\right) = \left(\frac{bp}{bq} + \mathbb{Z}\right) \otimes \left(\frac{a}{b} + \mathbb{Z}\right) &= \left(\frac{p}{bq} + \mathbb{Z}\right) \otimes b\left(\frac{a}{b} + \mathbb{Z}\right) \\
&= \left(\frac{p}{bq} + \mathbb{Z}\right) \otimes 0 = 0 \otimes 0 = 0.
\end{aligned}$$

Example 10.25. Let p and q be distinct prime integers. Then p has inverse modulo q, say $ap \equiv 1 \bmod q$, and q has an inverse modulo p, say $bq \equiv 1 \bmod p$. Given any simple tensor $n \otimes m$ in $\mathbb{Z}/(p) \otimes_{\mathbb{Z}} \mathbb{Z}/(q)$,

$$n \otimes m = ((bq)n) \otimes ((ap)m) = (pbn) \otimes (qam) = 0 \otimes 0.$$

Since all simple tensors are 0 and $\mathbb{Z}/(p) \otimes_{\mathbb{Z}} \mathbb{Z}/(q)$ is generated by simple tensors, we conclude that $\mathbb{Z}/(p) \otimes_{\mathbb{Z}} \mathbb{Z}/(q) = 0$.

Of course not all tensor products are 0. But showing an element in a tensor product is nonzero is somehow harder than showing an element is zero. Usually, to show that a particular element in $M \otimes_R N$ is nonzero, one shows there is a homomoprhism from $M \otimes_R N$ to some R-module L that takes that particular element no some nonzero element in L

Example 10.26. Consider the abelian group $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/(2)$. The map

$$\begin{aligned} 2\mathbb{Z} \times \mathbb{Z}/(2) &\longrightarrow \mathbb{Z}/(2) \\ (a,b) &\longmapsto \tfrac{ab}{2} \end{aligned}$$

is $\mathbb{Z}$-bilinear, and thus it induces an R-module homomorphism $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/(2) \longrightarrow \mathbb{Z}/(2)$. Via this map, $2 \otimes 1 \mapsto 1 \neq 0$, so $2 \otimes 1$ is nonzero in $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/(2)$, and $2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/(2) \neq 0$.

Exercise 34. Show that if $d = \gcd(m,n)$, then $\mathbb{Z}/(n) \otimes_{\mathbb{Z}} \mathbb{Z}/(m) \cong \mathbb{Z}/(d)$.

Similarly to the Hom functor, tensor behaves well with respect to arbitrary direct sums.

Theorem 10.27. *Let M be and R-module, and let $\{N_i\}_{i\in I}$ be an arbitrary family of R-modules. There is an isomorphism*

$$M \otimes_R \left(\bigoplus_{i\in I} N_i\right) \xrightarrow{\cong} \bigoplus_{i\in I} M \otimes_R N_i$$

which is natural, meaning that given two families of R-modules $\{A_i\}_{i\in I}$ and $\{B_j\}_{j\in J}$, and R-module homomorphisms $\sigma_{ij} \colon A_i \longrightarrow B_j$, the R-module homomorphisms

$$\begin{aligned} \bigoplus_{i\in I} A_i &\xrightarrow{\sigma} \bigoplus_{j\in J} B_j \\ (a_i)_{i\in I} &\longmapsto (\sigma_{ij}(a_i))_{j\in J} \end{aligned} \qquad \textit{and} \qquad \tilde{\sigma} = \bigoplus_{i\in I} \sigma_{ij} \colon \bigoplus_{i\in I} M \otimes_R A_i \longrightarrow \bigoplus_{j\in J} M \otimes_R B_j$$

give a commutative diagram

$$\begin{tikzcd} M \otimes_R \left(\bigoplus\limits_{i\in I} A_i\right) \arrow[r, "\cong"] \arrow[d, "1\otimes\sigma"'] & \bigoplus\limits_{i\in I} M \otimes_R A_i \arrow[d, "\tilde{\sigma}"] \\ M \otimes_R \left(\bigoplus\limits_{j\in J} B_j\right) \arrow[r, "\cong"'] & \bigoplus_{j\in J} M \otimes_R B_j. \end{tikzcd}$$

Proof. First, note that the function

$$\begin{aligned} M \times \left(\bigoplus_{i\in I} A_i\right) &\longrightarrow \bigoplus_{i\in I} (M \otimes_R A_i) \\ (m, (a_i)_i) &\longmapsto (m \otimes a_i) \end{aligned}$$

is R-bilinear, so it induces a homomorphism $M \otimes_R \left(\bigoplus\limits_{i\in I} A_i\right) \xrightarrow{\tau} \bigoplus\limits_{i\in I} (M \otimes_R A_i)$.

For each $k \in I$, let ι_k denote the inclusion map $A_k \subseteq \bigoplus_i A_i$. The universal property of the coproduct (which in the case of R-modules, means the direct sum) gives an R-module homomorphism

$$\begin{aligned} \bigoplus_{i\in I}(M \otimes_R A_i) &\xrightarrow{\lambda} M \otimes_R \bigoplus_{i\in I}(A_i) \\ (m \otimes a_i)_i &\longmapsto m \otimes \sum_i \iota_i(a_i) \end{aligned}$$

which we obtain by assembling the R-module homomorphisms $1 \otimes \iota_i$. It is routine to check that λ is the inverse of τ, which must then be an isomorphism. Finally, we can check naturality by checking commutativity of the square above, element by element:

$$\begin{tikzcd} m \otimes (a_i)_i \arrow[r, mapsto] \arrow[d, mapsto] & (m \otimes a_i)_i \arrow[d, mapsto] \\ m \otimes (\sigma_{ij}(a_i))_i \arrow[r, mapsto] & (m \otimes \sigma_{ij}(a_i)). \end{tikzcd}$$

□

Tensor is right exact.

Theorem 10.28. *Let M be an R-module. The functor $M \otimes_R - \colon R\textbf{-mod} \longrightarrow R\textbf{-mod}$ is right exact, meaning that for every exact sequence*

$$A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

the sequence

$$M \otimes_R A \xrightarrow{1\otimes i} M \otimes_R B \xrightarrow{1\otimes p} M \otimes_R C \longrightarrow 0$$

is exact.

Proof. We have three things to show:

- $(1 \otimes p)(1 \otimes i) = 0$.
  It is sufficient to show the map is 0 on simple tensors. And indeed,
  $$(1 \otimes p)(1 \otimes i)(m \otimes a) = (1 \otimes p)(m \otimes p(a)) = m \otimes ip(a) = m \otimes 0 = 0.$$
- $1 \otimes p$ is surjective.
  Consider any $m_1 \otimes c_1 + \cdots + m_n \otimes c_n \in M \otimes_R C$. Since p is surjective, we can find $b_1, \dots, b_n \in B$ such that $p(b_i) = c_i$. Therefore,
  $$(1\otimes p)(m_1 \otimes b_1 + \cdots + m_n \otimes b_n) = m_1 \otimes p(b_1) + \cdots + m_n \otimes p(b_n) = m_1 \otimes c_1 + \cdots + m_n \otimes c_n.$$
- $\ker(1 \otimes p) = \operatorname{im}(1 \otimes i)$
  Let $I = \operatorname{im}(1 \otimes i)$. We have already shown that $I \subseteq \ker(1 \otimes p)$, so $1 \otimes p$ induces a map $q\colon (M \otimes_R B)/I \longrightarrow M \otimes_R C$. Let $\pi\colon M \otimes_R B \longrightarrow (M \otimes_R B)/I$ be the canonical projection. By definition, $q\pi = 1 \otimes p$.

Consider the map

$$\begin{aligned} M \times C &\xrightarrow{f} (M \otimes_R B)/I\,, \\ (m,c) &\longmapsto m \otimes b \end{aligned}$$

where b is such that $p(b) = c$. First, we should check this map f is well-defined. To see that, suppose that $b' \in B$ is another element with $p(b') = c$, so that $p(b - b') = 0$. Then $b - b' \in \ker p = \operatorname{im} i$, so $m \otimes (b - b') \in \operatorname{im}(1 \otimes i) \subseteq I$. Therefore, $m \otimes b = m \otimes b'$ modulo I, and f is well-defined.

Moreover, we can easily check that f is R-bilinear, so f induces a homomorphism of R-modules $M \otimes_R C \longrightarrow (M \otimes_R B)/I$, which we will denote by $\hat{f}$. We will show that $\hat{f}$ is a left inverse of q, so q is injective. And indeed, given $m_i \in M$ and $b_i \in B$, we have

$$\hat{f}q\left(\sum_{i=1}^{n} m_i \otimes b_i\right) = f\left(\sum_{i=1}^{n} m_i \otimes p(b_i)\right) = \sum_{i=1}^{n} f(m_i \otimes p(b_i)) = \sum_{i=1}^{n} m_i \otimes b_i.$$

We conclude that q is injective, and thus

$$\ker(1 \otimes p) = \ker(q\pi) = \ker \pi = I = \operatorname{im}(1 \otimes i).$$

$\square$

Exercise 35. Let M be an R-module. The functor $- \otimes_R M \colon R\text{-}\mathbf{mod} \longrightarrow R\text{-}\mathbf{mod}$ is right exact, meaning that for every exact sequence

$$A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

the sequence

$$A \otimes_R M \xrightarrow{i \otimes 1} B \otimes_R M \xrightarrow{p \otimes 1} C \otimes_R M \longrightarrow 0$$

is exact.

However, tensor is not exact.

Example 10.29. Consider the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{i} \mathbb{Q} \xrightarrow{p} \mathbb{Q}/\mathbb{Z} \longrightarrow 0.$$

Applying the functor $\mathbb{Z}/2 \otimes_{\mathbb{Z}} -$, we get an exact sequence

$$\mathbb{Z}/2 \otimes_{\mathbb{Z}} \mathbb{Z} \longrightarrow \mathbb{Z}/2 \otimes_{\mathbb{Z}} \mathbb{Q} \longrightarrow \mathbb{Z}/2 \otimes_{\mathbb{Z}} \mathbb{Q}/\mathbb{Z} \longrightarrow 0.$$

However, we claim that $1 \otimes i$ is not injective. On the one hand, $\mathbb{Z}/2 \otimes_{\mathbb{Z}} \mathbb{Z} \cong \mathbb{Z}/2$. On the other hand, we have seen in Example 10.23 that $\mathbb{Z}/2 \otimes_{\mathbb{Z}} \mathbb{Q} = 0$, so the map $1 \otimes i$ cannot possibly be injective.

10.3 Hom-tensor adjunction

The Hom and tensor functors are closely related.

Theorem 10.30 (Hom-tensor adjunction I)**.** *Let M, N, and P be R-modules. There is an isomorphism of R-modules*

$$\operatorname{Hom}_R(M \otimes_R N, P) \cong \operatorname{Hom}_R(M, \operatorname{Hom}_R(N, P))$$

that is natural on M, N, and P.

Proof. The universal property of the tensor product says that to give an R-module homomorphism $M \otimes_R N \longrightarrow P$ is the same as giving an R-bilinear map $M \times N \longrightarrow P$. Given such a bilinear map f, the map $n \mapsto f(m \otimes n)$ is R-linear for each $m \in M$, so it defines an R-module homomorphism $N \longrightarrow P$. Now the assignment

$$\begin{aligned} M &\longrightarrow \operatorname{Hom}_S(N, P) \\ m &\longrightarrow (n \mapsto f(m \otimes n)) \end{aligned}$$

is R-linear, f is an R-module homomorphism, and $m \mapsto m \otimes n$ is R-linear on m.

Conversely, given an R-module homomorphism $f \in \operatorname{Hom}_R(M, \operatorname{Hom}_R(N, P))$, one can easily check (exercise!) that $(m, n) \mapsto f(m)(n)$ is an R-bilinear map, so it induces an R-module homomorphism $M \otimes_R N \longrightarrow P$.

So we have constructed a bijection of Hom-sets

$$\begin{tikzcd} \operatorname{Hom}_R(M \otimes_R N, P) \arrow[r, "\tau"] & \operatorname{Hom}_R(M, \operatorname{Hom}_R(N, P)) \, . \\ f \arrow[r, mapsto] & (m \mapsto (n \mapsto f(m \otimes n))) \\ (m \otimes n \mapsto g(m)(n)) & g \arrow[l, mapsto] \end{tikzcd}$$

It's routine to check that both of these bijections are indeed homomorphisms of R-modules, so we leave it as an exercise.

Finally, we have the following commutative diagrams:

$$\begin{tikzcd} A \arrow[d, "f"'] & & \operatorname{Hom}_R(A \otimes_R N, P) \arrow[r, "\cong"] \arrow[d, "(f \otimes 1_N)^*"'] & \operatorname{Hom}_R(A, \operatorname{Hom}_R(N, P)) \arrow[d, "f^*"] \, , \\ B & \arrow[r, rightsquigarrow] & \operatorname{Hom}_R(B \otimes_R N, P) \arrow[r, "\cong"'] & \operatorname{Hom}_R(B, \operatorname{Hom}_R(N, P)) \end{tikzcd}$$

$$\begin{tikzcd} A \arrow[d, "f"'] & & \operatorname{Hom}_R(M \otimes_R A, P) \arrow[r, "\cong"] \arrow[d, "(1_M \otimes f)^*"'] & \operatorname{Hom}_R(M, \operatorname{Hom}_R(A, P)) \arrow[d, "(f^*)_*"] \, , \\ B & \arrow[r, rightsquigarrow] & \operatorname{Hom}_R(M \otimes_R B, P) \arrow[r, "\cong"'] & \operatorname{Hom}_R(M, \operatorname{Hom}_R(B, P)) \end{tikzcd}$$

and

$$\begin{tikzcd} A \arrow[d, "f"'] & & \operatorname{Hom}_R(M \otimes_R N, A) \arrow[r, "\cong"] \arrow[d, "f_*"'] & \operatorname{Hom}_R(M, \operatorname{Hom}_R(N, A)) \arrow[d, "(f^*)^*"] \, . \\ B & \arrow[r, rightsquigarrow] & \operatorname{Hom}_R(M \otimes_R N, B) \arrow[r, "\cong"'] & \operatorname{Hom}_R(M, \operatorname{Hom}_R(N, B)) \end{tikzcd}$$

We leave checking these do indeed commute as an exercise. □

Corollary 10.31 (Tensor and Hom are adjoint functors)**.** *Let R be a ring, and M an R-module. The functor $- \otimes_R M : R\text{-}\boldsymbol{mod} \longrightarrow R\text{-}\boldsymbol{mod}$ is left adjoint to $\operatorname{Hom}_R(M, -) : R\text{-}\boldsymbol{mod} \longrightarrow R\text{-}\boldsymbol{mod}$.*

Proof. The adjointness translates into the fact that for all R-modules N and P there is a bijection
$$\operatorname{Hom}_R(N \otimes_R M, P) \cong \operatorname{Hom}_R(N, \operatorname{Hom}_R(M, P))$$
which is natural on N and P, which is a corollary of Theorem 10.30. $\square$

Later, when we talk about more general abelian categories, we will see that this adjunction *implies* that Hom is left exact and that tensor is right exact; in fact, this is a more general fact about adjoint pairs. For now, we want to discuss a more general version of this Hom-tensor adjunction.

One of the reasons tensor products are useful is the fact that we can use tensor products to extend modules to ring extensions.

Lemma 10.32. *Let $R \xrightarrow{f} S$ be a ring homomorphism. Then $S \otimes_R -$ determines a functor from R-modules to S-modules.*

Proof. First, we claim that $S \otimes_R M$ is an S-module for every R-module M. Since $S \otimes_R M$ is an R-module, it is in particular an abelian group. For each fixed $s \in S$, the multiplication map $S \xrightarrow{\cdot s} S$ is an R-module homomorphism, and thus it induces an R-module homomorphism $\mu_s \colon S \otimes_R M \longrightarrow S \otimes_R M$. For each $s \in S$ and each $\sum_i s_i \otimes m_i \in S \otimes_R M$, we define
$$s \cdot \left(\sum_i s_i \otimes m_i \right) := \mu_s \left(\sum_i s_i \otimes m_i \right).$$
We claim that this determines an S-module structure on $S \otimes_R M$.

- $1 \cdot (\sum_i s_i \otimes m_i) = \cdot(\sum_i s_i \otimes m_i)$, since μ_1 is the identity on $S \otimes_R M$.
- Since $S \xrightarrow{\cdot s_1} S \xrightarrow{\cdot s_2} S$ is the multiplication by $s_2 s_1$, we conclude that $\mu_{s_2}\mu_{s_1} = \mu_{s_2 s_1}$, and $s_2(s_1 t) = (s_2 s_2)t$ for all $t \in S \otimes_R M$.
- Similarly, $\mu_{s_1} + \mu_{s_2} = \mu_{s_1+s_2}$, so $(s_1 + s_2)t = s_1 t + s_2 t$ for all $t \in S \otimes_R M$.
- Finally, all the μ_s are R-module homomorphisms, so $s \cdot (t_1 + t_2) = s \cdot t_1 + s \cdot t_2$ for all $s \in S$ and all $t_1, t_2 \in S \otimes_R M$.

Now given any R-module homomorphism $f : M \longrightarrow N$, $\mathrm{id}_S \otimes_1$ is an R-module homomorphism, which we claim is also an S-module homomorphism via the S-module structure we just discussed. $\square$

Definition 10.33. Let $R \xrightarrow{f} S$ be a ring homomorphism. The **extension of scalars** from R to S is the functor $S \otimes_R - : R\text{-}\mathbf{mod} \longrightarrow S\text{-}\mathbf{mod}$ we discussed above. For each R-module M, we get an S-module $S \otimes_R M$ with
$$s \cdot \left(\sum_i s_i \otimes m_i \right) := \sum_i (s s_i) \otimes m_i,$$
and for each R-module homomorphism $f \colon M \longrightarrow N$ we get the S-module homomorphism $1 \otimes_R f$.

This is in closely related to restriction of scalars, which we discussed before, although we did not phrase it in this categorical language.

Definition 10.34. Let $R \xrightarrow{f} S$ be a ring homomorphism. The **restriction of scalars functor** from S to R is the functor $f^*: S\text{-}\mathbf{mod} \longrightarrow R\text{-}\mathbf{mod}$ that takes each S-module M to the R-module f^*M with underlying abelian group M and R-module structure

$$r \cdot m := f(r)m$$

induced by f. Moreover, for each S-module homomorphism $g: M \longrightarrow N$ we get the R-module homomorphism $f^*(g)\colon f^*(M) \longrightarrow f^*(N)$ defined by $f^*(g)(m) := g(n)$.

Exercise 36. Check that restriction of scalars as defined above is indeed a functor.

Theorem 10.35 (Hom-tensor adjunction II)**.** *Let S be an R-algebra, M be an R-module, and P and N be S-modules. There is an isomorphism of abelian groups*

$$\operatorname{Hom}_S(M \otimes_R N, P) \cong \operatorname{Hom}_R(M, \operatorname{Hom}_S(N, P)).$$

Moreover, this isomorphism is natural on M, N, and P, meaning that it induces natural isomorphisms

- *between* $\operatorname{Hom}_S(- \otimes_R N, P)$ *and* $\operatorname{Hom}_R(-, \operatorname{Hom}_S(N, P))$.
- *between* $\operatorname{Hom}_S(M \otimes_R -, P)$ *and* $\operatorname{Hom}_R(M, \operatorname{Hom}_S(-, P))$.
- *between* $\operatorname{Hom}_S(M \otimes_R N, -)$ *and* $\operatorname{Hom}_R(M, \operatorname{Hom}_S(N, -))$.

Proof. Consider the map

$$\begin{array}{rcl} \operatorname{Hom}_S(M \otimes_R N, P) & \xrightarrow{\tau} & \operatorname{Hom}_R(M, \operatorname{Hom}_S(N, P)) \, . \\ f & \longmapsto & m \mapsto (n \mapsto f(m \otimes n)) \end{array}$$

Fix f. For each $m \in M$, let τ_m be the map $N \longrightarrow P$ defined by $\tau_m(n) := f(m \otimes n)$. Note that τ_m is indeed a homomorphism of S-modules, since it is the composition of two S-module maps, f and $m \otimes_R \mathrm{id}_N$, where m is the constant map $M \longrightarrow M$ equal to m.

We should check that our proposed map τ is indeed a map of abelian groups. It is immediate from the definition that τ sends the 0-map to the 0-map. Moreover, given S-module homomorphisms $f, g\colon M \otimes N \longrightarrow P$, and any $n \in N$, we have

$$\begin{aligned} \tau_m(f+g)(n) &= (f+g)(m \otimes n) && \text{by definition} \\ &= f(m \otimes n) + g(m \otimes n) && \text{since } f \text{ and } g \text{ are } S\text{-module maps} \\ &= \tau_m(f)(n) + \tau_m(g)(n) && \text{by definition} \end{aligned}$$

so $\tau_m(f+g) = \tau_m(f) + \tau_m(g)$ for all $m \in M$, and thus $\tau(f+g) = \tau(f) + \tau(g)$.

Suppose that $\tau(f) = 0$. Then for every $m \in M$ and every $n \in N$,

$$0 = \tau(f)(m)(n) = \tau_m(f)(n) = f(m \otimes n),$$

so f vanishes at every simple tensor, and we must have $f = 0$. On the other hand, if we are given $g \in \mathrm{Hom}_R(M, \mathrm{Hom}_S(N, P))$, consider the map $M \times N \longrightarrow P$ defined by $\tilde{f}(m, n) = g(m)(n)$. Since g is a homomorphism of R-modules, it is R-linear on m. Moreover, for each fixed m, $g(m)$ is a homomorphism of S-modules, so in particular $g(m)$ is R-linear. Together, these say that $\tilde{f}$ is an R-bilinear map. Let f be the homomorphism of R-modules $M \otimes_R N \longrightarrow P$ induced by $\tilde{f}$. By definition, $f(m \otimes n) = \tilde{f}(m, n) = g(m)(n)$, so $\tau(f) = g$. We conclude that τ is a bijection.

We leave the statements about naturality as exercises. □

Corollary 10.36 (Adjointness of restriction and extension of scalars)**.** *Let f $R \longrightarrow S$ be a ring homomorphism. The restriction of scalars functor $f^* \colon S$-**mod** $\longrightarrow R$-**mod** is the right adjoint of the extension of scalars functor $f_! \colon R$-**mod** $\longrightarrow S$-**mod**.*

Proof. We need to show that for every R-module M and every S-module N there are bijections

$$\mathrm{Hom}_S(f_!(M), N) \cong \mathrm{Hom}_R(M, f^*(N))$$

which are natural on both M and N. By Theorem 10.35, we have natural bijections

$$\mathrm{Hom}_S(M \otimes_R S, N) \cong \mathrm{Hom}_R(M, \mathrm{Hom}_S(S, N)).$$

The module $M \otimes_R S$ is precisely $f_!(M)$. By Exercise 31, $\mathrm{Hom}_S(S, N) \cong N$ as an S-module. An isomorphism of S-modules $\mathrm{Hom}_S(S, N) \longrightarrow N$ is in particular an R-linear map, and thus also an isomorphism of R-modules. So $\mathrm{Hom}_S(S, N) \cong f^*(N)$ as R-modules. Therefore, the Hom-tensor adjuntion gives us the natural bijections we were looking for. □

The idea is that restriction of scalars and extension of scalars are the most efficient ways of making an R-module out of an S-module, and vice-versa.

Chapter 11

Enough (about) projectives and injectives

11.1 Projectives

While the Hom functors on R**-mod** are not exact, they are always left exact. Modules M that make $\mathrm{Hom}_R(M, -)$ or $\mathrm{Hom}_R(-, M)$ exact functors are special.

Definition 11.1. An R-module P is **projective** if given any surjective R-module homomorphism $A \xrightarrow{p} B$ and any R-module homomorphism $P \xrightarrow{f} B$, there exists an R-module homomorphism g such that the diagram

$$\begin{tikzcd} & P \arrow[dl, dashed, "g"'] \arrow[d, "f"] & \\ A \arrow[r, "p"'] & B \arrow[r] & 0 \end{tikzcd}$$

commutes.

Remark 11.2. The commutativity of the diagram

$$\begin{tikzcd} & P \arrow[dl, "g"'] \arrow[d, "f"] & \\ A \arrow[r, "p"'] & B \arrow[r] & 0 \end{tikzcd}$$

says that $p_*(g) = f$, where p_* is the map $\mathrm{Hom}_R(P, A) \longrightarrow \mathrm{Hom}_R(P, B)$ induced by p. Whenever this happens, we say that g is a **lifting** of f.

Free modules are projective.

Theorem 11.3. *If F is a free R-module, then F is projective.*

Proof. Suppose we are given R-module homomorphisms $A \xrightarrow{p} C$ and $P \xrightarrow{f} B$ such that p is surjective. Fix a basis $B = \{b_i\}_i$ for F. Since p is surjective, for each i we can choose

$a_i \in A$ such that $p(a_i) = f(b_i)$. Consider the function $u\colon B \longrightarrow A$ given by $u(b_i) = a_i$. The universal property of free modules says that there exists an R-module homomorphism $g\colon P \longrightarrow A$ that coincides with u for all basis elements. Now

$$pg(b_i) = pu(b_i) = p(a_i) = f(b_i),$$

so pg agrees with f for all basis elements. Since B generates F, we conclude that $pg = f$. □

Projective modules are precisely those that make the covariant Hom functor exact.

Theorem 11.4. *Let P be an R-module. The functor $\mathrm{Hom}_R(P, -)$ is exact if and only if P is projective.*

Proof. By Theorem 10.7, $\mathrm{Hom}_R(P, -)$ is left exact. So the statement is really that P is projective if and only for any short exact sequence

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0\,,$$

the induced map $p_* : \mathrm{Hom}_R(P, B) \longrightarrow \mathrm{Hom}_R(P, C)$ is surjective. So say we are given a surjective map

$$B \xrightarrow{p} C \longrightarrow 0\,.$$

The induced map p_* is surjective if and only of for every $f \in \mathrm{Hom}_R(P, C)$ there exists a lifting $g \in \mathrm{Hom}_R(P, B)$ of f, meaning $p_*(g) = f$. By Remark 11.2, the existence of such a g for all such surjective maps p is precisely the condition that P is projective. □

We can rephrase the condition that a module is projective or injective in terms of split exact sequences.

Definition 11.5. We say that a short exact sequence

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

splits or is a **split short exact sequence** if it is isomorphic to

$$0 \longrightarrow A \xrightarrow{i} A \oplus C \xrightarrow{p} C \longrightarrow 0$$

where i is the inclusion of the first component and p is the projection onto the second component.

Lemma 11.6 (Splitting Lemma)**.** *Consider the short exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

of R-modules. The following are equivalent:

a) There exists a homomorphism of R-modules $q\colon B \longrightarrow A$ such that $qf = \mathrm{id}_A$.

b) There exists a homomorphism of R-modules $r\colon C \longrightarrow B$ such that $gr = \mathrm{id}_C$.

c) The short exact sequence splits.

Proof. If the sequence splits, then consider an isomorphism of complexes

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "a"'] & B \arrow[r, "g"] \arrow[d, "b"'] & C \arrow[r] \arrow[d, "c"'] & 0 \\ 0 \arrow[r] & A \arrow[r, "i"] & A \oplus C \arrow[r, "p"] \arrow[l, bend left, "\pi"] & C \arrow[r] \arrow[l, bend left, "j"] & 0, \end{tikzcd}$$

meaning that the diagram commutes and a, b, and c are isomorphisms of R-modules, i is the inclusion in the first component, and p is the projection onto the second component. Let $\pi : A \oplus C \longrightarrow A$ be the projection onto the first component, and $j : C \longrightarrow A \oplus C$ be the inclusion onto the first component. Now consider the maps $q := a^{-1}\pi b$ and $r := b^{-1}jc$. Then

$$\begin{aligned} qf &= a^{-1}\pi bf \\ &= a^{-1}\pi ia && \text{by commutativity} \\ &= a^{-1}a && \text{because } \pi i = \mathrm{id}_A \\ &= 1_A \end{aligned}$$

and

$$\begin{aligned} gr &= gb^{-1}jc \\ &= c^{-1}(cg)b^{-1}jc && \text{multiplied by } c^{-1}c = 1_C \\ &= c^{-1}(pb)b^{-1}jc && \text{by commutativity} \\ &= c^{-1}pjc && \text{because } bb^{-1} = 1_B \\ &= c^{-1}c && \text{because } pj = \mathrm{id}_C \\ &= 1_C \end{aligned}$$

Therefore, c implies a and b.

Now suppose that a holds, and let's show that the sequence splits. First, we need to show that $B \cong A \oplus C$.

Every $b \in B$ can be written as

$$b = (b - fq(b)) + fq(b),$$

where $fq(b) \in \operatorname{im} f \cong A$, and

$$q(b - fq(b)) = q(b) - \underbrace{qf}_{\mathrm{id}_A}(q(b)) = q(b) - q(b) = 0,$$

so $b - fq(b) \in \ker q$. This shows that $B = \operatorname{im} f + \ker q$. Moreover, if $f(a) \in \ker q$, then $a = qf(a) = 0$, so $\operatorname{im} f \cap \ker q = 0$, and $B = \operatorname{im} f \oplus \ker q$. Now when we restrict g to $\ker q$, g becomes injective. We claim it is also surjective, and thus an isomorphism. Indeed, for any $c \in C$ we can pick $b \in B$ such that $g(b) = c$, since g is surjective, and we showed that we can write $b = f(a) + k$ for some $k \in \ker q$. Then

$$g(k) = \underbrace{gf}_{0}(a) + g(k) = g(b) = c.$$

Finally, note that $\operatorname{im} f \cong A$, so we conclude that $B \cong A \oplus C$, via the isomorphism φ given by

$$\begin{aligned} B &\longrightarrow \operatorname{im} f \oplus \ker q \longrightarrow A \oplus C \\ b &\longmapsto (fq(b), b - fq(b)) \longmapsto (q(b), g(b)). \end{aligned}$$

Since $gf = 0$ and $qf = \mathrm{id}_A$, $\varphi f(a) = (qf(a), 0) = (a, 0)$, so $\varphi f = i$, where $i\colon A \longrightarrow A \oplus C$ is the inclusion on the first factor. If $p\colon A \oplus C \longrightarrow C$ denotes the projection onto the second factor, $p\varphi = g$. Together, these two facts say that the following is a map of complexes:

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, equal] & B \arrow[r, "g"] \arrow[d, "\varphi"'] & C \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & A \arrow[r, "i"'] & A \oplus C \arrow[r, "p"'] & C \arrow[r] & 0. \end{tikzcd}$$

Since φ is an isomorphism, so is our map of complexes, and thus our original sequence is a split exact sequence. This shows that a implies c.

Now assume b holds. Every $b \in B$ can be written as

$$b = (b - rg(b)) + rg(b),$$

where $rg(b) \in \operatorname{im} r$ and

$$g(b - rg(b)) = g(b) - \underbrace{gr}_{\mathrm{id}_C}(g(b)) = g(b) - g(b) = 0,$$

so $b - rg(b) \in \ker g$. This shows that $B = \ker g + \operatorname{im} r$. Moreover, if $r(c) \in \ker g$, then

$$c = \mathrm{id}_C(c) = gr(c) = 0.$$

Therefore, $B = \ker g \oplus \operatorname{im} r$. Now r is injective, since $r(c) = 0 \implies c = gr(c) = 0$, and thus $\operatorname{im} r \cong C$. Since $\ker g = \operatorname{im} f \cong A$, we conclude that $B \cong A \oplus C$, via the isomorphism

$$\begin{aligned} A \oplus C &\xrightarrow{\ \psi\ } B \\ (a, c) &\longmapsto f(a) + r(c). \end{aligned}$$

Finally, let $i : A \longrightarrow A \oplus C$ denote the inclusion of the first factor, and $p : A \oplus C \longrightarrow C$ denote the projection onto the second factor. By construction, $\psi i = f$. Moreover,

$$g\psi(a, c) = \underbrace{gf}_{0}(a) + \underbrace{gr}_{\mathrm{id}_C}(c) = c,$$

so $g\psi = p$. Together, these say that the diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "i"] \arrow[d, equal] & A \oplus C \arrow[r, "p"] \arrow[d, "\psi"'] & C \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & A \arrow[r, "f"'] & B \arrow[r, "g"'] & C \arrow[r] & 0 \end{tikzcd}$$

commutes, and must then be an isomorphism of short exact sequences. $\square$

Not every short exact sequence is split.

Example 11.7. The short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2} \mathbb{Z} \longrightarrow \mathbb{Z}/2 \longrightarrow 0$$

is not split. Indeed, $\mathbb{Z}$ does not have any 2-torsion elements, so it is not isomorphic to $\mathbb{Z} \oplus \mathbb{Z}/2$.

An alternative explanation is that there is no splitting to the inclusion $\mathbb{Z} \xrightarrow{2} \mathbb{Z}$. On the one hand, every $\mathbb{Z}$-module map is given by multiplication by a fixed integer n, so a splitting $f : \mathbb{Z} \longrightarrow \mathbb{Z}$ would be of the form $f(a) = na$ for some fixed n. On the other hand, our proposed splitting f must send 2 to 1, but there is no integer solution n to $2n = f(2) = 1$.

More surprisingly, a short exact sequence of the form

$$0 \longrightarrow A \xrightarrow{f} A \oplus C \xrightarrow{g} C \longrightarrow 0$$

is not necessarily split, not unless f is the inclusion of the first component and g is the projection onto the second component.

Example 11.8. Consider the short exact sequence

$$0 \longrightarrow \mathbb{Z}/(2) \xrightarrow{f} \mathbb{Z}/(4) \xrightarrow{g} \mathbb{Z}/(2) \longrightarrow 0$$

where f is the inclusion of the subgroup generated by 2, so $f(1) = 2$, and g is the quotient onto that subgroup, meaning $g(1) = 1$. This is not a split short exact sequence, because $\mathbb{Z}/(4) \ncong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2)$. Now let $M = \bigoplus_{\mathbb{N}}(\mathbb{Z}/(2) \oplus \mathbb{Z}/(4))$ be the direct sum of infinitely many copies of $\mathbb{Z}/(2) \oplus \mathbb{Z}/(4)$. Then

$$\mathbb{Z}/(2) \oplus M \cong M \cong M \oplus \mathbb{Z}/(4),$$

and the sequence

$$0 \longrightarrow \mathbb{Z}/(2) \xrightarrow{h} \mathbb{Z}/(4) \oplus M \xrightarrow{t} \mathbb{Z}/(2) \oplus M \longrightarrow 0$$

with $h(a) = (f(a), 0)$ and $t(a, m) = (g(a), m)$ is still exact. The middle term is indeed isomorphic to the direct sum of the other two, and yet this is not a split exact sequence: a splitting $q : \mathbb{Z}/(4) \oplus M \longrightarrow \mathbb{Z}/(2)$ of h would restrict to a splitting $\mathbb{Z}/(4) \longrightarrow \mathbb{Z}/(2)$ of f, which we know cannot exist.

Given splittings q and r for a short exact sequence as in Lemma 11.6, we can quickly show that our short exact sequence splits using the Five Lemma.

Exercise 37 (The Five Lemma)**.** Consider the following commutative diagram of R-modules with exact rows:

$$\begin{tikzcd}
A' \arrow[r] \arrow[d, "a"] & B' \arrow[r] \arrow[d, "b"] & C' \arrow[r] \arrow[d, "c"] & D' \arrow[r] \arrow[d, "d"] & E' \arrow[d, "e"] \\
A \arrow[r] & B \arrow[r] & C \arrow[r] & D \arrow[r] & E
\end{tikzcd}$$

Show that if a, b, d, and e are isomorphisms, then c is an isomorphism.

Remark 11.9. Given a short exact sequence, suppose we have R-module homomorphisms q and r

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] & B \arrow[r, "g"] \arrow[l, bend left, "q"] & C \arrow[r] \arrow[l, bend left, "r"] & 0 \end{tikzcd}$$

such that $qf = \mathrm{id}_A$ and $rg = \mathrm{id}_C$. Then we get an induced map

$$\begin{aligned} B &\xrightarrow{\varphi} A \oplus C \\ b &\longmapsto (q(b), g(b)) \end{aligned}$$

such that the diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, equal] & B \arrow[r, "g"] \arrow[d, dashed, "\varphi"] & C \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & A \arrow[r, "i"'] & A \oplus C \arrow[r, "p"'] & C \arrow[r] & 0. \end{tikzcd}$$

commutes. The Five Lemma 37 guarantees that φ must be an isomorphism, so our diagram is an isomorphism of short exact sequences.

There are many ways in which R-**mod** behaves better than the category of groups, and this is one of them.

Remark 11.10. Lemma 11.6 does not hold if we replace R-modules with the category **Grp** of groups. For example, consider the symmetric group on 3 elements S_3 and the inclusion $A_3 \hookrightarrow S_3$ of the alternating group in S_3. Notice that A_3 is precisely the kernel of the sign map $\mathrm{sign}: S_3 \longrightarrow \mathbb{Z}/2$, so that

$$0 \longrightarrow A_3 \longrightarrow S_3 \longrightarrow \mathbb{Z}/2 \longrightarrow 0$$

is a short exact sequence. Moreover, this exact sequence is not split, since S_3 is not abelian but $A_3 \oplus \mathbb{Z}/2$ is, and thus $S_3 \ncong A_3 \oplus \mathbb{Z}/2$. However, any group homomorphism $u\colon \mathbb{Z}/2 \longrightarrow S_3$ defined by sending the generator to any two cycle is a splitting for our short exact sequence, meaning $\mathrm{sign} \circ u = \mathrm{id}_{\mathbb{Z}/2}$.

Funny enough, there is no splitting for the inclusion $A_3 \subseteq S_3$, since there are no nontrivial homomorphisms $S_3 \longrightarrow A_3$: A_3 has no elements of order 2, so a group homomorphism $S_3 \longrightarrow A_3$ must send every 2-cycle in S_3 must be sent to the identity, but 2-cycles generate S_3.

We can now characterize projective modules in term of split short exact sequences.

Theorem 11.11. *An R-module P is projective if and only if every short exact sequence*

$$0 \longrightarrow A \longrightarrow B \longrightarrow P \longrightarrow 0$$

splits.

Proof. Consider a short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} P \longrightarrow 0\,.$$

If P is projective, the identity map on P lifts to a map $P \longrightarrow B$, meaning that

$$\begin{tikzcd} & P \arrow[dl, dashed, "h"'] \arrow[d, equal] & \\ B \arrow[r, "g"'] & P \arrow[r] & 0 \end{tikzcd}$$

commutes. This says that our map h

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] & B \arrow[r, "g"] & P \arrow[r] \arrow[l, bend left, "h"] & 0 \end{tikzcd}$$

is a splitting for our short exact sequence, which must then be split, by Lemma 11.6.

Conversely, suppose that every short exact sequence

$$0 \longrightarrow A \longrightarrow B \longrightarrow P \longrightarrow 0$$

splits, and consider any diagram

$$\begin{tikzcd} & P \arrow[d, "f"] & \\ B \arrow[r, "p"'] & C \arrow[r] & 0. \end{tikzcd}$$

Let F be a free module that surjects onto P — for example, the free module on a set of generators of P — and fix a surjection $\pi\colon F \twoheadrightarrow P$. By assumption, the short exact sequence

$$\begin{tikzcd} 0 \arrow[r] & \ker p \arrow[r] & F \arrow[r, "\pi"] & P \arrow[r] \arrow[l, bend left, dashed, "h"] & 0 \end{tikzcd}$$

splits, so by Lemma 11.6 there exists h such that $\pi h = \mathrm{id}_P$. Now since F is free, we can define an R-module map $\hat{g}\colon F \longrightarrow B$ that such that

$$\begin{tikzcd} F \arrow[r, "\pi"'] \arrow[d, dashed, "\hat{g}"'] & P \arrow[l, bend right, "h"'] \arrow[d, "f"] & \\ B \arrow[r, "p"'] & C \arrow[r] & 0 \end{tikzcd}$$

commutes, by sending each basis element $b \in F$ to any lift of $f\pi(b)$ in B via p. Now set $g := \hat{g}h$. Now

$$\begin{aligned} pg &= p\hat{g}h && \text{by definition} \\ &= f\pi h && \text{by commutativity} \\ &= f && \text{since } \pi h = \mathrm{id}_P, \end{aligned}$$

so g is a lift of p by f. $\square$

We have seen that free modules are projective; what other modules are projective?

Definition 11.12. An R-module M is a **direct summand** of an R-module N if there exists an R-module A such that $A \oplus M \cong N$.

Remark 11.13. Saying that M is a direct summand of N is equivalent to giving a splitting s of the inclusion map $M \xrightarrow{i} N$, meaning that $si = \mathrm{id}_N$. As we've argued in Lemma 11.6, such a splitting s gives

$$N = \operatorname{im} i \oplus \ker s.$$

Essentially repeating the argument we used in Lemma 11.6, every element in N can be written as

$$n = (n - is(n)) + is(n),$$

where $is(n) \in \operatorname{im} i$ and $n - is(n) \in \ker s$, and $\ker s \cap \operatorname{im} i = 0$ because if $i(a) \in \ker s$ then $a = si(a) = 0$.

When we are dealing with graded modules over a graded ring, the kernels and images of graded maps are graded modules, and the equality $N = \operatorname{im} i \oplus \ker s$ is a graded direct sum of graded modules.

Theorem 11.14. *An R-module is projective if and only if it is a direct summand of a free R-module. In particular, a finitely generated R-module is projective if and only if it is a direct summand of R^n for some n.*

Proof. Let P be a projective module, and fix a free module F surjecting onto P. If P is finitely generated, we can take $F = R^n$ for some n. The short exact sequence

$$0 \longrightarrow \ker p \longrightarrow F \xrightarrow{\pi} P \longrightarrow 0$$

must split by Theorem 11.11, so P is a direct summand of F.

Now suppose P is a direct summand of a free module F. In particular, we have an inclusion map $i\colon P \longrightarrow F$ and a projection map $\pi\colon F \longrightarrow F$, and $\pi i = \mathrm{id}_P$. Given any diagram

$$\begin{tikzcd} & P \arrow[d, "f"] & \\ B \arrow[r, "p"'] & C \arrow[r] & 0, \end{tikzcd}$$

we can again define an R-module homomorphism

$$\begin{tikzcd} F \arrow[r, "\pi"'] \arrow[d, dashed, "h"'] & P \arrow[l, bend right, "i"'] \arrow[d, "f"] & \\ B \arrow[r, "p"'] & C \arrow[r] & 0, \end{tikzcd}$$

such that $ph = f\pi$. Setting $g := hi$, we do indeed obtain $pg = f$, since

$$\begin{aligned} pg &= phi && \text{by definition} \\ &= f\pi i && \text{because } ph = f\pi \\ &= f && \text{since } \pi i = \mathrm{id}_P\,. \square \end{aligned}$$

Corollary 11.15.

1) *Every direct summand of a projective module is projective.*
2) *Every direct sum of projective modules is projective.*

Proof.

1) Suppose $M \oplus A \cong P$ for some projective module P. By Theorem 11.14, there exists a free R-module F and an R-module B such that $P \oplus B \cong F$. Then $M \oplus A \oplus M \cong F$, and by Theorem 11.14 this implies M is projective.
2) Suppose $\{P_i\}_{i \in I}$ are all projective modules. By Theorem 11.14, there exist free modules F_i such that each P_i is a direct summand of F_i. Therefore, $\oplus P_i$ is a direct summand of $\oplus_i F_i$, which is also free. By Theorem 11.14, this implies that $\oplus P_i$ is projective.

□

Projective modules are not necessarily free.

Example 11.16. The ring $R = \mathbb{Z}/(6)$ can be written as a direct sum of the ideals

$$I = (2) \text{ and } J = (3).$$

Indeed, $R = I + J$ and $I \cap J = 0$, so $R = I \oplus J$. By Corollary 11.15, I and J are projective R-modules. However, I and J are *not* free. This can easily be explained numerically: every finitely generated free R-module is of the form R^n, so it has 6^n elements for some n, while I and J have 3 and 2 elements respectively.

However, over a local ring, every projective module is indeed free.

Theorem 11.17. *Let $(R, \mathfrak{m})$ be a local ring. Every finitely generated projective R-module is free.*

Proof. Let P be a finitely generated projective module. Let $n := \mu(P)$ be the minimal number of generators of P, and let $m_1, \ldots, m_n$ be a minimal generating set. The map

$$\begin{aligned} R^n &\xrightarrow{\quad\pi\quad} P \\ (r_1, \ldots, r_n) &\longrightarrow r_1 m_1 + \cdots + r_n m_n \end{aligned}$$

is surjective, and an element $(r_1, \ldots, r_n) \in \ker \pi$ corresponds to a relation $r_1 m_1 + \cdots + r_n m_n = 0$. As we saw in Proposition 4.32 and Definition 4.34, our assumption that $m_1, \ldots, m_n$ form a minimal generating set means that their images in $P/\mathfrak{m}P$ are linearly independent, and thus form a basis for $P/\mathfrak{m}P$. Therefore, if $(r_1, \ldots, r_n) \in \ker \pi$ then we must have $r_1, \ldots, r_n \in \mathfrak{m}$. In particular, $\ker \pi \subseteq \mathfrak{m}F$.

By Theorem 11.11, the short exact sequence

$$0 \longrightarrow \ker \pi \longrightarrow F \xrightarrow{\pi} P \longrightarrow 0$$

splits, and there exists an injective map $j \colon P \longrightarrow F$ such that $F = \operatorname{im} j \oplus \ker \pi$. We have shown that $\ker \pi \subseteq \mathfrak{m}F$, so we must have $\ker \pi = \mathfrak{m}(\ker \pi)$. Since $\ker \pi$ is a finitely generated module, NAK 4.31 implies that $\ker \pi = 0$. We conclude that π is an isomorphism, and P is free. □

Kaplansky [Kap58] showed that this holds even for modules that are not necessarily finitely generated, but only generated by countably many elements.

Theorem 11.18. *Let R be an $\mathbb{N}$-graded k-algebra with $R_0 = k$, and $\mathfrak{m} = R_+$. If M is a graded R-module. which is a direct sum of a finitely generated graded R-module, then M is free.*

Proof. Let $f_1, \ldots, f_s$ be a minimal set of homogeneous generators for M, of degrees $d_1, \ldots, d_s$, respectively. The surjective R-module homomorphism

$$\begin{aligned} F := R(-d_1) \oplus \cdots \oplus R(-d_s) &\xrightarrow{\quad \pi \quad} M \\ (r_1, \ldots, r_s) &\longrightarrow r_1 f_1 + \cdots + r_s f_s \end{aligned}$$

is a graded map of degree 0. We claim this is an isomorphism, so we just need to show that $\ker \pi = 0$. Our elements $f_1, \ldots, f_s$ give a basis for the $R/\mathfrak{m}$-vector space $M/\mathfrak{m}M$, so if $(r_1, \ldots, r_s) \in \ker \pi$, or equivalently $r_1 f_1 + \cdots + r_s f_s = 0$, then $r_1, \ldots, r_s \in \mathfrak{m}$. In particular, $\ker \pi \subseteq \mathfrak{m}F$.

Now suppose that $M \oplus N \cong G$ for some graded free R-module G, and let $g_1, \ldots, g_t$ be a homogenous basis for G. Let p be the degree 0 graded projection map $G \longrightarrow M$. Now both π and p are degree 0 surjective maps, so for each i there exists some homogeneous element $h_i \in F$ of the same degree such that $\pi(h_i) = p(g_i)$. Since G is a free graded module, we can define a degree 0 graded homomorphism of R-modules $G \xrightarrow{\alpha} F$ by setting $\alpha(g_i) = h_i$. Let $\beta \colon M \longrightarrow F$ be the restriction of α to M. Now notice that $\pi\beta$ is the identity on M by construction, and thus π is a splitting of β. In particular, as described in Remark 11.13, $F = \operatorname{im} \beta \oplus \ker \pi$. We showed above that $\ker \pi \subseteq \mathfrak{m}F$, so in particular $F = \operatorname{im} \beta + \mathfrak{m}F$. By NAK 4.30, we must have $F = \operatorname{im} \beta$. Then $\ker \pi = 0$ and π is an isomorphism. In particular, M is a graded free R-module. $\square$

Finally, we record an easy result that we have used repeatedly at this point.

Lemma 11.19. *For every R-module M, there exists a free module F surjecting onto M. If M is finitely generated, we can take F to be finitely generated.*

11.2 Injectives

Injective modules are dual to projectives.

Definition 11.20. An R-module I is **injective** if given an injective R-module homomorphism $i \colon A \longrightarrow B$ and an R-module homomorphism $f \colon A \longrightarrow I$, there exist an R-module homomorphism g such that

$$\begin{tikzcd} & I & \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commutes.

These are precisely the modules I such that $\mathrm{Hom}_R(-, I)$ is exact.

Theorem 11.21. *An R-module I is injective if and only if $\mathrm{Hom}_R(-, I)$ is exact, meaning that for every short exact sequence*

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

we get an exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(C, I) \xrightarrow{p^*} \mathrm{Hom}_R(B, I) \xrightarrow{i^*} \mathrm{Hom}_R(A, I) \longrightarrow 0.$$

Proof. By Theorem 10.7, $\mathrm{Hom}_R(-, I)$ is left exact, so for any short exact sequence

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

we get an exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(C, I) \xrightarrow{p^*} \mathrm{Hom}_R(B, I) \xrightarrow{i^*} \mathrm{Hom}_R(A, I).$$

So the content of the theorem is that I is injective if and only if every injective R-module homomorphism $i \colon A \longrightarrow B$, the induced map i^* is surjective. Now notice that i^* is surjective if and only if every $f \in \mathrm{Hom}_R(A, I)$ lifts to some $g \in \mathrm{Hom}_R(B, I)$, meaning

$$\begin{tikzcd} & I & \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commutes. That is precisely what we want for I to be injective. □

The product of injectives is injective.

Lemma 11.22. *Given any family $\{M_i\}_{i \in I}$ of injective modules, $\prod_{i \in I} M_i$ is injective.*

Proof. Let $\pi_j \colon \prod_{i \in I} M_i \longrightarrow M_j$ be the projection onto the jth factor. Given any diagram

$$\begin{tikzcd} & \prod_{i \in I} M_i & \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B, \end{tikzcd}$$

the fact that M_i is injective gives us R-module homomorphisms g_i such that

$$\begin{tikzcd} & M_i & \\ 0 \arrow[r] & A \arrow[u, "\pi_i f"] \arrow[r, "i"'] & B \arrow[ul, "g_i"'] \end{tikzcd}$$

commutes for each i. Now the R-module homomorphism

$$\begin{aligned} B &\xrightarrow{\ g\ } \textstyle\prod_{i \in I} M_i \\ b &\longmapsto (g_i(b)) \end{aligned}$$

makes the diagram

$$\begin{tikzcd} & \prod_{i\in I} M_i & \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B) - g \arrow[ul] \end{tikzcd}$$

commute, so $\prod_{i \in I} M_i$ is injective. $\square$

Direct summands of injective modules are also injective.

Lemma 11.23. *If $M \oplus N = E$ is an injective R-module, then so are M and N.*

Proof. Any diagram

$$\begin{tikzcd} & M & \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \end{tikzcd}$$

can be extended to a map $A \longrightarrow E$ by composing f with the inclusion of the first factor. Since E is injective, there exists h such that

$$\begin{tikzcd} & M \arrow[r, "j"'] & E \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \arrow[u, dashed, "h"'] \end{tikzcd}$$

commutes. Let $\pi : E \longrightarrow M$ be the projection onto M, so that $\pi j = \mathrm{id}_M$. Now if we set $g := \pi h$,

$$\begin{tikzcd} & M \arrow[r] & E \arrow[l, bend right, "\pi"'] \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \arrow[u, "h"'] \end{tikzcd}$$

$$\begin{aligned} gi &= \pi h i && \text{by definition} \\ &= \pi j f && \text{by commutativity} \\ &= f && \text{because } \pi j = \mathrm{id}_M\,. \end{aligned}$$

$\square$

We can reduce injectivity to the ability to extend maps from ideals to the entire ring.

Theorem 11.24 (Baer Criterion)**.** *An R-module E is injective if and only if every R-module homomorphism $I \longrightarrow E$ from an ideal I in R can be extended to the whole ring, meaning that there exists g making the diagram*

$$\begin{tikzcd} & E & \\ 0 \arrow[r] & I \arrow[u, "f"] \arrow[r] & R \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commute.

Proof. On the one hand, if E is injective then our condition is simply a special case of the definition of injective module. On the other hand, suppose that this condition holds, and consider any diagram

$$\begin{tikzcd} & E & \\ 0 \arrow[r] & M \arrow[u, "f"] \arrow[r] & N. \end{tikzcd}$$

To simplify notation, let's assume our map $M \longrightarrow N$ is indeed the inclusion of the submodule M, so we can write $m \in N$ for the image of m in N. Consider the set

$$X := \{(A, g) \mid A \text{ is a submodule of } N, M \subseteq A \subseteq N, \text{ and } g \text{ extends } f\}.$$

First, notice X is nonempty, since $(M, f) \in X$. Moreover, we can partially order X by setting $(A, g) \leqslant (B, h)$ if $A \subseteq B$ and $h|_A = g$. So we have a non-empty partially ordered set; let's show we can apply Zorn's Lemma to it.

Given a chain in X, meaning a sequence

$$(A_1, g_1) \leqslant (A_2, g_2) \leqslant \cdots$$

of nested submodules $A_1 \subseteq A_2 \subseteq \cdots$ and maps g_i that extend all g_j with $j \leqslant i$, let $A := \bigcup_i A_i$, and define

$$\begin{aligned} A &\xrightarrow{\quad g \quad} E \\ a &\longrightarrow g_i(a) \text{ if } a \in A_i. \end{aligned}$$

This map g is indeed a map of R-modules, since so are all the g_i, and it is well-defined, since the $g_i(a) = g_j(a)$ whenever $a \in A_i \cap A_j$. By construction, this map extends all the g_i, so we conclude that (A, g) is an upper bound for our chain. Moreover, $M \subseteq A \subseteq N$ follows immediately from our construction, and since each g_i extends f, so does g. We conclude that $(A, g) \in X$, and more generally that any chain in X has an upper bound in X. So Zorn's Lemma applies.

By Zorn's Lemma, X has a maximal element, say (A, g). Now we claim that $A = N$. Suppose not, and let $n \in N$ be an element not in A. One can easily check that

$$I := \{r \in R \mid rn \in A\}$$

is an ideal in R, and that

$$\begin{aligned} I &\xrightarrow{\ h \ } E \\ r &\longrightarrow g(rn) \end{aligned}$$

is an R-module homomorphism. By assumption, we can extend h to an R-module homomorphism $R \longrightarrow E$, which we will write as h as well. Now the R-module homomorphism

$$\begin{aligned} A + Rn &\xrightarrow{\quad \varphi \quad} E \\ a + rn &\longrightarrow g(a) + h(r) \end{aligned}$$

is well-defined by construction, since any $rn \in A$ satisfies $g(rn) = h(r)$, and if $rn = r'n$ then $h(r) = rn = r'n = h(r')$. Finally, this map agrees with g on A, and thus it agrees with f on

M, so $(A+Rn, \varphi) \in X$ and $(A, g) \leqslant (A+Rn, \varphi)$. By the maximality of (A, g), we conclude that $A + Rn = A$, and thus $n \in A$, which is a contradiction. We conclude that $A = N$. Therefore, g makes the diagram

$$\begin{tikzcd} & E & \\ 0 \arrow[r] & M \arrow[u, "f"] \arrow[r] & N. \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commute. $\square$

Over a Noetherian ring, a direct sum of injective modules is injective.

Corollary 11.25. *Let R be a Noetherian ring. If $\{M_j\}_{j\in J}$ are all injective R-modules, then so is $\bigoplus_{j\in J} M_j$.*

Proof. By Theorem 11.24, it is enough to show that any R-module map

$$\begin{tikzcd} & \bigoplus_{j\in J} M_j & \\ 0 \arrow[r] & I \arrow[u, "f"] \arrow[r] & R \end{tikzcd}$$

from an ideal I into $\bigoplus_{j\in J} M_j$ extends to R. Since R is Noetherian, I is finitely generated, so let $I = (a_1, \ldots, a_n)$. For each $i = 1, \ldots, n$, $f(a_i) = (b_{i,j})_{j\in J}$ has $b_{i,j} \neq 0$ only for finitely many values of $j \in J$. Then

$$K := \{j \in J \mid f(a_i)_j \neq 0 \text{ for some } i = 1, \ldots, n\}$$

is a finite set, and $f(I) \subseteq \bigoplus_{j\in K} M_j$. Direct sums of finitely many modules coincide with their product, so by Lemma 11.22, $\bigoplus_{j\in K} M_j$ is injective. Therefore, there exists g such that

$$\begin{tikzcd} & \bigoplus_{k\in K} M_k & \\ 0 \arrow[r] & I \arrow[u, "f"] \arrow[r] & R \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commutes. Now $\bigoplus_{k\in K} M_k$ is a submodule of $\bigoplus_{j\in J} M_j$, so we can think of g as an R-module homomorphism with codomain $\bigoplus_{j\in J} M_j$, and

$$\begin{tikzcd} & \bigoplus_{j\in J} M_j & \\ 0 \arrow[r] & I \arrow[u, "f"] \arrow[r] & R \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commutes. $\square$

It is very easy to see that every R-module is a quotient of a free module. The dual statement is true as well, but it is a little more delicate.

Definition 11.26. An R-module D is **divisible** if for every nonzero $r \in R$ and every $d \in D$ there exists $b \in D$ such that $rb = d$.

Remark 11.27. Given $r \in R$, and an R-module M, the multiplication by r map $M \xrightarrow{\cdot r} M$ is an R-module homomorphism. The module M is divisible if and only if multiplication by r is surjective for all nonzero $r \in R$.

Example 11.28. If R is a domain, the fraction field $\operatorname{Frac}(R)$ of R is an injective R-module.

Remark 11.29. We claim that any quotient of a divisible module is also divisible. Indeed, given $r \in R$ and any class $\overline{d} \in D/E$, let $d \in D$ be a lift of $\overline{d}$. By assumption, there exists $e \in D$ such that $re = d$. The image $\overline{e}$ of e in D/E is still a solution to $r\overline{e} = \overline{d}$.

Lemma 11.30. *Over a domain, every injective module is divisible.*

Proof. Suppose that E is an injective R-module, where R is a domain. Fix $r \in R$ and $a \in E$. Since R is a domain, we have $sr = s'r \implies s = s'$ for any $s, s', r \in R$, so the map of R-modules

$$\begin{aligned} (r) &\longrightarrow E \\ sr &\longrightarrow sa \end{aligned}$$

is well-defined. Since E is injective, we can extend this to a homomorphism $f \colon R \longrightarrow E$. Finally, $f(1) \in E$ is an element such that $e = f(r) = rf(1)$, and E is divisible. $\square$

This not true in general if we do not assume R is a domain.

Example 11.31. Let k be a field and $R = k[x]/(x^2)$. On the one hand, R is not a divisible R-module, since there is no $y \in R$ such that $xy = 1$. On the other hand, R is actually an injective module over itself, although we do not yet have the tools to justify that this is indeed an injective R-module.[1]

The converse of Lemma 11.30 does not hold in general.

Example 11.32. Consider the domain $R = k[x, y]$ and its fraction field Q. The R-module $M = Q/R$ is divisible but not injective.

The converse of Lemma 11.30 does hold for some special classes of rings.

Lemma 11.33. *Let R be a principal ideal domain.*

a) An R-module E is injective if and only if it is divisible.

b) Quotients of injective modules are injective.

Proof.

[1] Using fancy words we haven't learned yet, this ring R is an example of a complete intersection, which is a subclass of Gorenstein rings. One thing we do know how to justify is that $\dim R = 0$. Now it turns out that Gorenstein rings of dimension 0 are injective modules over themselves.

a) Given Lemma 11.30, we only need to show that divisible modules are injective. So let E be a divisible R-module, and consider any map $I \longrightarrow E$ from an ideal I to E. If $I = 0$, we could extend our map by taking the 0 map from R to E, so we might as well assume that $I \neq 0$. By assumption, $I = (a)$ for some $a \in R$, and since E is divisible, there exists $e \in E$ such that $f(a) = ae$. Now consider the multiplication by r map,

$$\begin{aligned} R &\xrightarrow{\;g\;} E \\ r &\longrightarrow re. \end{aligned}$$

For every $r \in R$, $g(ra) = rae = rf(a) = f(ra)$, so g extends f. Therefore, by Theorem 11.24, E is injective.

b) if E is injective, it is also divisible, by Lemma 11.30. Given any submodule $D \subseteq E$, any $e \in E$, and a nonzero $r \in R$, there exists $y \in E$ such that $ry = e$, and so this also holds in E/D. Then E/D is divisible, and thus injective by a.

□

Given an injective abelian group, we can always construct an injective R-module over our favorite ring R.

Lemma 11.34. *Given an injective abelian group D, $\operatorname{Hom}_{\mathbb{Z}}(R, D)$ is an injective R-module.*

Proof. Let $E := \operatorname{Hom}_{\mathbb{Z}}(R, D)$. This abelian group E is an R-module, via

$$r \cdot f := (a \mapsto f(ra)).$$

We claim that E is actually an injective R-module. To show that, we will prove that $\operatorname{Hom}_R(-, \operatorname{Hom}_{\mathbb{Z}}(R, D))$ is an exact functor, which is sufficient by Theorem 11.21. By Corollary 10.31, $\operatorname{Hom}_R(-, \operatorname{Hom}_{\mathbb{Z}}(R, D))$ is naturally isomorphic to $\operatorname{Hom}_{\mathbb{Z}}(- \otimes_{\mathbb{Z}} R, D)$. This last functor is the composition of

$$\operatorname{Hom}_{\mathbb{Z}}(- \otimes_{\mathbb{Z}} R, D) = \operatorname{Hom}_{\mathbb{Z}}(-, D) \circ (- \otimes_{\mathbb{Z}} R).$$

On the one hand, $- \otimes_{\mathbb{Z}} R$ is naturally isomorphic to the identity on R-**mod**, by Lemma 10.22, so it is exact. On the other hand, D is an injective $\mathbb{Z}$-module, so $\operatorname{Hom}_{\mathbb{Z}}(-, D)$ is exact by Theorem 11.21. The composition of exact functors is exact, and thus $\operatorname{Hom}_R(-, \operatorname{Hom}_{\mathbb{Z}}(R, D))$ is exact. □

Theorem 11.35. *Every R-module M is a submodule of some injective R-module.*

Proof. First, let us show that M includes in some injective abelian group. On the one hand, M is a quotient of some free abelian group, say $M \cong (\oplus_i \mathbb{Z})/K$. Now $\mathbb{Z}$ embeds in $\mathbb{Q}$, and thus M embeds into a quotient of $\oplus_i \mathbb{Q}$. By Example 11.28, $\mathbb{Q}$ is an injective abelian group, and by Corollary 11.25, $\oplus_i \mathbb{Q}$ is an injective abelian group. By Lemma 11.33, any quotient of $\oplus_i \mathbb{Q}$ is also injective, so we have shown that M embeds into an injective abelian group, say D. Let $i \colon M \longrightarrow D$ be the inclusion map.

Let $E := \operatorname{Hom}_{\mathbb{Z}}(R, D)$. By Lemma 11.34, E is an injective R-module. Since Hom is left exact, by Theorem 10.7, we have an inclusion $\operatorname{Hom}_{\mathbb{Z}}(R, M) \subseteq \operatorname{Hom}_{\mathbb{Z}}(R, D)$. Now consider the map

$$\begin{aligned} M &\xrightarrow{\psi} \operatorname{Hom}_{\mathbb{Z}}(R, M) \\ m &\longrightarrow (r \mapsto rm). \end{aligned}$$

This is an R-module homomorphism, since

- Given $a, b \in M$,

  $$\psi(a+b)(r) = r(a+b) = ra + rb = \psi(a)(r) + \psi(b)(r),$$

  so $\psi(a+b) = \psi(a) + \psi(b)$.
- Given $r \in R$, $m \in M$, and $s \in R$,

  $$\psi(rm)(s) = s(rm) = r(sm) = r\psi(m)(s),$$

  so $\psi(rm) = r\psi(m)$.

Moreover, if $\psi(m) = 0$ then $m = \psi(m)(1) = 0$. So ψ is injective, and thus composing ψ with our previous inclusion $\operatorname{Hom}_{\mathbb{Z}}(R, M) \subseteq \operatorname{Hom}_{\mathbb{Z}}(R, D)$ gives us an inclusion φ of M into the injective R-module $\operatorname{Hom}_{\mathbb{Z}}(R, D)$. However, the inclusion $\operatorname{Hom}_{\mathbb{Z}}(R, M) \subseteq \operatorname{Hom}_{\mathbb{Z}}(R, D)$ is only a map of abelian groups, so we should check that φ is indeed R-linear. For each $m \in M$, $\varphi(m)$ is the map $\varphi_m \colon R \longrightarrow D$ given by

$$\varphi_m(r) = i(rm).$$

For every $r \in R$, $m \in M$, and $s \in R$,

$$\begin{aligned} \varphi_{rm}(s) &= i(s(rm)) && \text{by definition} \\ &= i(r(sm)) \\ &= r \cdot i(sm) && \text{by definition of } R\text{-module structure on } \operatorname{Hom}_{\mathbb{Z}}(R, D) \\ &= r\varphi_m(s) && \text{by definition,} \end{aligned}$$

so $\varphi(rm) = r\varphi(m)$. □

We can even do better and talk about the smallest injective module that M embeds in.

Definition 11.36. Let $M \subseteq E$. We say E is an *essential extension* of M if every nonzero submodule $N \subseteq E$ intersects M nontrivially, meaning $E \cap M \neq 0$. More generally, an injective map $\alpha \colon M \longrightarrow E$ is an essential extension if $\alpha(M) \subseteq E$ is an essential extension in the sense above.

Exercise 38. The property of being an essential extension is transitive, meaning that if $A \subseteq B$ and $B \subseteq C$ is an essential extension, then $A \subseteq C$ is an essential extension.

Moreover, if $A \subseteq B$ is an essential extension, and C is a submodule of B with $A \subseteq C \subseteq B$, then $A \subseteq C$ and $C \subseteq B$ are both essential extensions.

A module M always has the trivial essential extension $M \subseteq M$. If only has more interesting essential extensions when it is not injective.

Theorem 11.37. *An R-module M is injective if and only if it has no proper essential extensions.*

Proof. Suppose that M is injective, and that $M \subseteq E$ is an essential extension. By Theorem 11.40, the short exact sequence

$$0 \longrightarrow M \longrightarrow E \longrightarrow E/M \longrightarrow 0$$

splits, so $E = M \oplus N$ for some submodule N. If $N \neq 0$, we must have $N \cap M \neq 0$, which is a contradiction. Therefore, $E = M$.

Now suppose M has no nontrivial essential extensions, and let E be an injective module such that $M \subseteq E$, which exists by Theorem 11.35. We claim that the extension $M \subseteq E$ is essential, and thus we must have $M = E$. Suppose not, so that there exists some nonzero submodule $N \subseteq E$ with $N \cap M = 0$. Then the set

$$S := \{N \subseteq E \text{ submodule } \mid N \cap M = 0\}$$

is nonempty. It is also partially ordered by inclusion, and we claim that we can apply Zorn's Lemma to find a maximal element in S. To see that, consider a chain

$$N_1 \subseteq N_2 \subseteq \cdots$$

of submodules of E such that $M \cap N_i \neq 0$. Then $N := \bigcup_i N_i$ is a submodule of E, and any element in $N \cap M$ must be in $N_i \cap M$ for some i, so $N \cap M = 0$. This shows that our increasing chain has an upper bound, and thus Zorn's Lemma says there is a maximal element in S. More precisely, there exists a submodule N of E maximal with respect to the property that $N \cap M = 0$, meaning that if $A \supsetneq N$, then $A \cap M \neq 0$.

Now consider the canonical projection map $\pi \colon E \longrightarrow E/N$. We assumed that $N \cap M = 0$, and since $N = \ker \pi$, π is injective. For simplicity, identify M with its image in E/N via π. Submodules of E/N are of the form A/N for some submodule $A \supseteq N$ of E, and if $A/N \cap M = 0$ then we must have $A \cap M = 0$. We conclude that $M \cong \pi(M) \subseteq E/N$ is an essential extension. By assumption, this essential extension must be trivial, and thus $\pi \colon M \longrightarrow E/N$ is surjective. Finally, this shows that $E \cong N \oplus M$, so M is a direct summand of an injective module. By Lemma 11.23, M is an injective module. □

Theorem 11.38. *Given R-modules $M \subseteq E$, the following are equivalent:*

1) $M \subseteq E$ is a maximal essential extension, meaning that if $M \subseteq E \subseteq N$ is such that $M \subseteq N$ is an essential extension, then $E = N$.

2) $M \subseteq E$ is an essential extension and E is injective.

3) E is injective and for every injective R-module I, if $M \subseteq I \subseteq E$ then $I = E$.

Proof. If 1 holds, and $E \subseteq N$ is an essential extension, then so is $M \subseteq N$, by Exercise 38. Therefore, E has no nontrivial essential extensions, and it must be injective by Theorem 11.37.

Suppose 2 holds, and let I be an injective R-module with $M \subseteq I \subseteq E$. Then $I \subseteq E$ is an essential extension, by Exercise 38, but since I is injective, Theorem 11.37 says we must have $I = E$.

Now if 3 holds, let us show that $M \subseteq E$ is an essential extension. First, consider the set

$$S := \{N \text{ submodule of } E \mid M \subseteq N \text{ is an essential extension}\}.$$

This set is nonempty, since it contains M, and it is partially ordered by inclusion. Given an ascending chain

$$N_1 \subseteq N_2 \subseteq \cdots$$

in S, let $N := \bigcup_i N_i$. This is a submodule of E and it contains M. Given a nonzero submodule $A \subseteq N$, take any nonzero $a \in A$. Then $a \in N_i$ for some i, so $Ra \subseteq N_i$ and we must have $Ra \cap M \neq 0$. Therefore, $A \cap M \neq 0$, and $M \subseteq N$ is an essential extension.

Now let us show that N has no proper essential extensions. We know it has no proper essential extensions within E, but not that it has no proper essential extensions elsewhere. Let $i : N \longrightarrow E$ denote the inclusion map, and suppose that $j : N \subseteq E'$ is an essential extension. Since E is injective, there exists g such that

$$\begin{tikzcd} & E & \\ 0 \arrow[r] & N \arrow[u, "i"] \arrow[r, "j"'] & E' \arrow[ul, "g"'] \end{tikzcd}$$

commutes. Since $N \subseteq E'$ is an essential extension, $\ker g \cap N = \ker g \cap \operatorname{im} j \neq 0$. But $gj = i$ is injective, so g must also be injective. So we have submodules $N \subseteq g(E') \subseteq E$, and since N is a maximal essential extension of M inside E', $M \subseteq g(E')$ is not an essential extension. On the other hand, $N \subseteq E'$ is an essential extension, and since g is injective, $N \subseteq g(E')$ is also an essential extension. By Exercise 38, $M \subseteq g(E')$ must also be an essential extension, which is a contraction unless $j(N) = E'$.

Ultimately, this says that N has no nontrivial essential extensions. By Theorem 11.37, N must be injective. But by assumption, that implies that $N = E$. So this shows that $M \subseteq E$ is an essential extension and also that it is maximal. $\square$

So now we can define the smallest injective module that M embeds in.

Definition 11.39. Let M be an R-module. An R-module $E(M)$ is said to be an **injective hull** or **injective envelope** of M if $M \subseteq E(M)$ is a maximal essential extension.

Exercise 39. Any two injective hulls of M must be isomorphic.

The story of the structure of injective modules then continues in a beautiful way. Over a Noetherian ring, it turns out that every injective module can be decomposed into a direct sum of injective modules of the form $E(R/P)$, where P is a prime ideal in R. Moreover, the injective modules $E(R/P)$ are the indecomposable injective modules, so the basic building blocks of injective modules. We can in fact compute the injective hull of any finitely generated R-module very explicitly. This story begins in Eben Matlis' beautiful PhD thesis [Mat58].

And finally, just like we did for projectives, we can characterize injectives in terms of split short exact sequences.

Theorem 11.40. *An R-module I is injective if and only if every short exact sequence*

$$0 \longrightarrow I \longrightarrow B \longrightarrow C \longrightarrow 0$$

splits.

Proof. Let I be an injective R-module, and consider any short exact sequence

$$0 \longrightarrow I \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0.$$

Since I is injective, there exists a map g making

$$\begin{tikzcd} & I \\ 0 \arrow[r] & I \arrow[u, equal] \arrow[r, "i"'] & B \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commute, and such a g gives a splitting for our short exact sequence.

Conversely, suppose that every short exact sequence $0 \longrightarrow I \longrightarrow B \longrightarrow C \longrightarrow 0$ splits, and consider a diagram

$$\begin{tikzcd} & I \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B. \end{tikzcd}$$

By Theorem 11.35, I embeds into some injective R-module E, say by the inclusion j. By assumption, the short exact sequence

$$0 \longrightarrow I \xrightarrow{j} E \longrightarrow \operatorname{coker} j \longrightarrow 0$$

splits, so there exists a map $q\colon E \longrightarrow I$ such that $qi = \mathrm{id}_I$. Since E is injective, we can lift i through jf, obtaining an R-module homomorphism ℓ such that

$$\begin{tikzcd} & I \arrow[r, "j"'] & E \arrow[l, bend right, "q"'] \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B \arrow[u, dashed, "\ell"'] \end{tikzcd}$$

commutes. Now $g := q\ell$ satisfies

$$\begin{aligned} gi =& q\ell i && \text{by definition} \\ =& qjf && \text{by commutativity} \\ =& f && \text{since } qj = \mathrm{id}_I, \end{aligned}$$

so

$$\begin{tikzcd} & I \\ 0 \arrow[r] & A \arrow[u, "f"] \arrow[r, "i"'] & B. \arrow[ul, dashed, "g"'] \end{tikzcd}$$

commutes. $\square$

11.3 Flat modules

Finally, we turn to the modules that make tensor exact.

Definition 11.41. An R-module M is said to be **flat** if $M \otimes_R -$ is an exact functor.

Remark 11.42. By Theorem 10.28, $M \otimes_R -$ is right exact. Therefore, M is flat if and only if for every injective R-module map $i\colon A \longrightarrow B$, $M \otimes_R A \xrightarrow{1\otimes i} M \otimes_R B$ is injective.

Lemma 11.43. *Given a family of R-modules $\{M_i\}_{i\in I}$, $\oplus_i M_i$ is flat if and only if every M_i is flat.*

Proof. Given a family of R-module homomorphisms $f_i\colon M_i \longrightarrow N_i$, there is an R-module homomorphism

$$\begin{aligned} \bigoplus\nolimits_{i\in I} M_i &\xrightarrow{(f_i)_{i\in I}} \bigoplus\nolimits_{i\in I} N_i \\ (m_i) &\longmapsto (f_i(m_i)) \end{aligned}$$

which is injective if and only if every f_i is injective.

Let $f : A \longrightarrow B$ be an injective R-module homomorphism. There is a commutative diagram

$$\begin{tikzcd} \left(\bigoplus\limits_{i\in I} M_i\right) \otimes_R A \arrow[r, "\cong"] \arrow[d, "\varphi:=1\otimes f"'] & \bigoplus\limits_{i\in I} M_i \otimes_R A \arrow[d, "(1\otimes f)_i=:\psi"] \\ \left(\bigoplus\limits_{i\in I} M_i\right) \otimes_R B \arrow[r, "\cong"'] & \bigoplus\nolimits_{i\in I} M_i \otimes_R B \end{tikzcd}$$

where the horizontal maps are the isomorphisms from Theorem 10.27. In particular, φ is injective if and only if ψ is injective. Moreover, ψ is injective if and only if each component is injective, meaning $1 \otimes f\colon M_i \otimes A \longrightarrow M_i \otimes B$ is injective for all i.

On the one hand, $\bigoplus_{i\in I} M_i$ is flat if and only if for every injective map f, the corresponding map ψ is injective. On the other hand, all the M_i are flat if and only if for every injective maps f, $1 \otimes f\colon M_i \otimes A \longrightarrow M_i \otimes B$ is injective for all i, or equivalently, as explained above, if φ is injective for any given injective map f. This translates into the equivalence we want to show. $\square$

All projectives are flat.

Theorem 11.44. *Every projective R-module is flat.*

Proof. First, recall that $R\otimes_R-$ is naturally isomorphic to the identity functor, by Lemma 10.22, and thus exact (see Remark 10.6). This shows that R is flat, and thus any free module, being a direct sum of copies of R, must also be flat, by Lemma 11.43. Finally, every projective module is a direct summand of a free module, by Theorem 11.14. Direct summands of flat modules are flat, by Lemma 11.43, so every projective module is flat. $\square$

We can test whether a given module if flat by looking at the finitely generated submodules.

Theorem 11.45. *If every finitely generated submodule of M is flat, then M is flat.*

Proof. Let $i\colon A \longrightarrow B$ be an injective map of R-modules. We want to show that

$$M \otimes_R A \xrightarrow{1\otimes i} M \otimes_R B$$

is injective. Suppose that $u \in \ker(1_M \otimes i)$. We are going to construct a finitely generated submodule $j : N \subseteq M$ and an element $v \in N \otimes_R M$ such that $v \in \ker(1_N \otimes i)$ and $u = (j \otimes 1_A)(v)$. Once we do that, our assumption will say that $1_N \otimes i$ is injective, so $v = 0$ and thus $u = 0$. This will show that $1_M \otimes i$ is also injective, and M is flat.

Let's say that $u = m_1 \otimes a_1 + \cdots + m_n \otimes a_n$. In Theorem 10.13, we constructed the tensor product $M \otimes_R B$ as a quotient of the free module F on $M \times B$ by the submodule S with all the necessary relations we need to impose. This gives us a short exact sequence

$$0 \longrightarrow S \longrightarrow F \xrightarrow{\pi} M \otimes_R B \longrightarrow 0.$$

The fact that $m_1 \otimes i(a_1) + \cdots + m_n \otimes i(a_n) = 0$ means we can rewrite this element as $\pi(s)$ for some $s \in S$. This element s is a linear combination of elements of finitely many $(m, b) \in M \times B$. Let $n_1, \ldots, n_s$ be all the M-coordinates of those elements.

Now we take N to be the finitely generated submodule of M generated by $m_1, \ldots, m_n$ and $n_1, \ldots, n_s$, and $v = m_1 \otimes a_1 + \cdots + m_n \otimes a_n \in N \otimes A$. Now

$$(j \otimes 1_A)(m_1 \otimes a_1 + \cdots + m_n \otimes a_n) = m_1 \otimes a_1 + \cdots + m_n \otimes a_n \in M \otimes_R A,$$

and

$$(1_N \otimes i)(m_1 \otimes a_1 + \cdots + m_n \otimes a_n) = m_1 \otimes i(a_1) + \cdots + m_n \otimes i(a_n) = 0,$$

as desired.

The reason we needed to add in these extra elements n_i is that a priori $N \otimes B$ is not necessarily a submodule of $M \otimes B$, so we do not necessarily have $m_1 \otimes i(a_1) + \cdots + m_n \otimes i(a_n) = 0$ in $(Rm_1 + \cdots + Rm_n) \otimes B$ without adding in all relations that make it true. □

Definition 11.46. Let R be a domain and M be an R-module. The **torsion submodule** of M is

$$T(M) := \{m \in M \mid rm = 0 \text{ for some regular element } r \in R\}.$$

The elements of $T(M)$ are called **torsion elements**, and we say that M **is torsion** if $T(M) = M$. Finally, M is **torsion free** if $T(M) = 0$.

Lemma 11.47. *If R is a domain and M is a flat R-module, then M is torsion free.*

Proof. Let $Q = \operatorname{frac}(R)$ be the fraction field of R, which is a torsion free R-module. Now $M \otimes_R Q$ is a Q-vector space, so isomorphic to a direct sum of copies of Q. In particular, $M \otimes_R Q$ is torsion free as an R-module. Since M is flat, the inclusion $R \subseteq Q$ induces an injective R-module map

$$0 \longrightarrow M \otimes_R R \longrightarrow M \otimes_R Q,$$

and since $M \cong M \otimes_R R$, by Lemma 10.22, we conclude that M is isomorphic to a submodule of $M \otimes_R Q$. Submodules of torsion free modules are also torsion free, so $M \otimes_R Q$ is torsion free. □

In general, the converse does not hold.

Example 11.48. Let k be a field and $R = k[x, y]$. Consider the ideal $\mathfrak{m} = (x, y)$. This is a submodule of the torsion free module R, and thus torsion free. However, it is not flat. For example, when we apply $R/\mathfrak{m} \otimes_R -$ to the inclusion $\mathfrak{m} \subseteq R$ we obtain a map of $R/\mathfrak{m}$-vector spaces

$$\mathfrak{m}/\mathfrak{m}^2 \longrightarrow R/\mathfrak{m} .$$

This map cannot possibly be injective: $\mathfrak{m}/\mathfrak{m}^2$ is a 2-dimensional $R/\mathfrak{m}$-vector space, and $R/\mathfrak{m}$ is 1-dimensional.

The converse does over a PID.

Lemma 11.49. *If R is a principal ideal domain, an R-module M is flat if and only if it is torsion free.*

Proof. Suppose M is a torsion free finitely generated R-module. The structure theorem for PIDs says that M must be isomorphic to a direct sum of copies of cyclic modules. The cyclic module R/I has torsion (all the elements are killed by I) unless $I = 0$. Therefore, M must be isomorphic to a direct sum of copies of R, and thus free. By Theorem 11.3 and Theorem 11.44, M is flat.

Now let M be any torsion free R-module. All of the finitely generated submodules of R are also torsion free, and thus flat. $\square$

Not all flat modules are projective.

Example 11.50. The $\mathbb{Z}$-module $\mathbb{Q}$ is torsion free and thus flat, by Lemma 11.49. However, $\mathbb{Q}$ is not a projective $\mathbb{Z}$-module. Suppose $\mathbb{Q}$ is a projective $\mathbb{Z}$-module. By Theorem 11.14, $\mathbb{Q}$ must be a direct summand of a free module, say $F = \bigoplus_I \mathbb{Z}$. Consider the inclusion $\iota\colon \mathbb{Q} \hookrightarrow F$, and pick $i \in I$ such that the image of $\mathbb{Q}$ contains some element with a nonzero entry in the i component. Now consider the projection $\pi : F \longrightarrow \mathbb{Z}$ onto the ith factor. By assumption, the composition $\pi i\colon \mathbb{Q} \longrightarrow \mathbb{Z}$ is nonzero. However, there are no nontrivial abelian group homomorphisms $\mathbb{Q} \longrightarrow \mathbb{Z}$, contradicting the fact that πi is nonzero. We conclude that $\mathbb{Q}$ is not projective.

We now turn to an important example: localization. In particular, we can finally prove Theorem 4.25, a theorem we claimed a long time ago.

Theorem 11.51 (Flatness of localization)**.** *Let R be a ring, and $W \ni 1$ a multiplicative subset of R. Then*

1) *For every R-module M, there is an isomorphism $W^{-1}R \otimes_R M \cong W^{-1}M$ of $W^{-1}R$-modules, and given an R-module map $M \xrightarrow{\alpha} N$, $W^{-1}R \otimes \alpha$ corresponds to $W^{-1}\alpha$ under these isomorphisms. In fact, we have a natural isomorphism between $W^{-1}(-)$ and $W^{-1}R \otimes_R -$.*
2) *$W^{-1}R$ is flat over R.*
3) *$W^{-1}(-)$ is an exact functor; i.e., it sends exact sequences to exact sequences.*

Proof.

1) The bilinear map

$$\begin{aligned} W^{-1}R \times M &\longrightarrow W^{-1}M \\ (\tfrac{r}{w}, m) &\longmapsto \tfrac{rm}{w} \end{aligned}$$

induces a map ψ from the tensor product that is clearly surjective. For an inverse map, set $\phi(\frac{m}{w}) := \frac{1}{w} \otimes m$. To see this is well-defined, suppose $\frac{m}{w} = \frac{m'}{w'}$, so there exists some $v \in W$ such that $v(mw' - m'w) = 0$. Then,

$$\phi\left(\frac{m}{w}\right) - \phi\left(\frac{m'}{w'}\right) = \frac{1}{w} \otimes m - \frac{1}{w'} \otimes m'.$$

We can multiply through by $\frac{vww'}{vww'}$ to get

$$\frac{vw'}{vww'} \otimes m - \frac{vw}{vww'} \otimes m' = \frac{1}{vww'} \otimes v(mw' - m'w) = 0.$$

To see this is a homomorphism, we note that

$$\begin{aligned} \phi(\frac{m}{w} + \frac{m'}{w'}) = \phi(\frac{mw' + m'w}{ww'}) &= \frac{1}{ww'} \otimes (mw' + m'w) = \frac{1}{ww'} \otimes mw' + \frac{1}{ww'} \otimes m'w \\ &= \frac{w'}{ww'} \otimes m + \frac{w}{ww'} \otimes m' = \frac{1}{w} \otimes m + \frac{1}{w'} \otimes m' = \phi\left(\frac{m}{w}\right) + \phi\left(\frac{m'}{w'}\right), \end{aligned}$$

and

$$\phi\left(r\frac{m}{w}\right) = \frac{1}{w} \otimes rm = r\left(\frac{1}{w} \otimes m\right) = r\phi\left(\frac{m}{w}\right).$$

The composition $\phi \circ \psi$ sends

$$\frac{r}{w} \otimes m \mapsto \frac{rm}{w} \mapsto \frac{1}{w} \otimes rm = \frac{r}{w} \otimes m.$$

Since this is the identity on simple tensors, it must be the identity.

For the claim about maps, we need check that $\psi_N \circ (W^{-1}R \otimes \alpha) = W^{-1}\alpha \circ \psi_M$ for every $M \xrightarrow{\alpha} N$. And indeed,

$$\begin{aligned} (\psi_N \circ (W^{-1}R \otimes \alpha))\left(\frac{r}{w} \otimes m\right) = \psi_N\left(\frac{r}{w} \otimes \alpha(m)\right) &= \frac{r\alpha(m)}{w} \\ &= \frac{\alpha(rm)}{w} = W^{-1}\alpha\left(\frac{rm}{w}\right) = (W^{-1}\alpha \circ \psi_M)\left(\frac{r}{w} \otimes m\right). \end{aligned}$$

Finally, we note that our isomorphisms $W^{-1}R \otimes_R M \cong W^{-1}M$ give a natural isomorphism between the localization functor W^{-1} and the tensor functor $W^{-1}R \otimes_R -$. Indeed, given a map of R-modules $M \xrightarrow{f} N$, the diagram

$$\begin{tikzcd} W^{-1}R \otimes M \arrow[r] \arrow[d, "\mathrm{id} \otimes f"'] & W^{-1}M \arrow[d, "W^{-1}(f)"] \\ W^{-1}R \otimes N \arrow[r] & W^{-1}N \end{tikzcd}$$

commutes, since it commutes for simple tensors:

$$\begin{tikzcd}
\frac{r}{w} \otimes m \arrow[r] \arrow[d, "\mathrm{id} \otimes f"'] & \frac{rm}{w} \arrow[d, "W^{-1}(f)"] \\
\frac{r}{w} \otimes f(m) \arrow[r] & \frac{rf(m)}{w} = \frac{f(rm)}{w}.
\end{tikzcd}$$

2) This follows from the earlier observation that $W^{-1}(-)$ preserves injective maps, Remark 4.23.

3) This is immediate from part 2). $\square$

Corollary 11.52 (Hom and localization)**.** *Let R be a Noetherian ring, W be a multiplicative set, M be a finitely generated R-module, and N an arbitrary R-module. Then,*

$$\mathrm{Hom}_{W^{-1}R}(W^{-1}M, W^{-1}N) \cong W^{-1}\,\mathrm{Hom}_R(M, N).$$

In particular, if $\mathfrak{p}$ is prime,

$$\mathrm{Hom}_{R_\mathfrak{p}}(M_\mathfrak{p}, N_\mathfrak{p}) \cong \mathrm{Hom}_R(M, N)_\mathfrak{p}.$$

Chapter 12

Resolutions

12.1 Projective resolutions

Definition 12.1. Let M be an R-module. A **projective resolution** is a complex

$$C_\bullet = \cdots \longrightarrow \underset{n}{P_n} \longrightarrow \cdots \longrightarrow \underset{1}{P_1} \longrightarrow \underset{0}{P_0} \longrightarrow 0$$

where all the P_i are projective, $\mathrm{H}_0(C) = M$, and $\mathrm{H}_i(C) = 0$ for all $i \neq 0$. We may also write a projective resolution for M as an exact sequence

$$\cdots \longrightarrow \underset{n}{P_n} \longrightarrow \cdots \longrightarrow \underset{1}{P_1} \longrightarrow \underset{0}{P_0} \longrightarrow M \longrightarrow 0$$

where all the modules P_i are projective.

You'll find both these definitions in the literature, often indicating the second option as an abuse of notation. We will be a bit sloppy and consider both equivalently, since at the end of the day they contain the same information.

Theorem 12.2. *Every R-module has a projective resolution.*

Proof. Let M be an R-module. We are going to construct a projective resolution quite explicitly. The first step is to find a projective module P_0 that surjects onto M. In fact, we can find a free module surjecting onto M, by Lemma 11.19. Now consider the kernel of that projection, say

$$0 \longrightarrow K_0 \xrightarrow{i_0} P_0 \xrightarrow{\pi_0} M \longrightarrow 0.$$

Set $\partial_0 := \pi_0$. There exists a free module P_1 surjecting onto K_0. Now the map $\partial_1 = i_0\pi_1$ satisfies $\operatorname{im}\partial_1 = K_0 = \ker\partial_0$.

$$\begin{tikzcd}
0 \arrow[dr] & & 0 & \\
& K_0 \arrow[ur, "i_0"] \arrow[dr] & & \\
P_1 \arrow[ur, "\pi_1"] \arrow[rr, "\partial_1"'] & & P_0 \arrow[r, "\partial_0"] & M.
\end{tikzcd}$$

Now the process continues analougously. We find a free module P_2 surjecting onto $K_1 := \ker\, \partial_1$, and set

$$\begin{tikzcd}
 & & 0 \arrow[dr] & & 0 & \\
 & & & K_0 \arrow[ur] \arrow[dr, "i_0"] & & \\
P_2 \arrow[rr, "\partial_2"] \arrow[dr, "\pi_2"'] & & P_1 \arrow[ur, "\pi_1"] \arrow[rr, "\partial_1"'] & & P_0 \arrow[r, "\partial_0"] & M. \\
 & K_1 \arrow[ur, "i_1"'] \arrow[dr] & & & & \\
0 \arrow[ur] & & 0. & & &
\end{tikzcd}$$

At each stage, $\pi_i \colon P_i \longrightarrow K_{i-1}$ is a surjective map, $K_i := \ker\, \partial_i$, i_i is the inclusion of the kernel of ∂_i into P_i, and we get short exact sequences

$$0 \longrightarrow K_{n+1} \xrightarrow{i_{n+1}} P_{n+1} \xrightarrow{\pi_{n+1}} K_n \longrightarrow 0.$$

In fact, $\operatorname{im}(i_{n+1}) = \ker \partial_{n+1} = \ker(i_n \pi_{n+1}) = \ker \pi_{n+1}$. We can continue this process indefinitely for as long as $P_n \neq 0$, and the resulting sequence will be a projective resolution for M. $\square$

In fact, we showed that every R-module has a free resolution.

Definition 12.3. A **free resolution** of an R-module M is a projective resolution for M by free modules, meaning each P_i is free.

We can think of a free resolution

$$\cdots \longrightarrow F_2 \longrightarrow F_1 \longrightarrow F_0 \longrightarrow M$$

as giving a detailed description of our module M. The first free module, F_0, gives us generators for M. The second free module, F_1, gives us generators for all the relations among our generators for M. The next module describes the relations among the relations among our generators. And so on.

More interestingly, we can do this minimally as long as some reasonable assumptions are satisfied.

Definition 12.4. Let $(R, \mathfrak{m})$ be either a local ring or an $\mathbb{N}$-graded graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. A complex

$$\cdots \longrightarrow F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0 \longrightarrow \cdots$$

is **minimal** if $\operatorname{im} \partial_{n+1} \subseteq \mathfrak{m} F_n$ for all n.

Remark 12.5. A complex

$$F = \cdots \longrightarrow F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0 \longrightarrow \cdots$$

is minimal if and only if the differentials in the complex $F \otimes_R R/\mathfrak{m}$ are all identically 0. If all the F_i are free, our complex is minimal if and only if all the entries in the matrices representing ∂_i are in $\mathfrak{m}$, whatever our bases are.

Lemma 12.6. *Let $(R, \mathfrak{m})$ be either a local ring or a $\mathbb{N}$-graded graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. Let M be a (graded) finitely generated R-module. A free resolution*

$$F = \cdots \longrightarrow F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0$$

for M is a minimal complex if and only if for all n, F_n is constructed by taking the free module on a minimal set of (homogeneous, in the graded case) generators for $\ker \partial_{n-1}$.

Proof. Suppose for some n we have chosen F_n to be the free module on some non-minimal set of generators $m_1, \ldots, m_s$ for $K_{n-1} := \ker \partial_{n-1}$. More precisely, there is a basis $e_1, \ldots, e_s$ for F_n so that $\partial_n(e_i) = m_i$, and the images of $m_1, \ldots, m_s$ in the vector space $K_{n-1}/\mathfrak{m}K_{n-1}$ are linearly dependent. Then we can choose $r_1, \ldots, r_s \in R$, not all in $\mathfrak{m}$, such that $r_1 m_1 + \cdots + r_s m_s = 0$ in R. In the graded case, we can take all these coefficients r_i to be homogeneous. At least one of these coefficients is not in $\mathfrak{m}$, and thus it must be invertible.[1] we can multiply by its inverse. So perhaps after reordering our elements, we get

$$m_s = r_1 m_1 + \cdots + r_{s-1} m_{s-1}.$$

Now

$$e_s - r_1 e_1 - \cdots - r_{s-1} e_{s-1} \in \ker \partial_n = \operatorname{im} \partial_{n+1}$$

is not in $\mathfrak{m}F_n$, so $\operatorname{im} \partial_{n+1} \notin \mathfrak{m}F_n$.

Now suppose that $\operatorname{im} \partial_{n+1} \nsubseteq \mathfrak{m}F_n$ for some n. Let $e_1, \ldots, e_s$ be a basis for F_n, so that $\partial_n(e_1), \ldots, \partial(e_s)$ form a generating set for $K_{n-1} := \ker \partial_{n-1}$. By assumption, $\ker \partial_n = \operatorname{im} \partial_{n+1}$ contains some (homogeneous, in the graded case) element that is not in $\mathfrak{m}F_n$. So there is an element $r_1 e_1 + \cdots + r_s e_s \in \ker \partial_n$ not in $\mathfrak{m}F_n$. In particular, some $r_i \notin \mathfrak{m}$, which we can assume without loss of generality to be r_1. Multiplying by the inverse of r_1, we get some $c_i \in R$ such that

$$e_1 - c_2 e_2 - \cdots - c_s e_s \in \ker \partial_n,$$

so

$$\partial_n(e_1) = c_2 \partial_n(e_2) + \cdots + c_s \partial_n(e_s).$$

This is a nontrivial relation among our chosen set of generators of K_{n-1}, which must then be non-minimal. □

So to construct a minimal free resolution of M, we simply take as few generators as possible in each step. Ultimately, we can talk about *the* minimal free resolution of M. To show that, we need some definitions and a lemma.

[1] In the graded case, homogeneous elements not in $\mathfrak{m}$ must be nonzero elements in R_0, and thus invertible.

Definition 12.7. Let (F, ∂) and (G, δ) be complexes of R-modules. The **direct sum** of F and G is the complex of R-modules $F \oplus G$ that has $(F \oplus G)_n = F_n \oplus G_n$, with differentials given by

$$\begin{array}{ccc} F_{n+1} & \xrightarrow{\partial_{n+1}} & F_n \\ \oplus & & \oplus \\ G_{n+1} & \xrightarrow[\delta_n]{} & D_n, \end{array}$$

together with the complex maps $F \hookrightarrow F \oplus G$ and $G \hookrightarrow F \oplus G$ given by the corresponding inclusion in each homological degree.

Remark 12.8. The direct sum of complexes is the coproduct in the category $\mathrm{Ch}(R)$.

Remark 12.9. The homology of a direct sum is the direct sum of the homologies, since

$$(\partial_n, \delta_n)(a, b) = (0, 0) \Leftrightarrow \partial_n(a) = 0 \text{ and } \delta_n(b) = 0,$$

and $(a, b) \in \operatorname{im}(\partial_n, \delta_n)$ if and only if $a \in \operatorname{im} \partial_n$ and $b \in \operatorname{im} \partial_n$, so

$$\mathrm{H}_n(F \oplus G) = \frac{\ker(\partial_n, \delta_n)}{\operatorname{im}(\partial_{n+1}, \delta_{n+1})} = \frac{\ker \partial_n}{\operatorname{im} \partial_{n+1}} \oplus \frac{\ker \delta_n}{\operatorname{im} \delta_{n+1}}.$$

Remark 12.10. Suppose that C is a subcomplex of D, and that we know that each C_n is a direct summand of D_n, say by $D_n = C_n \oplus B_n$. In order for C to be a free summand of D, we also need that the differentials of D behave well with C: for each n, we need that $\partial(B) \subseteq B$ and $\partial(C) \subseteq C$.

Definition 12.11. A complex C of R-modules is **trivial** if it is a direct sum of complexes of the form

$$\cdots \longrightarrow 0 \longrightarrow R \xrightarrow{1} R \longrightarrow 0 \longrightarrow \cdots.$$

Example 12.12. The complex

$$0 \longrightarrow R \xrightarrow{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} R^2 \xrightarrow{\begin{pmatrix} 0 & 1 \end{pmatrix}} R \longrightarrow 0 \qquad = \qquad \begin{array}{c} 0 \longrightarrow R \xrightarrow{1} R \longrightarrow 0 \\ \oplus \\ 0 \longrightarrow R \xrightarrow{1} R \longrightarrow 0 \end{array}$$

is trivial.

Remark 12.13. Trivial complexes are exact, since they are the direct sums of exact complexes, by Remark 12.9.

Lemma 12.14. *Let $(R, \mathfrak{m})$ be either a local ring or a $\mathbb{N}$-graded graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. Every (graded) complex*

$$\cdots \longrightarrow T_2 \xrightarrow{\partial_2} T_1 \xrightarrow{\partial_1} T_0 \longrightarrow 0$$

of finitely generated (graded) free R-modules that is exact everywhere must be trivial.

Proof. Since T_0 is projective, Theorem 11.11 says that the short exact sequence

$$0 \longrightarrow \ker \partial_1 \longrightarrow T_1 \xrightarrow{\partial_1} T_0 \longrightarrow 0$$

splits, so $T_1 \cong \ker \partial_1 \oplus T_0$. In fact, ∂_1 is the map $T_0 \oplus \ker \partial_1$ given by $(1,0)$, and our original exact sequence breaks off as

$$\begin{array}{c} \cdots \longrightarrow T_2 \xrightarrow{\partial_2} \ker \partial_1 \longrightarrow 0 \\ \oplus \\ 0 \longrightarrow T_0 \xrightarrow{1} T_0 \longrightarrow 0. \end{array}$$

In particular, since $0 \longrightarrow T_0 \xrightarrow{1} T_0 \longrightarrow 0$ is trivial and homology commutes with taking direct sums of complexes, by Remark 12.9, we conclude that

$$\cdots \longrightarrow T_2 \xrightarrow{\partial_2} \ker \partial_1 \longrightarrow 0$$

is also exact everywhere. We've also shown that $\ker \partial_1$ is a (graded) direct summand of the (graded) free R-module T_1. In the local case, $\ker \partial_1$ is a projective by Theorem 11.14, and thus free by Theorem 11.17. In the graded setting, Theorem 11.18 says that $\ker \partial_1$ is free. So we are back at our original situation, and we can repeat the same argument repeatedly to show that our complex breaks off as the direct sum of the trivial complexes

$$0 \longrightarrow \ker \partial_n \xrightarrow{1} \ker \partial_n \longrightarrow 0$$

and must therefore be trivial. □

Theorem 12.15. *Let*

$$P = \ \cdots \longrightarrow P_n \longrightarrow \cdots P_1 \xrightarrow{\partial_1} P_0 \xrightarrow{\partial_0} M \longrightarrow 0$$

be a complex of projective R-modules, and let

$$C = \ \cdots \longrightarrow C_n \longrightarrow \cdots C_1 \xrightarrow{\delta_1} C_0 \xrightarrow{\delta_0} N \longrightarrow 0$$

be an exact complex. Every (graded) R-module map $M \xrightarrow{f} N$ lifts to a map of complexes $P \xrightarrow{\varphi} C$, and any two such lifts are homotopic.

Moreover, when R is an $\mathbb{N}$-graded graded k-algebra with $R_0 = k$, M and N are finitely generated graded R-modules, P_n and C_n are finitely generated graded R-modules, and f is a graded homomorphism, the induced map of complexes is made out of graded R-module maps.

Proof. Since P_0 is projective and δ_0 is surjective, there exists an R-module homomorphism φ_0 such that

$$\begin{tikzcd} P_0 \arrow[r, "\partial_0"] \arrow[d, dashed, "\varphi_0"'] & M \arrow[r] \arrow[d, "f"] & 0 \\ C_0 \arrow[r, "\delta_0"'] & N \arrow[r] & 0 \end{tikzcd}$$

commutes. Notice in fact that

$$\begin{aligned}
\delta_0\varphi_0(\operatorname{im}\partial_1) &\subseteq \delta_0\varphi_0(\ker\partial_0) && \text{because } P \text{ is a complex}\\
&= f\partial_0(\ker\partial_0) && \text{by commutativity of the square above}\\
&= 0,
\end{aligned}$$

so $\varphi_0(\operatorname{im}\partial_1) \subseteq \ker\delta_0 = \operatorname{im}\delta_1$.

In the graded case, note that we can define φ_0 by sending the elements b_i in a homogeneous basis of P_0 to homogeneous $c_i \in C_0$ such that $\delta_0(c_i) = f\partial_0(b_i)$.

We now proceed by induction. Suppose we have constructed $P_{n-1} \xrightarrow{\varphi_{n-1}} C_{n-1}$ such that $\varphi_{n-1}(\operatorname{im}\partial_n) \subseteq \operatorname{im}\delta_n$. Since P_n is projective, there exists a map φ_n such that

$$\begin{tikzcd}
P_n \arrow[r, "\partial_n"] \arrow[d, dashed, "\varphi_n"'] & P_{n-1} \arrow[d, "\varphi_{n-1}"] \\
C_n \arrow[r, "\delta_n"'] & \operatorname{im}\delta_n
\end{tikzcd}$$

commutes. And again,

$$\begin{aligned}
\delta_n\varphi_n(\operatorname{im}\partial_{n+1}) &\subseteq \delta_n\varphi_n(\ker\partial_{n-1}) && \text{because } P \text{ is a complex}\\
&= \varphi_{n+1}\partial_n(\ker\partial_n) && \text{by commutativity of the square above}\\
&= 0,
\end{aligned}$$

so $\varphi_n(\operatorname{im}\partial_{n+1}) \subseteq \operatorname{im}\delta_n$. By induction, we obtain our map of complexes φ lifting f.

Now suppose we are given two such maps of complexes $P \longrightarrow C$ lifting f, say φ and ψ. Note that $\varphi - \psi$ and 0 are two liftings of the 0 map. We are going to show that any map lifting the 0 map $M \longrightarrow N$ must be nullhomotopic, which will then imply that φ and ψ are homotopic as well (essentially via the same homotopy!).

So let $\varphi \colon P \longrightarrow C$ be a map of complexes lifting the 0 map $M \longrightarrow N$.

$$\begin{tikzcd}
\cdots P_1 \arrow[r, "\partial_1"] \arrow[d, "\varphi_1"'] & P_0 \arrow[r, "\partial_0"] \arrow[d, "\varphi_0"'] & M \arrow[r] \arrow[d, "0"] & 0 \\
\cdots C_1 \arrow[r, "\delta_1"] & C_0 \arrow[r, "\delta_0"] & N \arrow[r] & 0
\end{tikzcd}$$

We will explicitly construct a nullhomotopy for φ using induction. Set $h_n = 0$ for all $n < 0$. The commutativity of the rightmost square says that $\delta_0\varphi_0 = 0$, so $\operatorname{im}\varphi_0 \subseteq \ker\delta_0 = \operatorname{im}\delta_1$. Since P_0 is projective, there exists an R-module homomorphism h_0 such that

$$\begin{tikzcd}
 & P_0 \arrow[dl, "h_0"'] \arrow[d, "\varphi_0"] \\
C_1 \arrow[r, "\delta_1"'] & \operatorname{im}\delta_1
\end{tikzcd}$$

commutes, and thus $\varphi_0 = \delta_1 h_0 + h_{-1}\partial_0$. Notice also that

$$\begin{aligned}
\delta_1(\varphi_1 - h_0\partial_1) &= \varphi_0\partial_1 - \delta_1 h_0\partial_1 && \text{because } \varphi \text{ is a map of complexes}\\
&= (\varphi_0 - \delta_1 h_0)\partial_1\\
&= 0 && \text{since } \varphi_0 = \delta_1 h_0,
\end{aligned}$$

so $\operatorname{im}(\varphi_1 - h_0\partial_1) \subseteq \ker \delta_1 = \operatorname{im} \delta_2$.

Now assume that we have constructed maps $h_0, \ldots, h_n$ such that $\varphi_n = h_{n-1}\partial_n + \delta_{n+1}h_n$ and $\operatorname{im}(\varphi_{n+1} - h_n\partial_{n+1}) \subseteq \operatorname{im} \delta_{n+2}$. Since P_{n+1} is projective, we can find a map h_{n+1} such that

$$\begin{tikzcd} & P_{n+1} \arrow[dl, "h_{n+1}"'] \arrow[d, "\varphi_{n+1} - h_n\partial_{n+1}"] \\ C_{n+2} \arrow[r, "\delta_{n+2}"'] & \operatorname{im} \delta_{n+2} \end{tikzcd}$$

commutes. Now

$$\begin{aligned} \delta_{n+2}(\varphi_{n+2} - h_{n+1}\partial_{n+2}) &= \varphi_{n+1}\partial_{n+2} - \delta_{n+2}h_{n+1}\partial_{n+2} && \text{since } \varphi \text{ is a map of complexes} \\ &= (\varphi_{n+1} - \delta_{n+2}h_{n+1})\partial_{n+2} \\ &= h_n\partial_{n+1}\partial_{n+2} = 0 && \text{since } \partial_{n+1}\partial_{n+2} = 0. \end{aligned}$$

So we again obtain $\operatorname{im}(\varphi_{n+2} - h_{n+1}\partial_{n+2}) \subseteq \ker \delta_{n+1} = \operatorname{im} \delta_{n+2}$. By induction, this process allows us to construct our homotopy h. $\square$

Theorem 12.16. *Let $(R, \mathfrak{m})$ be either a local ring or a $\mathbb{N}$-graded graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. If F is a minimal free resolution of M, any free resolution for M is isomorphic to a direct sum of F with a trivial complex. In particular, the minimal free resolution of M is unique up to isomorphism.*

Proof. Suppose that G is another free resolution of M. By Theorem 12.15, there are complex maps $\psi \colon G \longrightarrow F$ and $\varphi \colon F \longrightarrow G$ that lift the identity map on M. Then $\psi\varphi \colon F \longrightarrow F$ is a map of complexes that lifts the identity on M, and thus by Theorem 12.15 $\varphi\psi$ must be homotopic to the identity on F. Let h be a homotopy between $\psi\varphi$ and the identity, so that for all n,

$$\operatorname{id} - \psi_n\varphi_n = \partial_{n+1}h_n + h_{n-1}\partial_n.$$

Since F is minimal, we have $\operatorname{im} \partial_n \subseteq \mathfrak{m}F_{n-1}$ and $\operatorname{im} \partial_{n+1} \subseteq \mathfrak{m}F_n$, so $\operatorname{im}(\operatorname{id} - \psi_n\varphi_n) \subseteq \mathfrak{m}F_n$ for all n.

First we do the local case. Let A be the matrix representing $\psi_n\varphi_n$ in some fixed basis for F_n, and note that $\operatorname{id} - \psi_n\varphi_n$ is represented by $\mathrm{Id} - A$, so all the entries in $\mathrm{Id} - A$ must be in $\mathfrak{m}$. Our matrix A can be written as

$$A = \begin{pmatrix} 1 - a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & 1 - a_{22} & \cdots & a_{2s} \\ & \ddots & & \\ a_{s1} & \cdots & a_{s\,s-1} & 1 - a_{ss} \end{pmatrix}$$

for some $a_{ij} \in \mathfrak{m}$, so that $\det(A) = 1 + a$ for some $a \in \mathfrak{m}$. In particular, $\det(A)$ is invertible, and $\psi_n\varphi_n$ is an isomorphism.

In the graded case, we have to be a bit more careful: not all elements not in $\mathfrak{m}$ are invertible, only *homogeneous* ones. First, we fix $f_1, \ldots, f_s$ a basis of homogeneous elements for F_n with $\deg(f_1) \leqslant \deg(f_2) \leqslant \cdots \deg(f_s)$, and set $\Phi := \operatorname{id} - \psi_n\varphi_n$. Since our map Φ is degree-preserving, $\Phi(f_i)$ is homogeneous for each i, and so we can write $\Phi(f_i)$ as a linear combination of our basis elements $f_1, \ldots, f_s$ using only pieces of degree $\deg(\Phi(f_i))$. We obtain a matrix $C = (c_{ij})$ such that $c_{ij} \neq 0 \implies \deg(c_{ij}) = \deg(f_j) - \deg(f_i)$, and C represents Φ, meaning $\Phi(f_i) = c_{i1} f_1 + \cdots + c_{is} f_s$ for all i. Now all the entries of $C = \mathrm{Id} - A$ must be in $\mathfrak{m}$, so in particular we must have $a_{ii} = 1$ for all i. Moreover, since we chose our basis to have increasing degrees, $\deg(c_{ij}) = 0$ whenever $i < j$. Since we must also have $c_{ij} \in \mathfrak{m}$ whenever $i \neq j$, we conclude that $c_{ij} = 0$ for $i < j$. We conclude that A is an upper triangular matrix. Finally, $\det(A) = a_{11} \cdots a_{ss} = 1$, and A is invertible.

So we have shown in both cases that $\psi_n \varphi_n$ is an isomorphism for all n. By Exercise 24, $\psi\varphi$ is in fact an isomorphism of complexes, so let $F \xrightarrow{\xi} F$ be its inverse. Now we want to claim that φ splits as a map of complexes. Notice that $(\xi\psi)\varphi = \xi(\psi\varphi) = \mathrm{id}_F$, so $\xi\psi$ will be our proposed splitting for φ. We immediately get a splitting at the level of R-modules, since $G_n = \varphi(F_n) \oplus \ker(\xi_n \psi_n)$, so $\xi\psi$ provides at least splittings for the R-modules in each degree; we just need to prove that $G = \varphi(F) \oplus \ker(\xi\psi)$ as complexes. To do that, let $K := \ker(\xi\psi)$, and denote the differential in G by δ. We need to check that the differential δ satisfies $\delta(\varphi(F)) \subseteq \varphi(F)$ and $\delta(K) \subseteq K$.

Since φ is a map of complexes, $\delta\varphi = \varphi\partial$, so we do get $\delta(\varphi(F)) \subseteq \varphi(F)$. Given $a \in K_{n+1}$, we can write $\delta_{n+1}(a) = \varphi(b) + c$ for some $b \in F_n$ and K_n, since $G_n = \varphi(F_n) \oplus K_n$. Then

$$
\begin{aligned}
b &= \mathrm{id}(b) & \\
&= \xi_n \psi \varphi_n(b) & \text{since } \xi_n\psi_n \text{ is a splitting for } \varphi_n \\
&= \xi_n \psi_n(\varphi_n(b) + c) & \text{since } c \in K_n \\
&= \xi_n \psi_n \delta_{n+1}(a) & \text{by assumption} \\
&= \xi_n \delta_{n+1} \psi_n(a) & \text{since } \psi \text{ is a map of complexes} \\
&= \delta_{n+1}(\xi_n \psi_n)(a) & \text{since } \xi \text{ is a map of complexes} \\
&= 0 & \text{since } a \in K_n.
\end{aligned}
$$

We conclude that $\delta_{n+1}(a) \in K_n$, and $\delta(K) \subseteq K$. We have now shown that $G \cong F \oplus K$.

Finally, we are going to show that K is a trivial complex. First, we claim that K_n is free for all n. We have shown K_n is a (graded) direct summand of a (graded) free module. In the local case, Theorem 11.14 says that K_n is projective, and then Theorem 11.17 says that K_n must in fact be free. In the graded setting, Theorem 11.18 guarantees that K_n is free.

Since $G \cong F \oplus K$, we have $\mathrm{H}_n(G) \cong \mathrm{H}_n(F) \oplus \mathrm{H}_n(K)$. Since F and G are both (graded) free resolutions for M, so they have the same homology: $\mathrm{H}_n(F) = \mathrm{H}_n(G) = 0$ for all $n \neq 0$, and $\mathrm{H}_0(F) = \mathrm{H}_0(G) = M$. We conclude that K is exact everywhere. Finally, Lemma 12.14 shows that K is trivial. $\square$

We also want to keep track of the kernels of the differentials in a minimal free resolution.

Definition 12.17. Let $(R, \mathfrak{m})$ be either a local ring or an $\mathbb{N}$-graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. Let F be a minimal free resolution for the finitely generated (graded) R-module M. For each $n \geqslant 1$, the submodule

$$\Omega_n(M) := \operatorname{im} \partial_n = \ker \partial_{n-1}$$

is the n**th syzygy** of M.

Remark 12.18. For each n, we have a short exact sequence

$$0 \longrightarrow \ker \partial_n \longrightarrow F_n \longrightarrow \operatorname{im} \partial_n \longrightarrow 0.$$

But $\ker \partial_n = \Omega_n(M)$ and $\operatorname{im} \partial_n = \Omega_{n-1}(M)$, so we get a short exact sequence

$$0 \longrightarrow \Omega_n(M) \longrightarrow F_n \longrightarrow \Omega_{n-1}(M) \longrightarrow 0.$$

Syzygies are indeed well-defined up to isomorphism.

Remark 12.19. Suppose that F and G are two minimal free resolutions for M. By Theorem 12.16, there exists an isomorphism between F and G, say φ. Since φ is a map of complexes, $\varphi\partial^F = \partial^G\varphi$, and thus φ must send elements in $\ker \partial^F$ into elements in $\ker \partial^G$. Similarly, an inverse ψ to φ sends $\ker \partial^G$ into $\ker \partial^F$. In each homological degree, the induced maps $\ker \partial_n^F \longrightarrow \ker \partial_n^G$ and $\ker \partial_n^F \longrightarrow \ker \partial_n^G$ are inverse, and thus isomorphisms. In the graded case, one can show that we obtain graded isomorphisms, so that the graded syzygies are also well-defined up to isomorphism.

The number of generators in each homological degree is also an important invariant.

Definition 12.20. Let $(R, \mathfrak{m})$ be either a local ring or a $\mathbb{N}$-graded graded k-algebra with $R_0 = k$ and homogeneous maximal ideal $\mathfrak{m} = R_+$. Let F be a minimal free resolution for the finitely generated (graded) R-module M. The nth betti number of M is

$$\beta_i(M) := \text{rank } F_i = \mu(F_i).$$

In the graded case, the (i, j)th betti number of M, $\beta_{ij}(M)$, counts the number of generators of F_i in degree j. We often collect the betti numbers of a module in its **betti table**:

$$\begin{array}{r|cccc}
\beta(M) & 0 & 1 & 2 & \cdots \\
\hline
0 & \beta_{00}(M) & \beta_{01}(M) & \beta_{02}(M) & \\
1 & \beta_{11}(M) & \beta_{12}(M) & \beta_{13}(M) & \\
2 & \beta_{22}(M) & \beta_{23}(M) & & \\
\vdots & & & \ddots &
\end{array}$$

By convention, the entry corresponding to (i, j) in the betti table of M contains $\beta_{i,i+j}(M)$, and *not* $\beta_{ij}(M)$. This is how Macaulay2 displays betti tables as well, using the command `betti`.

Example 12.21. Suppose that $R = k[x, y, z]$ and that $M = R/(xy, xz, yz)$ corresponds to the variety defining the union of the three coordinate lines in $\mathbb{A}^3_k$. This variety has dimension 1 and degree 3. The minimal free resolution for M is

$$0 \longrightarrow R^2 \xrightarrow{\begin{pmatrix} z & 0 \\ -y & y \\ 0 & -x \end{pmatrix}} R^3 \xrightarrow{\begin{pmatrix} xy & xz & yz \end{pmatrix}} R \longrightarrow M.$$

From this minimal resolution, we can read the betti numbers of M:

- $\beta_0(M) = 1$, since M is a cyclic module;
- $\beta_1(M) = 3$, and these three quadratic generators live in degree 2;
- $\beta_2(M) = 2$, and these represent linear syzygies on quadrics, and thus live in degree 3.

Here is the graded free resolution of M:

$$0 \longrightarrow R(-3)^2 \xrightarrow{\begin{pmatrix} z & 0 \\ -y & y \\ 0 & -x \end{pmatrix}} R(-2)^3 \xrightarrow{\begin{pmatrix} xy & xz & yz \end{pmatrix}} R \longrightarrow M.$$

Notice that the graded shifts in lower homological degrees affect all the higher homological degrees as well. For example, when we write the map in degree 2, we only need to shift the degree of each generator by 1, but since our map now lands on $R(-2)^3$, we have to bump up degrees from 2 to 3, and write $R(-3)^2$. The graded betti number $\beta_{ij}(M)$ of M counts the number of copies of $R(-j)$ in homological degree i in our resolution. So we have

$$\beta_{00} = 1, \beta_{12} = 3, \text{and } \beta_{23} = 2.$$

We can collect the graded betti numbers of M in what is called a *betti table*:

$\beta(M)$	0	1	2
0	1	–	–
1	–	3	2

.

Example 12.22. Let k be a field, $R = k[x, y]$, and consider the ideal

$$I = (x^2, xy, y^3)$$

which has two generators of degree 2 and one of degree 3, so there are graded betti numbers β_{12} and β_{13}. The minimal free resolution for R/I is

$$0 \longrightarrow \begin{matrix} R(-3)^1 \\ \bigoplus \\ R(-4)^1 \end{matrix} \xrightarrow{\begin{pmatrix} y & 0 \\ -x & y^2 \\ 0 & -x \end{pmatrix}} \begin{matrix} R(-2)^2 \\ \bigoplus \\ R(-3)^1 \end{matrix} \xrightarrow{\begin{pmatrix} x^2 & xy & y^3 \end{pmatrix}} R \longrightarrow R/I.$$

$$\beta_{23}(R/I) = 1 \qquad \beta_{12}(R/I) = 2$$
$$\beta_{24}(R/I) = 1 \qquad \beta_{13}(R/I) = 1$$

So the betti table of R/I is

$\beta(M)$	0	1	2
0	1	–	–
1	–	2	1
2	–	1	1

.

We can also extract the Hilbert function of M from its betti numbers.

Theorem 12.23. *Let k be a field and R be an $\mathbb{N}$-graded k-algebra with $R_0 = k$. Let M be a finitely generated $\mathbb{Z}$-graded R-module. Then the Hilbert series of M satisfies*

$$h_M(t) = h_R(t) \sum_{i \geqslant 0, p \in \mathbb{Z}} (-1)^i \beta_{ip}(M) t^p.$$

When $R = k[x_1, \ldots, x_n]$ is a standard graded polynomial ring,

$$h_M(t) = \frac{\displaystyle\sum_{i \geqslant 0} \sum_{d \in \mathbb{Z}} (-1)^i \beta_{id}(M) t^d}{(1-t)^n}.$$

Proof. First, notice that the fact that the resolution is minimal implies that at each stage, the graded shifts in the copies of R in F_i must go up at least by 1. In particular, for each fixed each d, $\beta_{id} = 0$ for all $i \gg 0$.

The graded resolution of M breaks into graded pieces:

$$0 \longrightarrow \cdots \longrightarrow F_{i,d} \longrightarrow F_{i-1,d} \longrightarrow \cdots \longrightarrow F_{o,d} \longrightarrow M_d \longrightarrow 0.$$

Here we denote the piece of F_i in degree d by $F_{i,d}$. Notice that this is in fact finite, since β_{id} is eventually 0. As a simple application of the Rank-Nullity theorem from Linear Algebra, just like we did in Lemma 8.6, we obtain

$$\dim_k(M_d) = \sum (-1)^i \dim_k(F_{i,d}).$$

Multiplying by t^d and summing over all d, we get

$$\begin{aligned} H_M(t) = \sum_d \dim_k(M_d) t^d &= \sum_{i,d} (-1)^i \dim_k(F_{i,d}) t^d \\ &= \sum_{i \geqslant 0} (-1)^i \left(\sum_{i,d} \dim_k(F_{i,d}) t^d \right) = \sum_{i \geqslant 0} (-1)^i h_{F_i}(t). \end{aligned}$$

Now $F_i = \bigoplus_d R(-d)^{\beta_{id}}$, and $h_{R(-d)} = t^d h_R(t)$ by Example 8.5, so $h_{F_i} = \sum_d \beta id(M) t^d h_R(t)$. Then

$$h_M(t) = h_R(t) \sum_{i \geqslant 0} \sum_d (-1)^i \beta_{id}(M) t^d.$$

When $R = k[x_1, \ldots, x_n]$ is a standard graded polynomial ring, Example 8.4 says $h_R(t) = \frac{1}{(1-t)^n}$, so

$$h_M(t) = \frac{\displaystyle\sum_{i \geqslant 0} \sum_d (-1)^i \beta_{id}(M) t^d}{(1-t)^n}. \qquad \square$$

Example 12.24. We can now use the information we collected in Example 12.21 to calculate the Hilbert series of M:

$$h_M(t) = \frac{1t^0 - 3t^2 + 2t^3}{(1-t)^3} = \frac{1+2t}{(1-t)^1}$$

and since this last fraction is in lowest terms, we see that the dimension of M is 1 (the degree of the denominator) and that the degree of M is equal to $p(1) = 1 + 2 \cdot 1 = 3$.

The slogan is that we can get lots of information about M from its minimal free resolution. In fact, even if all we know is the betti numbers of M, there is lots of information to we can extract about M. For more about the beautiful theory of free resolutions and syzygies, see [Eis05]. For a detailed treatment of graded free resolutions, see [Pee11].

12.2 Injective resolutions

Injective resolutions are analogous to projective resolutions, but now we want to approximate our module M by injectives.

Definition 12.25. Let M be an R-module. An **injective resolution** of M is a complex

$$E = \; 0 \longrightarrow E_0 \longrightarrow E_1 \longrightarrow E_2 \longrightarrow \cdots$$

with $\mathrm{H}_0(E) = M$ and $\mathrm{H}_n(E) = 0$ for all $n \neq 0$. We may abuse notation and instead say that an injective resolution of M is an exact sequence

$$0 \longrightarrow M \longrightarrow \quad \longrightarrow E_0 \longrightarrow E_1 \longrightarrow E_2 \longrightarrow \cdots.$$

Remark 12.26. This is the first example we have encountered where we have a *cocomplex* rather than a complex. Its homology should technically be referred to as cohomology, and written with superscripts:

We can construct injective resolutions in a similar fashion to how we constructed projective resolutions.

Theorem 12.27. *Every R-module M has an injective resolution.*

Proof. By Theorem 11.35, every R-module embeds into an injective module. So we start by taking an injective R-module E_0 containing M, and look at the cokernel of the inclusion.

$$0 \longrightarrow M \xrightarrow{i_0} E_0 \xrightarrow{\pi_0} \operatorname{coker} i_0 \longrightarrow 0.$$

Now $\operatorname{coker} \pi_0$ includes in some other injective module E_1.

$$\begin{tikzcd}
0 \arrow[r] & M \arrow[r, "i_0"] & E_0 \arrow[rr, dashed, "\partial_0"] \arrow[dr, "\pi_0"'] & & E_1 \\
& & & \operatorname{coker} i_0 \arrow[ur, "i_1"'] \arrow[dr] & \\
& & 0 \arrow[ur] & & 0
\end{tikzcd}$$

Take $\partial_0 := i_1 \pi_0$. Since i_1 is injective,

$$\ker \partial_0 = \ker(i_1 \pi_0) = \ker \pi_0 = \operatorname{im} i_0.$$

Notice also that $\operatorname{coker} i_0 = \operatorname{im} \partial_0 = \ker(E_1 \longrightarrow \operatorname{coker} \partial_0)$. So we can now we continue in a similar fashion, by finding an injective module E_2 that $\operatorname{coker} \partial_0$ embeds into.

$$\begin{tikzcd}
 & & & & 0 \arrow[dr] & & 0 \\
 & & & & & \operatorname{coker} \partial_0 \arrow[ur] \arrow[dr, "i_2"] & \\
0 \arrow[r] & M \arrow[r, "i_0"] & E_0 \arrow[rr, dashed, "\partial_0"] \arrow[dr, "\pi_0"'] & & E_1 \arrow[rr, dashed, "\partial_1"'] \arrow[ur, "\pi_1"] & & E_2 \\
 & & & \operatorname{coker} i_0 \arrow[ur] \arrow[dr, "i_1"] & & & \\
 & & 0 \arrow[ur] & & 0 & &
\end{tikzcd}$$

By construction and since i_2 is injective, $\ker \partial_1 = \operatorname{im} \partial_0$, and our complex is exact at E_1. The process continues analogously. □

We can again define a minimal free resolution for M as one where at each step we take the injective hull of $\operatorname{coker} i_n$. Perhaps unsurprisingly, one can show that the minimal free resolution of a finitely generated module over a local ring is unique up to isomorphism. The analogues to the betti numbers are called Bass numbers, although now there are some major differences. When we construct a minimal free resolution, we have only to count copies of R in each homological degree, while there are many different building clocks for injective modules — the injective hulls of R/P, where P ranges over the prime ideals in R.

Example 12.28. Let's construct a minimal free resolution for the abelian group $\mathbb{Z}$. We start by including $\mathbb{Z}$ in $\mathbb{Q}$, and then note that the cokernel $\mathbb{Q}/\mathbb{Z}$ is actually injective, by Lemma 11.33. So $\mathbb{Q}/\mathbb{Z}$ embeds in itself, and our resolution stops there. So the short exact sequence

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Q} \longrightarrow \mathbb{Q}/\mathbb{Z} \longrightarrow 0$$

is in fact a minimal injective resolution for $\mathbb{Z}$.

Chapter 13

Abelian categories

An abelian category is a category that has just enough extra structure to behave like R-module: we have complexes and exact sequences, homology, the Snake Lemma, the long exact sequence in homology, and many other nice features. On the one hand, every abelian category embeds nicely in some R-**mod**, so it's in some ways sufficient to study R-**mod**. In other ways, the general nonsense definitions in an abelian category can sometimes give us a uniform, simple way to prove many results about R-**mod** (and $\mathrm{Ch}(R\text{-}\mathbf{mod})$, and other related categories) all at once.

13.1 What's an abelian category?

Definition 13.1. A category $\mathscr{A}$ is a **preadditive category** if:

- For all objects x and y in $\mathcal{A}$, $\mathrm{Hom}_{\mathcal{A}}(x, y)$ is an abelian group.
- For all objects x, y, and z in $\mathcal{A}$, the composition $\mathrm{Hom}_{\mathcal{A}}(x, y) \times \mathrm{Hom}_{\mathcal{A}}(z, x) \xrightarrow{\circ} \mathrm{Hom}_{\mathcal{A}}(z, y)$ is bilinear, meaning

$$g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2 \quad \text{and} \quad (g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f.$$

Example 13.2. Our favorite category R-**mod** is a preadditive category.

Definition 13.3. Let $\mathcal{A}$ and $\mathscr{B}$ be preadditive categories. An **additive functor** $\mathcal{A} \longrightarrow \mathscr{B}$ is a functor such that the map

$$\begin{aligned} \mathrm{Hom}_{\mathcal{A}}(x, y) &\longrightarrow \mathrm{Hom}_{\mathscr{B}}(F(x), F(y)) \\ f &\longmapsto F(f) \end{aligned}$$

is a homomorphism of abelian groups.

This extends our previous definition of additive functors in R-**mod**.

Definition 13.4. Let $\mathscr{C}$ be a category. An **initial object** in $\mathscr{C}$ is an object i such that for every object x in $\mathscr{C}$, $\mathrm{Hom}_{\mathscr{C}}(i, x)$ is a singleton, meaning there exists a unique arrow $i \longrightarrow x$. A **terminal object** in $\mathscr{C}$ is an object t such that for every object x in $\mathscr{C}$, $\mathrm{Hom}_{\mathscr{C}}(x, t)$ is a singleton, meaning there exists a unique arrow $x \longrightarrow t$. A **zero object** in $\mathscr{C}$ is an object that is both initial and terminal.

Exercise 40. Initial objects are unique up to unique isomorphism. Terminal objects are unique up to unique isomorphism.

So we can talk about *the* initial object, *the* terminal object, and *the* zero object, if they exist.

Example 13.5.

a) The 0 module is the zero object in R-**mod**.

b) In the category of rings, $\mathbb{Z}$ is the initial object, but there is no terminal object unless we allow the 0 ring.

c) There are no initial nor terminal objects in the category of fields.

Definition 13.6. Let $\mathscr{C}$ be a category with a zero object 0. Given two objects x and y in $\mathscr{C}$, the **zero arrow** from x to y is the unique arrow $x \longrightarrow y$ that factors through 0, meaning the composition of the unique arrows $x \longrightarrow 0 \longrightarrow y$. We will sometimes denote the 0 arrow by 0.

Notice in particular that if a category $\mathcal{A}$ has a zero object, then $\operatorname{Hom}_{\mathcal{A}}(x, y)$ is always nonempty, since it contains at least the 0 arrow.

Remark 13.7. Composing the 0 arrow with any other arrow always yields the 0 arrow.

Remark 13.8. In any preadditive category $\mathcal{A}$ with a zero object 0, the 0 arrow $x \longrightarrow y$ coincides with the 0 of the abelian group $\operatorname{Hom}_{\mathcal{A}}(x, y)$.

Remark 13.9. We can characterize the 0 object by the property that the zero arrow and the identity arrows on 0 coincide. To see this, notice that if $1_x = x \xrightarrow{0} x$, then given any arrow $x \xrightarrow{f} y$, $f = f \circ 1_x = f \circ 0 = 0$, and similarly any arrow $y \xrightarrow{f} x$ must be 0. Then x is terminal and initial, and it must be the zero object.

Definition 13.10. An **additive category** is a preadditive category $\mathcal{A}$ such that:

- $\mathcal{A}$ has a zero object.
- $\mathcal{A}$ has all finite products, meaning that given any two objects x and y in $\mathcal{A}$, there exists a product of x and y in $\mathcal{A}$.

Lemma 13.11. *In an additive category, finite coproducts exist and they agree with products.*

Proof. Let x and y be objects in our additive category, and consider their product, which we know exists:

$$\begin{tikzcd} & z \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ x & & y \end{tikzcd}$$

The universal property of the product give arrows ι_1 and ι_2 such that

$$\begin{tikzcd}
& x \arrow[d, dashed, "\iota_1"'] \arrow[ddl, bend right, "1_x"'] \arrow[ddr, bend left, "0"] & \\
& z \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
x & & y
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
& y \arrow[d, dashed, "\iota_2"'] \arrow[ddl, bend right, "0"'] \arrow[ddr, bend left, "1_y"] & \\
& z \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
x & & y
\end{tikzcd}$$

both commute. We claim that z together with ι_1 and ι_2 form a coproduct. So given an object w and arrows $x \xrightarrow{f} w$ and $y \xrightarrow{g} w$, we need to show that there exists a unique arrow h such that

$$\begin{tikzcd}
& z \arrow[dd, dashed, "h"'] & \\
x \arrow[ur, "\iota_1"] \arrow[dr, "f"'] & & y \arrow[ul, "\iota_2"'] \arrow[dl, "g"] \\
& w &
\end{tikzcd}$$

commutes.

To see such an h exists, consider $h := f\pi_1 + g\pi_2$. Then

$$h\iota_1 = f\underbrace{\pi_1\iota_1}_{1_x} + g\underbrace{\pi_2\iota_1}_{0} = f \quad \text{and} \quad h\iota_2 = f\underbrace{\pi_1\iota_2}_{0} + g\underbrace{\pi_2\iota_2}_{1_y} = f,$$

so indeed our proposed h does the job.

To show the uniqueness of such an h, we will use the fact that z together with π_1 and π_2 is a product for x and y. So suppose that h' is another arrow such that $h'\iota_1 = f$ and $h'\iota_2 = g$. Then $h - h'$ satisfies $(h - h')\iota_1 = f - f = 0$ and $(h - h')\iota_2 = g - g = 0$, so it's sufficient to show that the 0 arrow is the unique arrow ψ such that

$$\begin{tikzcd}
& z \arrow[dd, dashed, "\psi"'] & \\
x \arrow[ur, "\iota_1"] \arrow[dr, "0"'] & & y \arrow[ul, "\iota_2"'] \arrow[dl, "0"] \\
& w &
\end{tikzcd}$$

commutes. First, we claim that $\iota_1\pi_1 + \iota_2\pi_2$ is the identity arrow on z. And indeed, this map satisfies

$$\pi_1(\iota_1\pi_1 + \iota_2\pi_2) = \underbrace{\pi_1\iota_1}_{1_x}\pi_1 + \underbrace{\pi_2\iota_1}_{0}\pi_2 = \pi_1 \quad \text{and} \quad \pi_2(\iota_1\pi_1 + \iota_2\pi_2) = \underbrace{\pi_2\iota_1}_{0}\pi_1 + \underbrace{\pi_2\iota_1}_{1_y}\pi_2 = \pi_2,$$

and so does the identity arrow 1_z, so the universal property of the product guarantees that $\iota_1\pi_1 + \iota_2\pi_2 = 1_z$. Now if $\psi\iota_1 = 0$ and $\psi\iota_2 = 0$, then

$$\psi = \psi 1_z = \psi(\iota_1\pi_1 + \iota_2\pi_2) = \psi\iota_1\pi + \psi\iota_2\pi_2 = 0 + 0 = 0.$$

$\square$

Notation 13.12. In an additive category $\mathcal{A}$, $A \oplus B$ denotes the product $\equiv$ coproduct of the objects A and B.

Remark 13.13. If $\mathcal{A}$ is additive category, $A \oplus B$ is characterized by the existence of arrows $A \xrightarrow{i_A} A \oplus B$, $B \xrightarrow{i_B} A \oplus B$, $A \oplus B \xrightarrow{\pi_A} A$ and $A \oplus B \xrightarrow{\pi_A} B$ such that $\pi_A i_A = \mathrm{id}_A$, $\pi_B i_B = \mathrm{id}_B$, and $i_A \pi_A + i_B \pi_B = \mathrm{id}_{A \oplus B}$.

Lemma 13.14. *Let $F \colon \mathcal{A} \longrightarrow \mathcal{B}$ an additive functor between additive categories.*

a) If 0 is the 0 object, $F(0) = 0$. For any two objects x and y, $F(x \xrightarrow{0} y) = F(x) \xrightarrow{0} F(y)$.

b) F preserves finite products and coproducts.

Proof. We show the statement assuming F is covariant, and note that the contravariant case is essentially the same.

a) The statement about zero arrows follows immediately from the fact that F_{xy} is a group homomorphism and that the 0 arrow is the 0 element in the abelian group $\mathrm{Hom}_{\mathcal{A}}(F(x), F(y))$. Now the zero arrow and the identity arrows of 0 coincide, and so do their images by F. On the one hand, $F(1_0) = 1_{F(0)}$, and as we have shown, $F(1_0) = F(0) \xrightarrow{0} F(0)$. Then the identity and the zero on $F(0)$ coincide, so by Remark 13.9 we must have $F(0) = 0$.

b) Fix objects A and B and the canonical arrows $A \xrightarrow{i_A} A \oplus B$, $B \xrightarrow{i_B} A \oplus B$, $A \oplus B \xrightarrow{\pi_A} A$ and $A \oplus B \xrightarrow{\pi_A} B$. Any functor preserves identity arrows, so any additive functor F must send

$$F(\pi_A)F(i_A) = F(\pi_A i_A) = \mathrm{id}_{F(A)} \qquad\qquad F(\pi_B)F(i_B) = F(\pi_B i_B) = \mathrm{id}_{F(B)}$$

and

$$F(i_A)F(\pi_A) + F(i_B)F(\pi_B) = \mathrm{id}_{F(A \oplus B)}\,.$$

By Remark 13.13, this implies that $F(A \oplus B)$ is the product $\equiv$ coproduct of $F(A)$ and $F(B)$.

$\square$

Exercise 41. Let $\mathcal{A}$ be an additive category. Show that an arrow f is a mono if and only if $fg = 0$ implies $g = 0$. Similarly, an arrow f is an epi if and only if $gf = 0$ implies $g = 0$.

We can now define kernels and cokernels.

Definition 13.15. Let $\mathcal{A}$ be an additive category and f an arrow $x \longrightarrow y$. The **kernel** of f is an arrow $k \xrightarrow{i} x$ such that

- $k \xrightarrow{i} x \xrightarrow{f} y$ is 0.
- Given any $g \in \mathrm{Hom}_{\mathcal{A}}(z, x)$ such that $z \xrightarrow{g} x \xrightarrow{f} y$ is 0, there exists a unique arrow φ such that $i\varphi = g$, meaning that

$$\begin{tikzcd}
k \arrow[r, "i"] & x \arrow[r, "f"] & y \\
 & z \arrow[ul, dashed, "\exists!\varphi"] \arrow[u, "g"'] \arrow[ur, "0"'] &
\end{tikzcd}$$

commutes. We denote the kernel of f by $\ker f$.

We sometimes refer to the kernel as not just an arrow but the pair (object, arrow). Also, we might use the notation $\ker f \longrightarrow x$ for the kernel of $x \xrightarrow{f} y$. We might also abuse notation and refer to the object that is the source of $\ker f$ as the kernel of f. Nevertheless, the kernel of f is technically an arrow, not an object. A good reason for identifying the arrow $\ker f$ with its source object is the following rewriting of the definition above:

Remark 13.16. If $k_1 \xrightarrow{i_1} x$ and $k_1 \xrightarrow{i_2} x$ are both kernels of f, then there exist unique arrows $k_1 \to k_2$ and $k_2 \to k_1$ such that

$$\begin{tikzcd} k_1 \arrow[r, "i_1"] \arrow[dr, dashed, bend left=10, "h"] & x \arrow[r, "f"] & y \\ & k_2 \arrow[u, "i_2"'] \arrow[ul, dashed, bend left=10, "g"] & \end{tikzcd}$$

commutes. As we will see below in Remark 13.17, kernels are always mono. But then $i_1 gh = i_2 h = i_1$, and since i_1 is a mono, we must have $gh = 1$. Similarly, $hg = 1$, and g and h are isomorphisms.

This shows that if $k \xrightarrow{i} x$ is the kernel of $f \in \mathrm{Hom}_{\mathcal{A}}(x, y)$, the object k is, up to isomorphism, the unique object that satisfies the following universal property: for every object z and every arrow $z \xrightarrow{g} x$ such that $fg = 0$, there exists a unique arrow $z \xrightarrow{h} k$ such that $ih = g$.

Remark 13.17. We claim that a kernel, if it exists, is always a mono. Indeed, suppose that

$$\begin{tikzcd} z \arrow[r, shift left, "g_1"] \arrow[r, shift right, "g_2"'] & k \arrow[r, "i"] & x \arrow[r, "f"] & y \end{tikzcd}$$

are such that $ig_1 = ig_2$. Then $i(g_1 - g_2) = 0$, so it's sufficient to show that $ig = 0$ implies $g = 0$. But then

$$\begin{tikzcd} k \arrow[r, "i"] & x \arrow[r, "f"] & y \\ z \arrow[u, "g"] \arrow[ur, "0"'] & & \end{tikzcd}$$

commutes, and $f \circ 0 = 0$, so 0 factors uniquely through the kernel. But both g and $z \xrightarrow{0} k$ are such factorizations, so $g = 0$.

Definition 13.18. Let $\mathcal{A}$ be an additive category and $f \in \mathrm{Hom}_{\mathcal{A}}(x, y)$. The **cokernel** of f is an arrow $y \xrightarrow{p} c$ such that

- $x \xrightarrow{f} y \xrightarrow{p} c$ is 0.
- Given any $g \in \mathrm{Hom}_{\mathcal{A}}(y, z)$ such that $x \xrightarrow{f} y \xrightarrow{g} z$ is 0, there exists a unique arrow φ such that $i\varphi = g$, meaning that

$$\begin{tikzcd} x \arrow[r, "f"] & y \arrow[r, "p"] \arrow[d, "g"'] & c \arrow[dl, dashed, "\exists! \varphi"] \\ & z & \end{tikzcd}$$

commutes.

We denote the cokernel of f by $\operatorname{coker} f$.

We will sometimes use the notation $y \longrightarrow \operatorname{coker} f$ for the kernel of $x \xrightarrow{f} y$, although once again the cokernel of f is an arrow rather than an object.

Example 13.19.

a) The kernels and cokernels in R**-mod** are what we think they are: the inclusion of the usual kernel, and the projection onto the usual cokernel.

b) It's not always true that all arrows have kernels or cokernels. For example, the category of finitely generated R-modules over some non-Noetherian ring R is additive, but it does not have all kernels. If I is some infinitely generated ideal in R, the kernel of the canonical projection $R \longrightarrow R/I$ does not exist in our category. In fact, this is an epi but not a cokernel — it should be the cokernel of the inclusion map $I \longrightarrow R$, but this is not an arrow in our category.

While not all epis are cokernels and not all monos are kernels, the converse is true. Just like we saw for kernels, cokernels, if they exist, are always epi, and they are unique in the sense we described in Remark 13.16.

Remark 13.20. Let $\mathcal{A}$ be an additive category, and f any arrow such that $\ker \operatorname{coker} f$ and $\operatorname{coker} \ker f$ exist. Since $f \circ \ker f = 0$, f factors uniquely through $\operatorname{coker} \ker f$, say by $\operatorname{coker} \ker f \xrightarrow{g} y$. Now $\operatorname{coker} f \circ g \circ (\operatorname{coker} \ker f) = \operatorname{coker} f \circ f = 0$. Since $\operatorname{coker} \ker f$ is an epi, by Exercise 44, we must have $\operatorname{coker} f \circ g = 0$. Then g factors uniquely through $\ker \operatorname{coker} f$, so we get a unique arrow such that

$$\begin{tikzcd}
\ker f \arrow[r] & x \arrow[r, "f"] \arrow[d] & y \arrow[r] & \operatorname{coker} f \\
& \operatorname{coker} \ker f \arrow[r, dashed] & \ker \operatorname{coker} f \arrow[u] &
\end{tikzcd}$$

commutes.

Definition 13.21. An **abelian category** is an additive category $\mathcal{A}$ such that

- The category $\mathcal{A}$ contains all kernels and cokernels of arrows in $\mathcal{A}$.
- Every mono is a kernel of its cokernel.
- Every epi is the cokernel of its kernel.
- For every f, the canonical arrow $\operatorname{coker} \ker f \longrightarrow \ker \operatorname{coker} f$ is an isomorphism.

Ultimately, an abelian category is one that has just enough structure so that we can extend many of the desired properties of R**-mod**. In particular, we will see that we can define complexes and their homology in any abelian category, and that the Snake Lemma and the long exact sequence in homology hold.

Remark 13.22. Let $\mathcal{A}$ be an abelian category, and f any arrow. As described in Remark 13.20, we have a commutative diagram

$$\begin{tikzcd}
\ker f \arrow[r] & x \arrow[r, "f"] \arrow[d, "p"'] & y \arrow[r] & \operatorname{coker} f \\
& \operatorname{coker} \ker f \arrow[r, "\psi"'] & \ker \operatorname{coker} f \arrow[u, "i"'] &
\end{tikzcd}$$

where ψ is now assumed to be an iso. Now kernels are mono and cokernels are epi, by Exercise 44, and composing an epi (respectively, mono) with an iso gives us an epi (respectively, mono). Therefore, we can factor f as a composition mono $\circ$ epi.

Definition 13.23. Let $\mathcal{A}$ be an abelian category, and consider an arrow $x \xrightarrow{f} y$. The **image** of f is $\operatorname{im} f := \ker(\operatorname{coker} f)$.

Following Remark 13.22, the source of $\operatorname{im} f = \ker \operatorname{coker} f$ is the unique (up to unique isomorphism) object such that f factors as

$$x \xrightarrow{\text{epi}} \operatorname{im} f \xrightarrow{\text{mono}} y\ .$$

Exercise 42. Let $\mathcal{A}$ be an abelian category. Show that f is a mono if and only if $\ker f = 0$, and f is an epi if and only if $\operatorname{coker} f = 0$.

Remark 13.24. If $\mathcal{A}$ is an abelian category, its opposite category $\mathcal{A}^{\mathrm{op}}$ is also abelian.

Example 13.25.

a) The category R**-mod** of R-modules is an abelian category.

b) The category of free R-modules is additive but not abelian, as kernels and cokernels do not exist in general.

c) The category of finitely generated R-modules is abelian if and only if R is Noetherian, which is exactly the condition we need to guarantee the existence of kernels and cokernels. For a general ring R, the category of Noetherian R-modules is abelian.

d) The category of Hilbert spaces with continuous linear functions is an additive category. The monos are injective linear maps, and the epis are maps with dense image. The kernels are the usual kernels, while the cokernel of $f \colon X \longrightarrow Y$ is given by the orthogonal projection $Y \longrightarrow \overline{f(X)}^{\perp}$. However, this is not an abelian category, since a mono might not be the kernel of its cokernel. Indeed, if $X \hookrightarrow Y$ is a dense inclusion that is not surjective, then this mono is not the kernel of its cokernel.

Remark 13.26. Suppose that g factors through f, meaning that there exists h such that

$$\begin{tikzcd}
x \arrow[r, "f"] & y \\
& z \arrow[u, "g"'] \arrow[ul, dashed, "h"]
\end{tikzcd}$$

commutes. Then $(\operatorname{coker} f) \circ g = (\operatorname{coker} f) \circ f \circ h = 0$, so g factors through $\ker(\operatorname{coker} f) = \operatorname{im} f$, meaning

$$\begin{tikzcd}
x \arrow[r, "f"] & y \\
\operatorname{im} f \arrow[ur] & z \arrow[u, "g"'] \arrow[l, dashed, "h"]
\end{tikzcd}$$

also commutes.

Exercise 43. The kernel of $x \xrightarrow{0} y$ is the identity arrow 1_x, its cokernel is the identity arrow 1_y, and $\operatorname{im}(x \xrightarrow{0} y) = 0$.

Exercise 44. Let $\mathcal{A}$ be an abelian category, g an epi, and f a mono. Then $\ker(fg) = \ker g$, $\operatorname{coker}(fg) = \operatorname{coker} f$, and $\operatorname{im}(fg) = \operatorname{im} f = f$.

13.2 Complexes and homology in an abelian category

Definition 13.27. Let $\mathcal{A}$ be an abelian category. A **chain complex** or simply **complex** (C, ∂) (which we sometimes write just a C) is a sequence of objects and arrows

$$\cdots \longrightarrow C_n \xrightarrow{\partial_n} C_{n-1} \longrightarrow \cdots$$

such that $\partial_{n-1}\partial_n = 0$ for all n. A map of complexes $f : C \longrightarrow D$ between two chain complexes is a sequence of arrows f_n such that the diagram

$$\begin{tikzcd}
\cdots \arrow[r] & C_n \arrow[r, "\partial_n"] \arrow[d, "f_n"'] & C_{n-1} \arrow[r] \arrow[d, "f_{n-1}"'] & \cdots \\
\cdots \arrow[r] & D_n \arrow[r, "\partial_n"'] & D_{n-1} \arrow[r] & \cdots
\end{tikzcd}$$

commutes. The **category of (chain) complexes over** $\mathcal{A}$, denoted $\operatorname{Ch}(\mathcal{A})$, is the category that has objects all chain complexes in $\mathcal{A}$ and arrows all the chain complex maps.

Lemma 13.28. *If $\mathcal{A}$ is an abelian category, so is* $\operatorname{Ch}(\mathcal{A})$.

Proof sketch. First, note that $\operatorname{Ch}(\mathcal{A})$ is a preadditive category: given two maps of complexes f and g, $f + g$ is obtained degreewise, by taking $(f + g)_n := f_n + g_n$. The facts that $\operatorname{Hom}_{\operatorname{Ch}(\mathcal{A})}(C, D)$ is an abelian group and that composition is bilinear follows from the analogous facts in $\mathcal{A}$. The 0 object is the 0 complex, which has the 0 object in $\mathcal{A}$ in each degree. Given two complexes C and D, their product is taken degreewise:

$$C \times D = \cdots \longrightarrow C_n \times D_n \xrightarrow{\partial_n^C \times \partial_n^D} C_{n-1} \times D_{n-1} \longrightarrow \cdots$$

and the projection maps in each degree assemble to make a map of complexes. So $\operatorname{Ch}(\mathcal{A})$ is an additive category.

Let $C \xrightarrow{f} D$ be a map of complexes. The universal property of $\ker \partial_n$ gives us a unique arrow δ_{n+1} such that

$$\begin{tikzcd}
\ker f_{n+1} \arrow[r] \arrow[d, dashed, "\delta_{n+1}"'] & C_{n+1} \arrow[r, "f_{n+1}"] \arrow[d, "\partial_{n+1}"'] & D_{n+1} \arrow[d, "\partial_{n+1}"] \\
\ker f_n \arrow[r] & C_n \arrow[r, "f_n"'] & D_{n-1}
\end{tikzcd}$$

commutes. The commutativity of
$$\begin{tikzcd}
\ker f_{n+1} \arrow[r] \arrow[d, dashed, "\delta_{n+1}"'] & C_{n+1} \arrow[d, "\partial_{n+1}"] \\
\ker f_n \arrow[r] \arrow[d, dashed, "\delta_n"'] & C_n \arrow[d, "\partial_n"] \\
\ker f_{n-1} \arrow[r] & C_{n-1}
\end{tikzcd}$$
and the fact that $\partial_n \partial_{n+1} = 0$ together with the fact that $\ker f_{n-1}$ is a mono imply that $\delta_n \delta_{n+1} = 0$. Finally, we conclude that

$$\cdots \longrightarrow \ker f_n \xrightarrow{\delta_n} \ker f_{n-1} \longrightarrow \cdots$$

is a complex in $\mathrm{Ch}(\mathcal{A})$, and the canonical maps $\ker f_n \longrightarrow C_n$ assemble into a map of complexes. One can check that the universal property of the kernels $\ker f_n$ forces this complex we just constructed to be $\ker f$. In particular, $\mathrm{Ch}(\mathcal{A})$ has all kernels. Similarly, we construct cokernels in $\mathrm{Ch}(\mathcal{A})$, building on the fact that $\mathcal{A}$ has all cokernels.

Finally, it remains to show that every mono is the kernel of its cokernel and every epi is the cokernel of its kernel. This boils down to the fact that f is a mono if and only if all the f_n are monos, and dually that f is an epi if and only if all the f_n are epis. The conclusion will then follow from our construction of kernels and cokernels and the fact that $\mathcal{A}$ is abelian. Our claim follows from Exercise 42 and the fact that $f = 0$ if and only if all $f_n = 0$. □

Definition 13.29. Let $\mathcal{A}$ be an abelian category. For each C in $\mathrm{Ch}(\mathcal{A})$, we define its cycles $Z_n(C)$ and boundaries $B_n(C)$ by

$$Z_n(C) := \operatorname{source} \ker \partial_n \quad \text{and} \quad B_n(C) := \operatorname{source} \operatorname{im} \partial_{n+1}.$$

Remark 13.30. Let $\mathcal{A}$ be an abelian category, and $C \xrightarrow{f} D \xrightarrow{g} E$ be arrows in $\mathcal{A}$ such that $gf = 0$. By Remark 13.22, we can factor f as an epi followed by $\operatorname{im} f$.

$$\begin{tikzcd}
C \arrow[rr, "f"] \arrow[dr, "\text{epi}"'] & & D \arrow[r, "g"] & E \\
& \operatorname{im} f \arrow[ur] & &
\end{tikzcd}$$

Since $g \circ \operatorname{im} f \circ \text{epi} = gf = 0$, we must have $g \circ \operatorname{im} f = 0$, so $\operatorname{im} f$ factors uniquely through $\ker g$. Most importantly, there is a canonical arrow $\operatorname{im} f \longrightarrow \ker g$.

Definition 13.31. Let $\mathcal{A}$ be an abelian category. A sequence of arrows $C \xrightarrow{f} D \xrightarrow{g} E$ in $\mathcal{A}$ is **exact** if $gf = 0$ and $\ker g = \operatorname{im} f$.

Remark 13.32. In our definition of exact sequence, we really mean that the canonical arrow $\operatorname{im} f \longrightarrow \ker g$ we described in Remark 13.30 is an isomorphism. But notice that is equivalent to saying that the arrow $\operatorname{im} f$ is a kernel for g, and $\ker g$ is an image for f, hence the equality we wrote above, which is a more compact way of saying this.

This immediately generalizes to define an exact sequence, and once again a short exact sequence is one of the form

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0.$$

Exercise 45. Show that $0 \longrightarrow A \xrightarrow{f} B$ is exact if and only if f is a mono, and $B \xrightarrow{g} C \longrightarrow 0$ is exact if and only if g is an epi. Moreover, $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is a short exact sequence if and only if

- f is a mono.
- g is an epi.
- $f = \operatorname{im} f = \ker g$.
- $\operatorname{coker} f = g$.

Remark 13.33. Let $\mathcal{A}$ be an abelian category and (C, ∂) a complex in $\mathrm{Ch}(R)$. Since $\partial_n \partial_{n+1} = 0$ for all n, we get a canonical arrow $B_n(C) \longrightarrow Z_n(C)$ for each n.

Exercise 46. Given an additive category $\mathcal{A}$, B_n and Z_n are additive functors $\mathrm{Ch}(\mathcal{A}) \longrightarrow \mathcal{A}$. In particular, an arrow $C \xrightarrow{f} D$ induces arrows $Z_n(C) \xrightarrow{Z_n(f)} Z_n(D)$ and $B_n(c) \xrightarrow{B_n(f)} B_n(D)$.

Definition 13.34. Let $\mathcal{A}$ be an abelian category and (C, ∂) a complex in $\mathrm{Ch}(R)$. The nth homology of C is the object

$$\mathrm{H}_n(C) := \text{ target of } \operatorname{coker}(B_n(C) \longrightarrow Z_n(C)),$$

where $B_n(C) \longrightarrow Z_n(C)$ is the canonical arrow we described in Remark 13.30.

In fact, the nth homology is an additive functor $\mathrm{Ch}(ab) \longrightarrow \mathcal{A}$. But first, we need to make sense of what homology does to maps of complexes.

Let $\mathcal{A}$ be an abelian category and $C \xrightarrow{f} D$ a map of complexes in $\mathrm{Ch}(\mathcal{A})$. Fix an integer n. We get induced arrows $B_n(f)$ and $Z_n(f)$, since B_n and Z_n are additive functors. This gives us a commutative diagram

$$\begin{tikzcd}
B_n(C) \arrow[r, "\alpha"] \arrow[d, "B_n(f)"'] & Z_n(C) \arrow[r] \arrow[d, "Z_n(f)"] & \operatorname{coker} \alpha \\
B_n(C) \arrow[r, "\beta"'] & Z_n(C) \arrow[r] & \operatorname{coker} \beta
\end{tikzcd}$$

where α and β are the canonical arrows. To construct $\mathrm{H}_n(f)$, we claim that there is a unique arrow $\operatorname{coker} \alpha \longrightarrow \operatorname{coker} \beta$ making the diagram commute. This is all explained in the commutative diagram

$$\begin{tikzcd}
 & Z_n(C) \arrow[d, "\operatorname{coker} \alpha"] \arrow[ddr, bend left, "\operatorname{coker} \beta \circ (Z_n(f))"] & \\
B_n(C) \arrow[ur, "\alpha"] \arrow[r, "0"] \arrow[drr, bend right, "0"'] & \mathrm{H}_n(C) \arrow[dr, dashed, "\mathrm{H}_n(f)"] & \\
 & & \mathrm{H}_n(D)
\end{tikzcd}$$

where $\operatorname{coker} \beta \circ (Z_n(f)) \circ \alpha = \operatorname{coker} \beta \circ \beta \circ B_n(f) = 0$, which gives us a unique factorization $\mathrm{H}_n(f)$ through $\operatorname{coker} \alpha$.

Exercise 47. Given any abelian category $\mathcal{A}$, H_n is an additive functor $\mathrm{Ch}(\mathcal{A}) \longrightarrow \mathcal{A}$.

Similarly, we can define homotopies.

Definition 13.35. Let $\mathcal{A}$ be an abelian category and $f, g \colon F \longrightarrow G$ be maps complexes in $\mathrm{Ch}(\mathcal{A})$. A **homotopy**, sometimes referred to as a **chain homotopy**, between f and g is a sequence of arrows $h_n \colon F_n \longrightarrow G_{n+1}$

$$\begin{tikzcd}
\cdots \arrow[r, "\delta_{n+2}"] & F_{n+1} \arrow[r, "\delta_{n+1}"] \arrow[d, shift right, "g_{n+1}"'] \arrow[d, shift left, "f_{n+1}"] & F_n \arrow[r, "\delta_n"] \arrow[dl, "h_n"] \arrow[d, shift right, "g_n"'] \arrow[d, shift left, "f_n"] & F_{n-1} \arrow[r, "\delta_{n-1}"] \arrow[dl, "h_{n-1}"] \arrow[d, shift right, "g_{n-1}"'] \arrow[d, shift left, "f_{n-1}"] & \cdots \\
\cdots \arrow[r, "\delta_{n+2}"'] & G_{n+1} \arrow[r, "\delta_{n+1}"'] & G_n \arrow[r, "\delta_n"'] & G_{n-1} \arrow[r, "\delta_{n-1}"'] & \cdots
\end{tikzcd}$$

such that

$$\delta_{n+1} h_n + h_{n-1} \delta_n = f_n - g_n$$

for all n. If there exists a homotopy between f and g, we say that f and g are **homotopic**. If f is homotopic to the zero map, we say it is **null-homotopic**. If $f : F \longrightarrow G$ and $g \colon G \longrightarrow F$ are maps of complexes such that fg is homotopic to the identity arrow 1_G and gf is homotopic to the identity arrow 1_F, we say that f and g are **homotopy equivalences** and F and G are **homotopy equivalent**.

Exercise 48. Homotopy is an equivalence relation in $\mathrm{Ch}(\mathcal{A})$.

Exercise 49. Let $\mathcal{A}$ be an abelian category. Homotopic maps of complexes in $\mathrm{Ch}(\mathcal{A})$ induce the same map on homology.

Remark 13.36. Let F be an additive functor between abelian categories. Then F must send complexes to complexes, and it induces a functor $\mathrm{Ch}(\mathcal{A}) \longrightarrow \mathrm{Ch}(\mathcal{A})$, which we also call F. Now if h is a homotopy between two maps of complexes, F preserves the identities $\delta_{n+1} h_n + h_{n-1} \delta_n = f_n - g_n$ for all n, so $F(h)$ is a homotopy between $F(f)$ and $F(g)$.

Finally, we set up some notation we will use later.

Definition 13.37. We will denote the full subcategory of $\mathrm{Ch}(\mathcal{A})$ of complexes C such that $C_n = 0$ for all $n < k$ by $\mathrm{Ch}_{\geqslant k}(\mathcal{A})$.

13.3 Functors

Definition 13.38. Let $\mathcal{A}$ be an abelian category. A subcategory $\mathcal{B}$ of $\mathcal{A}$ is an **abelian subcategory** of $\mathcal{A}$ if $\mathcal{B}$ is abelian and the inclusion $\mathcal{B} \subseteq \mathcal{A}$ is an exact functor.

Exercise 50. Let $\mathcal{B}$ be a full subcategory of the abelian category $\mathcal{A}$.

a) $\mathcal{B}$ is an additive category if and only if $\mathcal{B}$ contains 0 and is closed under finite coproducts.

b) $\mathcal{B}$ is an abelian subcategory of A if and only if $\mathcal{B}$ is additive and closed under kernels and cokernels.

Definition 13.39. Let $T : \mathcal{A} \longrightarrow \mathcal{B}$ be an additive covariant functor between abelian categories. We say T is **left exact** if it takes every exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

to the exact sequence

$$0 \longrightarrow T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C),$$

and **right exact** if it takes every exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

to the exact sequence

$$T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C) \longrightarrow 0\,.$$

Finally, T is an **exact functor** if it preserves short exact sequences, meaning every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is taken to the short exact sequence

$$0 \longrightarrow T(A) \xrightarrow{T(f)} T(B) \xrightarrow{T(g)} T(C) \longrightarrow 0\,.$$

A contravariant additive functor $T\colon \mathcal{A} \longrightarrow \mathcal{B}$ between abelian categories is **left exact** if it takes every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

to the exact sequence

$$0 \longrightarrow T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A),$$

and **right exact** if it takes every exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

to the exact sequence

$$T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A) \longrightarrow 0\,.$$

Finally, T is an **exact functor** if it preserves short exact sequences, meaning every short exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is taken to the short exact sequence

$$0 \longrightarrow T(C) \xrightarrow{T(g)} T(B) \xrightarrow{T(f)} T(A) \longrightarrow 0\,.$$

Theorem 13.40. *Let $\mathcal{A}$ be an abelian category, and fix an object x in $\mathcal{A}$. The functors*

$$\begin{aligned} \mathcal{A} &\longrightarrow \boldsymbol{Ab} \\ y &\longmapsto \mathrm{Hom}_{\mathcal{A}}(x,y) \end{aligned} \qquad \textit{and} \qquad \begin{aligned} \mathcal{A} &\longrightarrow \boldsymbol{Ab} \\ y &\longmapsto \mathrm{Hom}_{\mathcal{A}}(y,x) \end{aligned}$$

are left exact.

Proof. We will show that $\mathrm{Hom}_{\mathcal{A}}(x,-)$ is left exact. Notice that the contravariant functor $\mathrm{hom}_{\mathcal{A}}(-,x)$ can be viewed as the covariant functor $\mathrm{Hom}_{\mathcal{A}^{\mathrm{op}}}(x,-)$. Since $\mathcal{A}^{\mathrm{op}}$ is also an abelian category, it will then follow that $\mathrm{Hom}_{\mathcal{A}^{\mathrm{op}}}(x,-)$ is also left exact, or equivalently, that $\mathrm{Hom}_{\mathcal{A}}(-,x)$ is left exact.

So let

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

be an exact sequence in $\mathrm{Ch}(\mathcal{A})$. We want to show that

$$0 \longrightarrow \mathrm{Hom}_{\mathcal{A}}(x,A) \xrightarrow{f_*} \mathrm{Hom}_{\mathcal{A}}(x,B) \xrightarrow{g_*} \mathrm{Hom}_{\mathcal{A}}(x,C)$$

is exact, and notice this last line lives in the category of abelian groups.

- Exactness at A is equivalent to f being a mono. By assumption, f is a mono, so $f_*(h) = fh$ is injective.
- Since $gf = 0$, so is $g_* f_* = (gf)_*$.
- We want to show that $\ker g_* = \mathrm{im}\, f_*$, and these are now maps of abelian groups. So we need to show that every $h \in \mathrm{Hom}_{\mathcal{A}}(x,C)$ such that $gh = 0$ factors uniquely through f, meaning $h = \mathrm{im}\, f_*$. Our assumption that the original sequence is exact implies that $f = \mathrm{im}\, f = \ker g$. The universal of property of the kernel gives us that whenever $gh = 0$, h must factor through $\ker g = f = \mathrm{im}\, f$.

□

Exercise 51. Let I be any small category. If $\mathcal{A}$ is an abelian category, then so is the category $\mathcal{A}^I$ of functors $I \longrightarrow \mathcal{A}$.

Theorem 13.41 (Yoneda Embedding for abelian categories)**.** *Let $\mathcal{A}$ be an abelian category. The covariant functor*

$$\begin{aligned} \mathcal{A} &\longrightarrow \boldsymbol{Ab}^{\mathcal{A}^{op}} \\ x &\longrightarrow \mathrm{Hom}_{\mathcal{A}}(-,x) \end{aligned}$$

from $\mathcal{A}$ to the category of contravariant functors $\mathcal{A} \longrightarrow \boldsymbol{Ab}$ is an embedding into a full subcategory and it preserves exactness, meaning that whenever

$$\mathrm{Hom}_{\mathcal{A}}(-,x) \longrightarrow \mathrm{Hom}_{\mathcal{A}}(-,y) \longrightarrow \mathrm{Hom}_{\mathcal{A}}(-,z)$$

is exact, so is $x \longrightarrow y \longrightarrow z$.

Proof. First, it is clear that our functor is injective on objects, as our axioms for a category include the assumption that the Hom-sets are all disjoint. We claim that

$$\begin{aligned} \mathrm{Nat}(\mathrm{Hom}_{\mathcal{A}}(-,x), \mathrm{Hom}_{\mathcal{A}}(-,y)) &\longrightarrow \mathrm{Hom}_{\mathcal{A}}(x,y) \\ \eta &\longmapsto \eta_x(1_x) \end{aligned}$$

is a natural bijection. This is essentially Theorem 9.36, so we leave it as an exercise. But in particular, the fact that this is a bijection says that our functor is indeed full and faithful.

To show that it reflects exactness, suppose that

$$\mathrm{Hom}_{\mathcal{A}}(-,x) \xrightarrow{f_*} \mathrm{Hom}_{\mathcal{A}}(-,y) \xrightarrow{g_*} \mathrm{Hom}_{\mathcal{A}}(-,z)$$

is exact. Then $g_* f_* = 0$, so $gf = g_* f_*(1_x) = 0$.

It remains to show that $\ker g = \operatorname{im} f$. Let ψ be the canonical arrow $\operatorname{im} f \longrightarrow \ker g$. The exactness of

$$\mathrm{Hom}_{\mathcal{A}}(-,x) \xrightarrow{f_*} \mathrm{Hom}_{\mathcal{A}}(-,y) \xrightarrow{g_*} \mathrm{Hom}_{\mathcal{A}}(-,z)$$

together with the fact that $g_*(\ker g) = 0$ imply that $\ker g$ factors through f. By Remark 13.26, $\ker g$ must also factor through $\operatorname{im} f$, say by φ. The universal property of the kernels $\ker g$ and $\operatorname{im} f$ will give us that ψ and φ are inverse isos. □

Corollary 13.42. *Let (L,R) be an adjoint pair of additive functors $\mathcal{A} \underset{R}{\overset{L}{\rightleftarrows}} \mathcal{B}$ between abelian categories. Then L is right exact, and R is left exact.*

Proof. Consider a short exact sequence

$$0 \longrightarrow x \longrightarrow y \longrightarrow z \longrightarrow 0$$

in $\mathcal{B}$, and let w be an object in $\mathcal{A}$. The adjointness of the pair (L,R) gives us a commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & \mathrm{Hom}_{\mathcal{A}}(w,Rx) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathcal{A}}(w,Ry) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathcal{A}}(w,Rz) \arrow[d] \\ 0 \arrow[r] & \mathrm{Hom}_{\mathcal{B}}(Lw,x) \arrow[r] & \mathrm{Hom}_{\mathcal{B}}(Lw,y) \arrow[r] & \mathrm{Hom}_{\mathcal{B}}(Lw,z) \end{tikzcd}$$

where the vertical maps are bijections of sets. For every w in $\mathcal{A}$, $\mathrm{Hom}_{\mathcal{B}}(Lw,-)$ is left exact, by Theorem 13.40, so the bottom row of the diagram above is exact. We claim this implies that the top row must also be exact. Our vertical maps are a priori only bijection on sets, but it is easy to see that these natural bijections restrict to a bijection between the images of each pair of corresponding maps. Moreover, for any objects A and B, the natural bijection $\mathrm{Hom}_{\mathcal{A}}(A,RB) \cong \mathrm{Hom}_{\mathcal{A}}(LA,B)$ must always send 0 to 0, since

$$\begin{tikzcd} 0 = \mathrm{Hom}_{\mathcal{A}}(A,R(0)) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathcal{A}}(A,R(B)) \arrow[d] \\ 0 = \mathrm{Hom}_{\mathcal{A}}(L(A),0) \arrow[r] & \mathrm{Hom}_{\mathcal{A}}(LA,B) \end{tikzcd}$$

commutes. It is then routine to check that our bijections also restrict to bijections between the kernels of each pair of corresponding maps. The exactness of the bottom row then induces exactness of the top row. By Theorem 13.41, Hom reflects exactness, and we conclude that

$$0 \longrightarrow Rx \longrightarrow Ry \longrightarrow Rz$$

must also be exact.

Finally, notice that L^{op} is the right adjoint to R^{op}, so L^{op} is left exact. Therefore, L must be right exact. □

Theorem 13.43 (Freyd-Mitchell embedding theorem)**.** *Let $\mathcal{A}$ be a small abelian category. There exists a ring R, possibly not commutative, and an exact, fully faithful embedding $\mathcal{A} \longrightarrow R$-**mod**.*

The full details of the proof are rather complicated, and can be found in [Fre03]. Here is a very rough map of the proof. By Theorem 13.41, we already have a fully faithful embedding of $\mathcal{A}$ in $\mathbf{Ab}^{\mathcal{A}^{\text{op}}}$, so it is sufficient to show that there is a fully faithful embedding of $\mathbf{Ab}^{\mathcal{A}^{\text{op}}}$ into some R-**mod**. The idea is to quotient $\mathbf{Ab}^{\mathcal{A}^{\text{op}}}$ by an abelian subcategory L that contains all the kernels and cokernels of the arrows $\operatorname{Hom}_{\mathcal{A}}(-, y) \longrightarrow \operatorname{Hom}_{\mathcal{A}}(-, z)$ for all epis $y \longrightarrow z$ in such a way that the composite of the embedding in Theorem 13.41 with this quotient remains an embedding. Then one shows that this quotient category has all coproducts and also what is called a projective generator. Roughly speaking, this is a projective object such that every object P such that for every object M there exists an arrow $P \longrightarrow M$. Then one shows that this implies that this category is equivalent to a full abelian subcategory of R-**mod** for some R.

Most of the theorems we have proved about R-**mod** extend to any abelian category. Some of those theorems can in fact be deduced from the fact that they are true over R-**mod**. In particular, short exact sequences of complexes in any abelian category induce a long exact sequence in homology.

Theorem 13.44 (Snake Lemma)**.** *Consider an abelian category $\mathcal{A}$ and a commutative diagram*

$$\begin{tikzcd}
 & A' \arrow[r, "i'"] \arrow[d, "f"'] & B' \arrow[r, "p'"] \arrow[d, "g"] & C' \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & A \arrow[r, "i"'] & B \arrow[r, "p"'] & C &
\end{tikzcd}.$$

If the rows of the diagram are exact, then there exists an exact sequence

$$\ker f \longrightarrow \ker g \longrightarrow \ker h \xrightarrow{\partial} \operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h.$$

Theorem 13.45 (Long exact sequence in homology)**.** *Given a short exact sequence in* $\operatorname{Ch}(R)$

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0,$$

there are connecting arrows $\partial : \mathrm{H}_n(C) \longrightarrow \mathrm{H}_{n-1}(A)$ such that

$$\cdots \longrightarrow \mathrm{H}_{n+1}(C) \xrightarrow{\partial} \mathrm{H}_n(A) \xrightarrow{\mathrm{H}_n(f)} \mathrm{H}_n(B) \xrightarrow{\mathrm{H}_n(g)} \mathrm{H}_n(C) \xrightarrow{\partial} \mathrm{H}_{n-1}(A) \longrightarrow \cdots$$

is an exact sequence.

Theorem 13.46 (The Five Lemma)**.** *Given an abelian category $\mathcal{A}$, consider the following commutative diagram in $\mathcal{A}$ with exact rows:*

$$\begin{tikzcd}
A' \arrow[r] \arrow[d, "a"'] & B' \arrow[r] \arrow[d, "b"'] & C' \arrow[r] \arrow[d, "c"'] & D' \arrow[r] \arrow[d, "d"'] & E' \arrow[d, "e"'] \\
A \arrow[r] & B \arrow[r] & C \arrow[r] & D \arrow[r] & E
\end{tikzcd}$$

If b and d are epi and e is a mono, then c is an epi. If b and d are mono and a is epi, then c is mono.

One can prove these by invoking the Freyd-Mitchell theorem and checking that one can go back and forth with our statements between some small subcategory of $\mathcal{A}$ containing our diagram and all the necessary kernels, cokernels, etc, and some R-**mod** where that category embeds. Alternatively, one can use what are called *members*, as in [ML98, VIII.4.5].

Chapter 14

Ext and Tor

While Hom and tensor are not exact functors, we can measure their lack of exactness using their derived functors Ext and Tor. These are the poster child examples of what are called derived functors, which can be constructed over any abelian category provided we have enough projective or injective objects.

14.1 Preliminaries

Definition 14.1. Let $\mathcal{A}$ be an abelian category. An object P in $\mathcal{A}$ is **projective** if $\mathrm{Hom}_{\mathcal{A}}(P,-)$ is an exact functor. An object E in $\mathcal{A}$ is **injective** if $\mathrm{Hom}_{\mathcal{A}}(-,E)$ is exact.

This generalizes the notion of projective and injective modules.

Remark 14.2. Let $\mathcal{A}$ be an abelian category. An object P is projective if and only if every arrow $P \longrightarrow Y$ factors through every epi $X \longrightarrow Y$:

$$\begin{tikzcd} & P \arrow[dl, dashed] \arrow[d] & \\ X \arrow[r] & Y \arrow[r] & 0 \end{tikzcd}$$

and an object E is injective if and only if every arrow $X \longrightarrow E$ factors through every mono $X \longrightarrow Y$:

$$\begin{tikzcd} & E & \\ 0 \arrow[r] & X \arrow[u] \arrow[r] & Y \arrow[ul, dashed] \end{tikzcd}$$

Exercise 52. Let $\mathcal{A}$ be an abelian category, and denote its coproduct by $\oplus$.

a) Show that $\mathrm{Hom}_{\mathcal{A}}(x \oplus y, z) = \mathrm{Hom}_{\mathcal{A}}(x, z) \oplus \mathrm{Hom}_{\mathcal{A}}(y, z)$.

b) Show that if P and Q are projective, then so is $P \oplus Q$.

Definition 14.3. An abelian category $\mathcal{A}$ has **enough projectives** if for every object M there exists a projective object P and an epi $P \longrightarrow M$. We say that $\mathcal{A}$ has **enough injectives** if for every object M there exists an injective object E and a mono $M \longrightarrow E$.

Lemma 11.19 and Theorem 11.35 say that R-**mod** has enough injectives and enough projectives.

Example 14.4. The category of finite abelian groups has no projectives beside 0.

Definition 14.5. Let M be an object in the abelian category $\mathcal{A}$. A **projective resolution** of M is a complex

$$\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \longrightarrow 0$$

where all the P_n are projective, $\mathrm{H}_0(P) = M$, and $\mathrm{H}_n(P) = 0$ for all $n \neq 0$. An **injective resolution** of M is a cocomplex

$$0 \longrightarrow E^0 \longrightarrow E^1 \longrightarrow \cdots$$

such that every E^n is injective, $\mathrm{H}^n(E) = 0$ for all n, and $\mathrm{H}^0(E) = M$.

Theorem 14.6. *If $\mathcal{A}$ has enough projectives, every object in $\mathcal{A}$ has a projective resolution. Similarly, if $\mathcal{A}$ has enough injectives, every object in $\mathcal{A}$ has an injective resolution.*

This generalizes Theorem 12.2 in a natural way, and the proof is essentially the same.

Proof. Given be an object M in $\mathcal{A}$, let's construct a projective resolution explicitly. We start by picking an epi $P_0 \xrightarrow{\varepsilon} M$ from a projective P_0. Since ϵ is an epi, it is the cokernel of its kernel, so

$$0 \longrightarrow \ker\varepsilon \longrightarrow P_0 \xrightarrow{\ \varepsilon\ } M \longrightarrow 0$$

is a short exact sequence. Now we find an epi $P_1 \xrightarrow{\varepsilon_1} K_0 := \ker\varepsilon$, and set $P_1 \xrightarrow{\partial_1} P_0$ to be the composition

$$\begin{tikzcd}
P_1 \arrow[rr, dashed, "\partial_1"] \arrow[dr, "\varepsilon_1"'] & & P_0 \arrow[r, "\varepsilon"] & M \arrow[r] & 0 \\
& \ker\varepsilon \arrow[ur, "i_0"'] \arrow[dr] & & & \\
0 \arrow[ur] & & 0 & &
\end{tikzcd}.$$

We proceed the same way, at each step taking a projective P_n and an epi $\varepsilon_n \colon P_n \longrightarrow \ker\partial_{n-1}$, and setting ∂_{n+1} to be the composition $(\ker\partial_{n-1}) \circ \varepsilon_n$. By construction, $\partial_n = i_{n-1}\varepsilon_n$, where ε_n is an epi and $\ker\partial_{n-1}$ is mono. By Exercise 44, $\operatorname{im}\partial_n = i_{n-1} = \ker\partial_{n-1}$. □

We can also characterize injectives in term of split short exact sequences, as we did for modules. In particular, the Splitting Lemma 11.6 extends to any abelian category.

Definition 14.7. Let $\mathcal{A}$ be an abelian category. A short exact sequence

$$0 \longrightarrow A \xrightarrow{\ f\ } B \xrightarrow{\ g\ } C \longrightarrow 0$$

splits if one of the following equivalent conditions hold:

1) There exists an arrow $C \xrightarrow{r} B$ such that $gr = \mathrm{id}_C$.

2) There exists an arrow $C \xrightarrow{r} B$ such that $gr = \mathrm{id}_C$.
3) There exists an isomorphism of complexes between our sequence and

$$0 \longrightarrow A \longrightarrow A \oplus C \longrightarrow C \longrightarrow 0$$

where the arrows are the canonical arrows that come with the (co)product $A \oplus C$.

Theorem 14.8. *Let $\mathcal{A}$ be an abelian category. Every short exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{f} C \longrightarrow 0$$

where A is injective or C is projective splits.

The proofs are exactly the same as in the case of R-**mod**, Theorem 11.11 and Theorem 11.40.

Proof. If C is projective, there exists h such that

$$\begin{tikzcd} & C \arrow[dl, dashed, "h"'] \arrow[d, equal] & \\ B \arrow[r, "g"'] & C \arrow[r] & 0 \end{tikzcd}$$

commutes, so $gh = \mathrm{id}_C$ and g is a splitting. If A is injective, there exists h such that

$$\begin{tikzcd} & A & \\ 0 \arrow[r] & A \arrow[u, equal] \arrow[r, "f"'] & B. \arrow[ul, dashed, "h"'] \end{tikzcd}$$

commutes, so $hf = \mathrm{id}_A$, and h is a splitting. □

More generally, we can talk about split exact complexes.

Definition 14.9. A complex C in $\mathrm{Ch}(\mathcal{A})$ is **split** if there are arrows $s_n \colon C_n \longrightarrow C_{n+1}$ such that the differential ∂ satisfies $\partial = \partial s \partial$. A complex is **split exact** if it is both exact and split.

Remark 14.10. A split short exact sequence is precisely a short exact sequence that is a split complex.

Exercise 53. Additive functors preserve split complexes, meaning that if C is a split complex, then so is $F(C)$ for any additive functor F. In particular, additive functors preserve split short exact sequences.

Lemma 14.11. *Let $\mathcal{A}$ be an abelian category, (P, ∂) in $\mathrm{Ch}_{\geqslant 0}(\mathcal{A})$ with each P_i projective, $P_0 \xrightarrow{\partial_0} M$ an arrow in $\mathcal{A}$ such that $\partial_0\partial_1 = 0$ and (Q, δ) a projective resolution of N. Given any $M \xrightarrow{f}$ in $\mathcal{A}$, there exists a map of complexes $P \xrightarrow{\varphi} Q$ such that*

$$\begin{tikzcd}
P_0 \arrow[r, "\partial_0"] \arrow[d, "\varphi_0"'] & M \arrow[d, "f"] \\
Q_0 \arrow[r, "\delta_0"'] & N
\end{tikzcd}$$

commutes, which is unique up to homotopy.

Proof. Since P_0 is projective and δ_0 is an epi, there exists φ_0 such that

$$\begin{tikzcd}
P_0 \arrow[r, "\partial_0"] \arrow[d, dashed, "\varphi_0"'] & M \arrow[r] \arrow[d, "f"] & 0 \\
Q_0 \arrow[r, "\delta_0"'] & N \arrow[r] & 0
\end{tikzcd}$$

commutes.

We proceed inductively, assuming we have $\varphi_0, \ldots, \varphi_{n-1}$ with $\varphi_{n-2}\partial_{n-1} = \delta_{n-2}\, \varphi_{n-1}$. Since P_n is projective, there exists φ_n such that

$$\begin{tikzcd}
P_n \arrow[r, "\partial_n"] \arrow[d, dashed, "\varphi_n"'] & P_{n-1} \arrow[r, "\partial_{n-1}"] \arrow[d, "\varphi_{n-1}"] & P_{n-2} \arrow[d, "\varphi_{n-2}"] \\
Q_n \arrow[r, "\delta_n"'] & Q_n \arrow[r, "\delta_{n-1}"'] & Q_{n-1}
\end{tikzcd}$$

commutes. Commutativity gives $\delta_{n-1}\varphi_{n-1}\partial_n = \varphi_{n-2}\partial_{n-1}\partial_n = 0$, so $\varphi_{n-1}\partial_n$ factors through the kernel of δ_{n-1}.

$$\begin{tikzcd}
P_n \arrow[drr, dashed] \arrow[dd, "\partial_n"] & Q_n \arrow[dr, two heads] \arrow[dd, "\delta_n"] & \\
& & Z_{n-1}(Q) \arrow[dl] \\
P_{n-1} \arrow[r, "\varphi_{n-1}"'] & Q_n &
\end{tikzcd}$$

Since Q is a projective resolution of N, the arrow $Q_n \longrightarrow Z_{n-1}(Q)$ above is an epi, so the arrow $P_n \longrightarrow Z_{n-1}(Q)$ we just constructed factors through Q_n, giving us φ_n such that

$$\begin{tikzcd}
P_n \arrow[r, dashed, "\varphi_n"] \arrow[drr, dashed] \arrow[dd, "\partial_n"] & Q_n \arrow[dr, two heads] \arrow[dd, "\delta_n"] & \\
& & Z_{n-1}(Q) \arrow[dl] \\
P_{n-1} \arrow[r, "\varphi_{n-1}"'] & Q_n &
\end{tikzcd}$$

commutes.

Now suppose we are given two such maps of complexes φ and ψ lifting f, say φ and ψ. Note that $\varphi - \psi$ and 0 are two liftings of the 0 map. We are going to show that any map lifting the 0 map $M \longrightarrow N$ must be nullhomotopic, which will then imply that φ and ψ are homotopic as well (essentially via the same homotopy!).

So let $\varphi\colon P \longrightarrow C$ be a map of complexes lifting the 0 map $M \longrightarrow N$.

$$\begin{tikzcd}
\cdots P_1 \arrow[r, "\partial_1"] \arrow[d, "\varphi_1"'] & P_0 \arrow[r, "\partial_0"] \arrow[d, "\varphi_0"'] & M \arrow[r] \arrow[d, "0"] & 0 \\
\cdots C_1 \arrow[r, "\delta_1"'] & C_0 \arrow[r, "\delta_0"'] & N \arrow[r] & 0
\end{tikzcd}$$

We will construct a nullhomotopy for φ inductively. Set $h_n = 0$ for all $n < 0$. The commutativity of the rightmost square says that $\delta_0\varphi_0 = 0$, so $\operatorname{im}\varphi_0 \subseteq \ker\delta_0 = \operatorname{im}\delta_1$. Since $\partial_0\varpi_0 = 0$, φ_0 factors through $Z_0(Q)$. But $Q_1 \twoheadrightarrow Z_0(Q)$ is an epi and P_0 is projective, there exists H_0 such that

$$\begin{tikzcd}
 & Q_1 \arrow[dr, two heads] \arrow[dd, "\delta_1"] & \\
 & & Z_0(Q) \arrow[dl] \\
P_0 \arrow[uur, dashed, "H_0"] \arrow[urr, dashed] \arrow[r, "\varphi_0"'] \arrow[d, "\partial_0"'] & Q_0 \arrow[d, "\delta_0"] & \\
M \arrow[r, "f"'] & N &
\end{tikzcd}$$

commutes. So H_0 satisfies $\delta_0 H_0 = \varphi_0$. Set $H_{-1} = 0$.

Now suppose we have constructed $H_0, \ldots, H_{n-1}$ such that $\delta_n H_{n-1} + H_{n-2}\partial_{n-1} = \varphi_{n-1}$. Then

$$\begin{aligned}
\delta_n\varphi_n &= \varphi_{n-1}\partial_n && \text{since } \varphi \text{ is a map of complexes} \\
&= (\delta_n H_{n-1} + H_{n-2}\partial_{n-1})\partial_n && \text{by assumption} \\
&= \delta_n H_{n-1}\partial_n + H_{n-2}\partial_{n-1}\partial_n && \\
&= \delta_n H_{n-1}\partial_n && \text{since } \partial_{n-1}\partial_n = 0
\end{aligned}$$

so $\delta_n(\varphi_n - H_{n-1}\partial_n) = 0$. Therefore, φ_n factors through $Z_n(Q)$, and since Q is a projective resolution of N, $Q_{n+1} \longrightarrow Z_n(Q)$ is an epi. Therefore, the factorization of $\varphi_n - H_{n-1}\partial_n$ through $Z_n(Q)$ also factors through Q_n, and we end up with an arrow H_n such that

$$\begin{tikzcd}
 & Q_{n+1} \arrow[dr, two heads] \arrow[dd, "\delta_{n+1}"] & \\
 & & Z_n(Q) \arrow[dl] \\
P_n \arrow[uur, dashed, "H_n"] \arrow[urr, dashed] \arrow[r, "\varphi_n"'] \arrow[d, "\partial_n"'] & Q_n \arrow[d, "\delta_n"] & \\
P_{n-1} \arrow[ur, "H_{n-1}"'] \arrow[r, "\varphi_{n-1}"'] & Q_{n-1} &
\end{tikzcd}$$

commutes. This H_n must then satisfy $\delta_{n-1}H_n + H_{n-1}\partial_n = \varphi_n$, and ultimately H is a homotopy between φ and 0. $\square$

Theorem 14.12 (Horseshoe Lemma)**.** *Let $\mathcal{A}$ be an abelian category, P be a projective resolution of A, and R be a projective resolution of C. If*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is an exact sequence, there exists a projective resolution Q of B and maps of complexes F and G lifting f and g such that

$$0 \longrightarrow P \xrightarrow{F} Q \xrightarrow{G} R \longrightarrow 0$$

is an exact sequence in $\mathrm{Ch}(\mathcal{A})$*.*

Proof. First, a word on notation: $\oplus$ denotes the coproduct in $\mathcal{A}$, and given arrows $x \xrightarrow{f} z$ and $y \xrightarrow{g} z$, we will write $f \oplus g$ for the unique arrow $x \oplus y \longrightarrow z$ induced by f and g. Moreover, we will denote the differential of P by ∂^P, and the differential of R by ∂^R.

Set $Q_n = P_n \oplus R_n$. Recall that the product and coproduct in $\mathcal{A}$ coincide, by Lemma 13.11, so let $F_n \colon P_n \longrightarrow Q_n$ and $G_n \colon Q_n \longrightarrow R_n$ be the canonical arrows. One can show that in fact we get short exact sequences

$$0 \longrightarrow P_n \xrightarrow{F_n} Q_n \xrightarrow{G_n} R_n \longrightarrow 0$$

for all n. Moreover, Q_n is projective for all n, by Exercise 52. We will construct the missing differentials ∂^Q inductively.

Since R_0 is projective and g is an epi, there exists γ such that

$$\begin{tikzcd}
0 \arrow[r] & P_0 \arrow[r, "F_0"] \arrow[d, "\partial_0"'] & Q_0 \arrow[r, "G_0"] & R_0 \arrow[r] \arrow[d, "\partial_0"] \arrow[dl, dashed, "\gamma"] & 0 \\
0 \arrow[r] & A \arrow[r, "f"] & B \arrow[r, "g"] & C \arrow[r] & 0
\end{tikzcd}$$

commutes. Set $\partial_0^Q := (f\partial_0^P) \oplus \gamma$. The universal property of the coproduct guarantees that

$$\begin{tikzcd}
0 \arrow[r] & P_0 \arrow[r, "F_0"] \arrow[d, "\partial_0"'] & Q_0 \arrow[r, "G_0"] \arrow[d, "\partial_0^Q"'] & R_0 \arrow[r] \arrow[d, "\partial_0"] \arrow[dl, dashed, "\gamma"] & 0 \\
0 \arrow[r] & A \arrow[r, "f"'] & B \arrow[r, "g"'] & C \arrow[r] & 0
\end{tikzcd}$$

commutes. By the Five Lemma 13.46, ∂_0^Q is epi. By the Snake Lemma 13.44,

$$\ker \partial_0^P \longrightarrow \ker \partial_0^Q \longrightarrow \ker \partial_0^R$$

is exact. We then proceed by induction, and at each step we apply the base case to

$$\begin{tikzcd}
0 \arrow[r] & P_{n+1} \arrow[r, "F_{n+1}"] \arrow[d, "\partial_{n+1}^P"'] & Q_{n+1} \arrow[r, "G_{n+1}"] & R_n \arrow[r] \arrow[d, "\partial_{n+1}^R"] & 0 \\
0 \arrow[r] & \ker \partial_n^P \arrow[r] & \ker \partial_n^Q \arrow[r] & \ker \partial_n^R \arrow[r] & 0
\end{tikzcd}$$

where the vertical arrows are epi because P and R are projective resolutions and thus exact. $\square$

Remark 14.13. By duality, if $\mathcal{A}$ has enough injectives, $0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$ is exact, and E_A and E_C are injective resolutions for A and C, there exists an injective resolution E_B of B and a short exact sequence of complexes $0 \longrightarrow E_A \longrightarrow E_B \longrightarrow E_C \longrightarrow 0$ extending the given one.

We finally have all the tools we need to construct derived functors, and in particular, Ext and Tor.

14.2 Derived functors

We start with the general construction of derived functors, although we will soon focus on concrete examples, most importantly Ext and Tor, the derived functors of hom and tensor.

Definition 14.14 (Derived functors)**.** Let $\mathcal{A}$ and $\mathcal{B}$ be abelian categories, and let $F\colon \mathcal{A} \longrightarrow \mathcal{B}$ be a covariant right exact functor. If $\mathcal{A}$ has enough projectives, the **left derived functors** of F are a sequence of functors $L_iF\colon \mathcal{A} \longrightarrow \mathcal{B}$, $i \geqslant 0$, defined as follows:

- For each object A in $\mathcal{A}$, fix a projective resolution P of A, and set

$$L_iF(A) := \mathrm{H}_i(F(P)).$$

- Given an arrow $A \xrightarrow{f} B$, fix projective resolutions $P \longrightarrow A$ and $Q \longrightarrow B$, and a map of complexes $P \xrightarrow{\varphi} Q$ lifting f. Then

$$L_iF(f) := \mathrm{H}_i(F(\varphi)).$$

Now let $F\colon \mathcal{A} \longrightarrow \mathcal{B}$ be a covariant left exact functor. If $\mathcal{A}$ has enough injectives, the **right derived functors** of F are a sequence of functors $R^iF\colon \mathcal{A} \longrightarrow \mathcal{B}$, $i \geqslant 0$, defined as follows:

- For each object A in $\mathcal{A}$, fix an injective resolution E of A, and set

$$R^iF(A) := \mathrm{H}^i(F(E)).$$

- Given an arrow $A \xrightarrow{f} B$, fix injective resolutions $A \longrightarrow E$ and $B \longrightarrow I$, and a map of complexes $P \xrightarrow{\varphi} Q$ extending f. Then

$$R^iF(f) := \mathrm{H}^i(F(\varphi)).$$

Let $F\colon \mathcal{A} \longrightarrow \mathcal{B}$ be a contravariant left exact functor. If $\mathcal{A}$ has enough projectives, the **right derived functors** of F are a sequence of functors $R^iF\colon \mathcal{A} \longrightarrow \mathcal{B}$, $i \geqslant 0$, defined as follows:

- For each object A in $\mathcal{A}$, fix a projective resolution P of A, and set

$$R^i F(A) := \mathrm{H}^i(F(P)).$$

- Given an arrow $A \xrightarrow{f} B$, fix projective resolutions $P \longrightarrow A$ and $Q \longrightarrow B$, and a map of complexes $P \xrightarrow{\varphi} Q$ extending f. Then

$$R^i F(f) := \mathrm{H}^i(F(\varphi)).$$

Finally, let $F \colon \mathcal{A} \longrightarrow \mathcal{B}$ be a contravariant right exact functor. If $\mathcal{A}$ has enough injectives, the **left derived functors** of F are a sequence of functors $L_i F \colon \mathcal{A} \longrightarrow \mathcal{B}$, $i \geqslant 0$, defined as follows:

- For each object A in $\mathcal{A}$, fix an injective resolution E of A, and set

$$L_i F(A) := \mathrm{H}_i(F(E)).$$

- Given an arrow $A \xrightarrow{f} B$, fix injective resolutions $A \longrightarrow E$ and $B \longrightarrow I$, and a map of complexes $E \xrightarrow{\varphi} I$ extending f. Then

$$L_i F(f) := \mathrm{H}_i(F(\varphi)).$$

It is not clear a priori that this construction is well-defined, but we will soon show that is indeed the case.

Remark 14.15. If F is exact, then $\mathrm{H}_i(F(C)) = F(\mathrm{H}_i(C))$, so $L_i F = 0$ for all $i > 0$.

Remark 14.16. If P is projective, then $0 \longrightarrow P \longrightarrow 0$ is a projective resolution of P, and thus $L_i F(P) = 0$ for all $i > 0$. Similarly, if E is injective then $R^i F(E) = 0$.

Proposition 14.17. *Let $\mathcal{A}$ be an abelian category with enough projectives, and F a covariant right exact functor.*

a) $L_i F(A)$ is well-defined up to isomorphism for every object A.

b) $L_i F(f)$ is well-defined for every arrow f.

c) $L_i F$ is an additive functor for each i.

d) $L_0 F = F$.

Proof.

a) Let P and Q be projective resolutions of A. Theorem 12.15 gives us maps of complexes $P \xrightarrow{\varphi} Q$ and $Q \xrightarrow{\psi} P$ such that $\varphi\psi$ is homotopic to 1_Q and $\psi\varphi$ is homotopic to 1_P. Additive functors preserve homotopies, by Remark 13.36, so $F(\varphi)F(\psi)$ and $F(\psi)F(\varphi)$ are homotopic to the corresponding identity arrows. Homotopic maps induce the same map in homology, by Exercise 49. Therefore, $F(\varphi)$ and $F(\psi)$ induce isomorphisms in homology.

b) Fix projective resolutions P and Q of M and N. Any two lifts φ and ψ of $f\colon M \longrightarrow N$ to $P \longrightarrow Q$ are homotopic, by Lemma 14.11. Additive functors preserve homotopies, by Remark 13.36, so $F(\varphi)$ and $F(\psi)$ are homotopic. Homotopic maps induce the same map in homology, by Exercise 49, so $L_iF(\varphi) = L_iF(\psi)$ for each i.

c) Given an arrow f, fix a lift φ of f to projective resolutions of the source and target. Since F is an additive functor, $\mathrm{H}_i(F(\varphi))$ is a homomorphism for each i, and thus $L_iF(f)$ is a homomorphism between the corresponding Hom-groups, which as we've seen is independent of our choice of φ.

d) Let A be any object and P be a projective resolution of A. Since P is right exact, and

$$P_1 \longrightarrow P_0 \longrightarrow A \longrightarrow 0$$

is exact, then so is

$$F(P_1) \longrightarrow F(P_0) \longrightarrow F(A) \longrightarrow 0.$$

We claim that $\mathrm{H}_0(F(P)) = F(A)$. In R**-mod**, this is very simple to justify: the last sequence above says that $F(A) = \operatorname{coker}(F(P_1) \longrightarrow F(P_0))$, and $\mathrm{H}_0(F(P)) = P_0/\operatorname{im}(F(P_1) \longrightarrow F(P_0)) = \operatorname{coker}(F(P_1) \longrightarrow F(P_0))$.

The argument in a general abelian category is essentially the same, modulo understanding that our definitions were set just right to make this work as desired. By Exercise 43, $\ker(F(P_0) \longrightarrow 0) = 1_{F(P_0)}$, so the canonical arrow $\operatorname{im} F(\partial_1) \longrightarrow F(P_0)$ is precisely the image of $\operatorname{im} F(\partial_1)$. By exactness of the last sequence we wrote above, $\operatorname{im} F(\partial_1) = \ker(F(P_0) \longrightarrow F(A))$. On the other hand, exactness at $F(A)$ says that $F(P_0) \longrightarrow F(A)$ is an epi, by Exercise 45. Every epi is the cokernel of its kernel, so $F(P_0) \longrightarrow F(A)$ is the cokernel of $\operatorname{im} F(\partial_1)$, which we saw was exactly the canonical arrow $B_1(F(P)) \longrightarrow Z_0(F(P))$. Therefore, $\mathrm{H}_0(F(P)) = F(A)$, the target of the cokernel of $B_1(F(P)) \longrightarrow Z_0(F(P))$. $\square$

Exercise 54. Let $\mathcal{A}$ be an abelian category with enough injectives, and F a covariant left exact functor.

a) $R^iF(A)$ is well-defined up to isomorphism.

b) $R^iF(f)$ is well-defined for every arrow f.

c) $R^iF(f)$ is an additive functor for every i.

d) $R^0F = F$.

Remark 14.18. If $\mathcal{A}$ is an abelian category with enough injectives, then $\mathcal{A}^{\mathrm{op}}$ is an abelian category with enough projectives. This gives us a relationship between left derived and right derived functors: $R^iF = (L_iF^{\mathrm{op}})^{\mathrm{op}}$.

And now we are ready to prove the most important result about derived functors: they fix the lack of exactness of the functor we are deriving, by inducing a long exact sequence in homology from any given short exact sequence.

Theorem 14.19. *Let $\mathcal{A}$ be an abelian category with enough projectives and F a right exact covariant functor. Any short exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

induces a natural long exact sequence which

$$\cdots \longrightarrow L_2F(C) \longrightarrow L_1F(A) \longrightarrow L_1F(B) \longrightarrow L_1F(C) \longrightarrow F(A) \longrightarrow F(B) \longrightarrow F(C) \longrightarrow 0.$$

Similarly, if F is a left exact covariant functor, we obtain a long exact sequence

$$0 \longrightarrow F(A) \longrightarrow F(B) \longrightarrow F(C) \longrightarrow R^1F(A) \longrightarrow R^1F(B) \longrightarrow R^1F(C) \longrightarrow R^2F(A) \longrightarrow \cdots .$$

If F is a contravariant left exact functor, we obtain a natural long exact sequence

$$0 \longrightarrow F(C) \longrightarrow F(B) \longrightarrow F(A) \longrightarrow R^1F(C) \longrightarrow R^1F(B) \longrightarrow R^1F(A) \longrightarrow R^2F(C) \longrightarrow \cdots .$$

Proof. We give a proof for the case of right exact functors, and the remaining cases follow by duality. We start by fixing projective resolutions P of A and R of C. By Theorem 14.12, we can choose a projective resolution Q of B and lifts of f and g such that

$$0 \longrightarrow P \longrightarrow Q \longrightarrow R \longrightarrow 0$$

is a short exact sequence in $\mathrm{Ch}(\mathcal{A})$. By Proposition 14.17, L_iF does not depend on the choice of resolution, so we can compute $L_iF(A)$, $L_iF(B)$, and $L_iF(C)$ from P, Q, and R. Now notice that for each n, R_n is projective, so

$$0 \longrightarrow P_n \longrightarrow Q_n \longrightarrow R_n \longrightarrow 0$$

is a split short exact sequence. Now additive functors preserve split short exact sequences, by Exercise 53, so

$$0 \longrightarrow F(P_n) \longrightarrow F(Q_n) \longrightarrow F(R_n) \longrightarrow 0$$

is a short exact sequence for all n. Then

$$0 \longrightarrow F(P) \longrightarrow F(Q) \longrightarrow F(R) \longrightarrow 0$$

is also a short exact sequence, now in $\mathrm{Ch}(\mathcal{A})$. Note, however, that this sequence is not necessarily split anymore, since the splittings at each level do not necessarily assemble into a map of complexes. The Long Exact Sequence in homology Theorem 13.45 now gives us the long exact sequence we desire.

It remains to check naturality. What is left to check is that given a commutative diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d, "a"'] & B \arrow[r] \arrow[d, "b"'] & C \arrow[r] \arrow[d, "c"'] & 0 \\
0 \arrow[r] & A' \arrow[r] & B' \arrow[r] & C' \arrow[r] & 0
\end{tikzcd}$$

and chosen lifts of the original short exact sequences to projective resolutions, there are maps of complexes such that

$$\begin{tikzcd} 0 \arrow[r] & P \arrow[r] \arrow[d, "\alpha"'] & Q \arrow[r] \arrow[d, "\beta"'] & R \arrow[r] \arrow[d, "\gamma"'] & 0 \\ 0 \arrow[r] & P' \arrow[r] & Q' \arrow[r] & R' \arrow[r] & 0 \end{tikzcd}$$

commutes. Our derived functors L_iF will preserve these maps of complexes and the commutativity of the diagram above, so we get commutative diagrams

$$\begin{tikzcd} L_iF(C) \arrow[r] \arrow[d, "L_iF(\gamma)"'] & L_{i-1}F(A) \arrow[d, "L_{i-1}F(\alpha)"] \\ L_iF(C') \arrow[r] & L_{i-1}F(A) \end{tikzcd}$$

for each i. First, notice that we know that a, b, and c can be lifted to maps of complexes by Lemma 14.11, and that any two lifts of each a, b, or c are unique up to homotopy. So let's start by fixing lifts α of a and γ of c, and we will construct an appropriate lift β of b. Since the short exact sequences

$$0 \longrightarrow P_n \longrightarrow Q_n \longrightarrow R_n \longrightarrow 0$$

split for each n, we might as well assume that $Q_n = P_n \oplus R_n$ and that the arrows $P \longrightarrow Q$ and $Q \longrightarrow R$ are given by the canonical arrows to and from the product $\equiv$ coproduct in each homological degree. We cannot, however, assume $Q = P \oplus R$ as complexes, only that $Q_n = P_n \oplus R_n$ in each homological degree n. The commutativity of

$$\begin{tikzcd} 0 \arrow[r] & P_n \arrow[r] \arrow[d, "\partial_n^P"'] & P_n \oplus R_n \arrow[d, "\partial_n^Q"] \\ 0 \arrow[r] & P_{n-1} \arrow[r] & P_{n-1} \oplus R_{n-1} \end{tikzcd}$$

does imply that $\partial^Q(P) \subseteq P$, so we can say that ∂^Q is of the form

$$\partial_n^Q = \begin{pmatrix} \partial_n^P & \mu_n \\ 0 & \partial_n^R \end{pmatrix}$$

for each n. Since this is a differential, we have

$$(\partial_n^Q)^2 = 0 \implies \partial_{n-1}^P \mu_n + \mu_{n-1}\partial_n^R = 0.$$

Similarly, all this applies to $\partial_n^{Q'}$, which must be of the form

$$\partial_n^{Q'} = \begin{pmatrix} \partial_n^{P'} & \mu_n' \\ 0 & \partial_n^{R'} \end{pmatrix}.$$

We claim that we can define $\beta_n = \begin{pmatrix} \alpha_n & \nu_n \\ 0 & \gamma_n \end{pmatrix}$ for each n such that β is a map of complexes, meaning

$$\partial_n^{Q'}\beta_n = \beta_{n-1}\partial_n^Q.$$

Writing the corresponding products of matrices, we must have

$$\begin{pmatrix} \partial_n^{P'} & \mu_n' \\ 0 & \partial_n^{R'} \end{pmatrix} \begin{pmatrix} \alpha_n & \nu_n \\ 0 & \gamma_n \end{pmatrix} = \begin{pmatrix} \alpha_{n-1} & \nu_{n-1} \\ 0 & \gamma_{n-1} \end{pmatrix} \begin{pmatrix} \partial_n^{P} & \mu_n \\ 0 & \partial_n^{R} \end{pmatrix} \implies \begin{cases} \alpha \text{ is a map of complexes} \\ \partial_n^{P'}\nu_n + \mu_n'\gamma_n = \alpha_{n-1}\mu_n + \nu_{n-1}\partial_n^R \\ 0 = 0 \\ \gamma \text{ is a map of complexes} \end{cases}$$

The only nontrivial statement we want to guarantee is that $\partial_n^{P'}\nu_n + \mu_n'\gamma_n = \alpha_{n-1}\mu_n + \nu_{n-1}\partial_n^R$. We can solve this inductively for each n, and construct an appropriate ν_n inductively. Given ν_{n-1}, set

$$\Gamma_n := \alpha_{n-1}\mu_n + \nu_{n-1}\partial_n^R - \mu_n'\gamma_n,$$

We want to construct ν_n such that
$$\begin{tikzcd} R_n \arrow[r, "\nu_n"] \arrow[dr, "\Gamma_n"'] & P_n' \arrow[d] \\ & P_{n-1}' \end{tikzcd}$$
commutes, assuming we have constructed ν_{n-1}. First, we claim that $\partial_{n-1}^{P'}\Gamma_n = 0$.

$$\begin{aligned} \partial_{n-1}^{P'}\Gamma_n =& \partial_{n-1}^{P'}\alpha_{n-1}\mu_n + \partial_{n-1}^{P'}\nu_{n-1}\partial_n^R - \partial_{n-1}^{P'}\mu_n'\gamma_n \\ =& \mu_{n-1}'\partial_n^{P'}\gamma_n + \partial_{n-1}^{P'}\alpha_{n-1}\mu_n + \partial_{n-1}^{P'}\nu_{n-1}\partial_n^R && \text{since } \mu_{n-1}'\partial_n^{P'} = \partial_{n-1}^{P}\mu_n \end{aligned}$$

By induction,

$$\partial_{n-1}^{P'}\nu_{n-1} + \mu_{n-1}'\gamma_{n-1} = \alpha_{n-2}\mu_{n-1} + \nu_{n-2}\partial_{n-1}^R.$$

Using this to replace $\partial_{n-1}^{P'}\nu_{n-1}$ in the equation above, we get

$$\begin{aligned} \partial_{n-1}^{P'}\Gamma_n =& \mu_{n-1}'\partial_n^{P'}\gamma_n + \partial_{n-1}^{P'}\alpha_{n-1}\mu_n + (\alpha_{n-2}\mu_{n-1} + \nu_{n-2}\partial_{n-1}^R - \mu_{n-1}'\gamma_{n-1})\partial_n^R \\ =& \alpha_{n-2}\mu_{n-1}\partial_n^R + \partial_{n-1}^{P'}\alpha_{n-1}\mu_n + \nu_{n-2}\partial_{n-1}^R\partial_n^R - \mu_{n-1}'(\partial_n^{P'}\gamma_n + \gamma_{n-1}\partial_n^R) \\ =& \alpha_{n-2}\partial_{n-1}^{P}\mu_n + \partial_{n-1}^{P'}\alpha_{n-1}\mu_n + \nu_{n-2}\partial_{n-1}^R\partial_n^R - \mu_{n-1}'(\partial_n^{P'}\gamma_n + \gamma_{n-1}\partial_n^R) \end{aligned}$$

We showed above that $\partial_n^{P'}\gamma_n + \gamma_{n-1}\partial_n^R = 0$. Moreover, $\partial_{n-1}^R\partial_n^R = 0$. We conclude that

$$\begin{aligned} \partial_{n-1}^{P'}\Gamma_n =& \alpha_{n-2}\partial_{n-1}^{P}\mu_n + \partial_{n-1}^{P'}\alpha_{n-1}\mu_n \\ =& \alpha_{n-2}\partial_{n-1}^{P}\mu_n + \alpha_{n-2}\partial_n^{P'}\mu_n && \text{since } \alpha \text{ is a map of complexes} \\ =& \alpha_{n-2}(\partial_{n-1}^{P}\mu_n + \partial_n^{P'}\mu_n) \\ =& 0 && \text{since } \partial_{n-1}^{P}\mu_n + \partial_n^{P'}\mu_n = 0. \end{aligned}$$

So this concludes the proof that $\partial_{n-1}^{P'}\Gamma_n = 0$. Therefore, Γ_n must factor through the $\ker \partial_{n-1}^{P'}$:

$$\begin{tikzcd} P_n' \arrow[rr, "\partial_n"] & & P_{n-1}' \arrow[r, "\partial_{n-1}"] & P_{n-2}' \\ & \ker\partial_{n-1} \arrow[ur] & R_n \arrow[u, "\Gamma_n"'] \arrow[l, dashed, "\psi_n"] & \end{tikzcd}$$

On the other hand, P' is a resolution and thus exact, so $\operatorname{im} \partial_n = \ker \partial_{n-1}$, and ∂_n factors through $\ker \partial_{n-1}$ as

$$\begin{tikzcd}
P'_n \arrow[rr, "\partial_n"] \arrow[dr, "\varphi_n"'] & & P'_{n-1} \arrow[r, "\partial_{n-1}"] & P'_{n-2} \\
& \ker \partial_{n-1} \arrow[ur] & R_n \arrow[l, dashed, "\psi_n"] \arrow[u, "\Gamma_n"'] &
\end{tikzcd}$$

via some epi φ_n. Finally, R_n is projective, so there exists ν_n such that

$$\begin{tikzcd}
& R_n \arrow[dl, dashed, "\nu_n"'] \arrow[d, "\psi_n"] \\
P'_n \arrow[r, "\varphi_n"'] & \ker \partial_{n-1}
\end{tikzcd}$$

commutes — this was the ν_n we were searching for. □

Theorem 14.19 can be phrased in a fancier way by saying that the derived functors of F are a (co)homological δ-functor. In general, a δ-functor $\mathcal{A} \longrightarrow \mathcal{B}$ is a sequence of additive functors that produce a long exact sequence given a short exact sequence, in a functorial way — meaning that each map of complexes between two short exact sequences gives rise to a commutative diagram between the corresponding long exact sequences. It turns out that the left/right derived functors of F form what is called a universal δ-functor, which boils down to it having a certain universal property. While we will not discuss δ-functors in detail, the topic can be found in any standard reference — for example, see [Wei94].

Theorem 14.20. *Let $T_i : \mathcal{A} \longrightarrow \mathcal{B}$ be a sequence of additive covariant functors between abelian categories, where $\mathcal{A}$ has enough projectives, and $F : \mathcal{A} \longrightarrow \mathcal{B}$ a right exact functor. Suppose the following hold:*

- *For every short exact sequence $0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$ in $\mathcal{A}$, we get a long exact sequence*
$$\cdots \longrightarrow T_2(C) \longrightarrow T_1(A) \longrightarrow T_1(B) \longrightarrow T_1(C) \longrightarrow T_0(A) \longrightarrow T_0(B) \longrightarrow T_0(C) \longrightarrow 0.$$
- *T_0 is naturally isomorphic to F.*
- *$T_n(P) = 0$ for every projective object P in $\mathcal{A}$, and all $n \geqslant 1$.*

Then T_n is naturally isomorphic to L_nF for all n.

14.3 A first look at Ext and Tor

It's time to return to R-**mod** and study some concrete examples of derived functors: Ext, the derived functor of Hom, and Tor, the derived functor of tensor. There are two Hom functors, each with its own derived functor: given R-modules M and N, we may take a projective resolution P of M, and compute $\mathrm{H}^i(\mathrm{Hom}_R(P, N))$, or we could take an injective resolution E of N, and compute $\mathrm{H}^i(\mathrm{Hom}_R(M, E))$. It turns out these two completely different sounding constructions give us isomorphic R-modules, which we call $\operatorname{Ext}^i_R(M,N)$. Similarly, we have two constructions for $\operatorname{Tor}_i(M,N)$: we may take a projective resolution P of M, and compute $\mathrm{H}_i(P \otimes N)$, or take a projective resolution Q of N, and compute $\mathrm{H}_i(M \otimes_R Q)$. Again, it turns out that these two definitions are equivalent, and the resulting module is $\operatorname{Tor}^R_i(M,N)$.

To show that for each of Ext and Tor these two seemingly unrelated definitions agree, we will need some more tools.

Definition 14.21. Let $\mathcal{A}$ be an abelian category. A **double complex** in $\mathcal{A}$ is a family of objects $\{C_{p,q}\}_{p,q\in\mathbb{Z}}$ together with arrows $d^h : C_{p,q} \longrightarrow C_{p-1,q}$ and $d^v : C_{p,q} \longrightarrow C_{p,q-1}$

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d] & \\
\cdots & C_{p-1,q+1} \arrow[l] \arrow[d, "d^v"] & C_{p,q+1} \arrow[l, "d^h"'] \arrow[d, "d^v"] & \cdots \arrow[l] \\
\cdots & C_{p-1,q} \arrow[l] \arrow[d] & C_{p,q} \arrow[l, "d^h"] \arrow[d] & \cdots \arrow[l] \\
 & \vdots & \vdots &
\end{tikzcd}$$

satisfying

$$d^h d^h = 0 \qquad d^v d^v = 0 \qquad d^h d^v + d^v d^h = 0.$$

Given a double complex C, its **total complex** is given by

$$\operatorname{Tot}^{\oplus}(C)_n := \bigoplus_{p+q=n} C_{p,q}$$

with differential $d = d^h + d^v$. Similarly, its **product total complex** is given by

$$\operatorname{Tot}^{\prod}(C)_n := \prod_{p+q=n} C_{p,q}$$

with differential $d = d^h + d^v$.

If we fix p, $C_{p,\bullet}$ is a complex with differential d^v. Similarly, if we fix q, $C_{\bullet,q}$ is a complex with differential d^h. Moreover, the total complex of a chain complex is indeed a complex.

Remark 14.22. Let C be a double complex with differentials d^v and d^h. Then

$$(d^h + d^v)^2 = \underbrace{d^h d^v}_{0} + \underbrace{d^h d^h + d^v d^h}_{0} + \underbrace{d^v d^v}_{0} = 0,$$

so $(\operatorname{Tot}^{\oplus}(C), d)$ and $(\operatorname{Tot}^{\prod}(C), d)$ are indeed complexes.

The most important examples — and the ones we will need as tools to prove our two definitions of Ext agree — are the tensor and the Hom double complex.

Definition 14.23. Let R be a ring and C and D be complexes in $\mathrm{Ch}(R)$. The **tensor product double complex** of C and D is the double complex $C \otimes D$ given by taking $(C \otimes D)_{p,q} = C_p \otimes D_q$, $d^h = \partial^C \otimes_R 1_D$, and $d^v = (-1)^p 1_C \otimes_R \partial^D$.

We call the total complex of the tensor product double complex of C and D the **tensor product** of C and D in $\mathrm{Ch}(R)$, and denote it by $C \otimes D$.

So the tensor product total complex has $\mathrm{Tot}^{\oplus}(C \otimes D)_n = \bigoplus_{p+q=n} C_n \otimes_R D_n$ and differential $d(x \otimes y) = \partial(x) \otimes y + (-1)^p x \otimes \partial(y)$ for $x \in C_p$ and $y \in D_q$.

Definition 14.24. Let R be a ring and C and D be complexes in $\mathrm{Ch}(R)$. The **Hom double complex** of C and D is the double complex $\mathrm{Hom}(C, D)$ given by $(\mathrm{Hom}(C, D))_{p,q} := \mathrm{Hom}_R(C_{-p}, D_q)$,

$$\begin{array}{ccc} \mathrm{Hom}_R(C_{-p}, D_q) & \xrightarrow{d^h} & \mathrm{Hom}_R(C_{-p-1}, D_q) \\ f & \longmapsto & f \circ \partial^C \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathrm{Hom}_R(C_{-p}, D_q) & \xrightarrow{d^v} & \mathrm{Hom}_R(C_{-p}, D_{q-1}) \\ f & \longmapsto & (-1)^{p+q+1}\partial^D \circ f \end{array}.$$

We call the product total complex of the Hom double complex of C and D the **(internal) Hom complex** of C and D, and denote it by $\mathrm{Hom}(C, D)$.

So the Hom complex of C and D is the complex

$$\mathrm{Hom}(C, D)_n = \prod_{p+q=n} \mathrm{Hom}_R(C_{-p}, D_q)$$

with differential $d(f) = f \circ \partial^C + (-1)^{p+q+1}\partial^D \circ f$ for each $f \in \mathrm{Hom}_R(C_{-p}, D_q)$.

Remark 14.25. Given C and D in $\mathrm{Ch}(R)$, what is a 0-cycle in the Hom complex $\mathrm{Hom}(C, D)$? A 0-cycle is a sequence of maps of R-modules $f_k : C_k \longrightarrow D_k$ satisfying $f\partial^C - \partial^D f = 0$, so the 0-cycles are precisely the maps of complexes $C \longrightarrow D$. Similarly, a sequence of maps f_k is a 0-boundary if there exists a sequence of maps $h_k : C_k \longrightarrow D_{k+1}$ such that $f_k = \partial^D h_k + h_{k-1}\partial^C$. In other words, a 0-boundary indicates a homotopy relation — if $f - g$ is a 0-boundary, f and g are homotopic maps.

Lemma 14.26 (Acyclic Assembly Lemma)**.** *Let C be a double complex in R-**mod**.*

a) If C is an upper half plane double complex with exact rows, meaning $C_{p,q} = 0$ whenever $q < 0$, then $\mathrm{Tot}^{\oplus}(C)$ is exact.

b) If C is a right half plane double complex with exact columns, meaning $C_{p,q} = 0$ whenever $p < 0$, then $\mathrm{Tot}^{\oplus}(C)$ is exact.

c) If C is an upper half plane double complex with exact columns, meaning $C_{p,q} = 0$ whenever $q < 0$, then $\mathrm{Tot}^{\prod}(C)$ is exact.

d) If C is a right half plane double complex with exact rows, meaning $C_{p,q} = 0$ whenever $p < 0$, then $\mathrm{Tot}^{\prod}(C)$ is exact.

Proof. Notice that a $\Leftrightarrow$ b and c $\Leftrightarrow$ d by switching the indexes. Moreover, we claim that it is sufficient to show c, since it implies b.

First, we need some notation. Given a double complex C, consider the nth truncation $\tau_n(C)$ of C defined by

$$\tau_n(C)_{p,q} := \begin{cases} C_{p,q} & \text{if } q > n \\ \ker(\, C_{p,n} \xrightarrow{d^v} C_{p,n-1}\,) & \text{if } q = n \\ 0 & \text{if } q < n. \end{cases}$$

There is a natural inclusion $\tau_n(C) \longrightarrow C$ which induces an isomorphism in homology for $i \geqslant n$.

So suppose that C is a right half plane double complex with exact columns, and assume that c holds. Then $\tau_n(C)$ still has exact columns, so by c, $\mathrm{Tot}^{\prod}(\tau_n(C))$ is exact. On the other hand, notice that up to a vertical shift, $\tau_n(C)$ is a first quadrant double complex, and for each fixed m, there are only finitely many values of p and q with $p+q=m$ and such that $\tau_n(C)_{p,q} \neq 0$. Therefore, $\mathrm{Tot}^{\prod}(\tau_n(C_{p,\bullet})) = \mathrm{Tot}^{\oplus}(\tau_n(C_{p,\bullet}))$, so $\mathrm{Tot}^{\oplus}(\tau_n(C_{p,\bullet}))$ is exact. We claim that this implies that $\mathrm{Tot}^{\oplus}(C)$ is exact. One can make this precise by saying $\mathrm{Tot}^{\oplus}(C) = \mathrm{colim}_n(\mathrm{Tot}^{\oplus}(C))$. We haven't discussed colimits, but this is actually easy to check explicitly. The point is that any element $a \in Z_k(\mathrm{Tot}^{\oplus}(C))$, when we write a explicitly as $a = (a_{p,q}) \in \oplus_{p+q=k} C_{p,q}$ in terms of its coordinates in each $C_{p,q}$, only finitely many $a_{p,q}$ are nonzero. Let q be the smallest such that $a_{p,q} \neq 0$, and fix any $n < q$. Then $a \in Z_k(\mathrm{Tot}^{\oplus}(\tau_n(C))) = B_k(\mathrm{Tot}^{\oplus}(\tau_n(C))) \subseteq B_k(\mathrm{Tot}^{\oplus}(C))$. So $\mathrm{Tot}^{\oplus}(C)$ is exact, so b holds.

All we have left to do is to show c, meaning that the product total complex of any upper half plane double complex C with exact columns is exact. We are going to show that $\mathrm{H}_0(\mathrm{Tot}^{\prod}(C)) = 0$, and the remaining homologies follow by shifting C left and right. Consider a 0-cycle in $\mathrm{Tot}^{\prod}(C)$, meaning a sequence of elements $c_p \in C_{-p,p}$ for each $p \geqslant 0$ such that $c = (c_p) \in Z_0(\mathrm{Tot}^{\prod}(C))$. So

$$d(c) = 0 \Leftrightarrow d^v(c_p) + d^h(c_{p-1}) = 0 \text{ for all } p.$$

We will construct $b_{-p,p+1} \in C_{-p,p+1}$ for each p such that $d^v(b_{-p,p+1}) + d^h(b_{-p+1,p}) = c_p$, proving that $c \in B_0(\mathrm{Tot}^{\prod}(C))$.

Set $b_{1,0} = 0 \in C_{1,0}$ when $p=-1$. Since $C_{0,-1} = 0$, we must have $d^v(c_0) = 0 \in C_{0,-1}$. We also assumed that the columns are exact, so in particular the 0th column is exact. We can then find $b_{0,1} \in C_{0,1}$ such that $d^v(b_{0,1}) = c_0$, and thus $d^v(b_{0,1}) + d^h(b_{1,0}) = c_0$.

Now we proceed by induction. Suppose we have constructed $b_{-s+1,s}$ for $-1 \leqslant s \leqslant p$ with the desired property that $d^v(b_{-s,s+1}) + d^h(b_{-s+1,s}) = c_s$ for all $s \leqslant p$. Then

$$\begin{aligned} d^v(c_{-p,p} - d^h(b_{-p+1,p})) &= d^v(c_p) + d^h d^v(b_{-p+1,p}) && \text{since } d^v d^h + d^h d^v = 0 \\ &= d^v(c_p) + d^h(c_{p-1} - d^h(b_{-p+2,p-1})) && \text{as } d^v(b_{-p+1,p}) + d^h(b_{-p+2,p-1}) = c_{p-1} \\ &= d^v(c_p) + d^h(c_{p-1}) - d^h d^h(b_{-p+2,p-1}) && \\ &= d^v(c_p) + d^h(c_{p-1}) && \text{since } d^h d^h = 0 \\ &= 0. && \end{aligned}$$

The last equality comes simply from the fact that $(d^v + d^h)(c) = 0$. So we have shown that $d^v(c_{-p,p} - d^h(b_{-p+1,p})) = 0$. Since the columns are exact, we can find $b_{-p,p+1} \in C_{-p,p+1}$ such that

$$d^v(b_{-p,p+1}) = c_{-p,p} - d^h(b_{-p+1,p}).$$

Equivalently,

$$d^v(b_{-p,p+1}) + d^h(b_{-p+1,p}) = c_{-p,p}. \qquad \square$$

We are also going to need a few other constructions with complexes.

Definition 14.27. Let $\mathcal{A}$ be an abelian category and C be a complex in $\mathrm{Ch}(\mathcal{A})$. The **suspension** of C is the complex $\Sigma C := C[-1]$ with $(\Sigma C)_n = C_{n-1}$ and $\partial^{\Sigma C} = -\partial^C$. Given an integer k, the k**th suspension** of C is the complex $\Sigma^k C := \underbrace{\Sigma \cdots \Sigma}_{c \text{ times}} C$, so $\partial^{\Sigma^k C} = (-1)^k \partial^C$.

Definition 14.28. Let $\mathcal{A}$ be an abelian category and $f\colon C \longrightarrow D$ be a map of complexes. The **cone** of f is the complex $\mathrm{cone}(f)$ with $\mathrm{cone}(f)_n = C_{n-1} \oplus D_n$ and differential given by

$$\partial_n := \begin{pmatrix} -\partial_C & 0 \\ -f & \partial_D \end{pmatrix} : \quad \begin{tikzcd} C_{n-1} \arrow[r, "-\partial^C"] \arrow[dr, "-f"] & C_{n-2} \\ \oplus & \oplus \\ D_n \arrow[r, "\partial^D"'] & D_{n-1} \end{tikzcd}$$

Exercise 55. Show that giving a map of complexes $\mathrm{cone}(\, C \xrightarrow{f} D \,) \longrightarrow E$ is the same as giving

- a map of complexes $D \xrightarrow{g} E$, and
- a homotopy between gf and 0.

Remark 14.29. Given any map of complexes $C \xrightarrow{f} D$, there is a short exact sequence

$$0 \longrightarrow D \longrightarrow \mathrm{cone}(f) \longrightarrow \Sigma C \longrightarrow 0$$

determined by the canonical arrows to and from the product $\equiv$ coproduct. The connecting arrows from the Snake Lemma

$$\mathrm{H}_{n-1}(C) = \mathrm{H}_n(\Sigma^{-1} C) \xrightarrow{\delta} \mathrm{H}_{n-1}(D)$$

are exactly $\mathrm{H}_{n-1}(f)\colon \mathrm{H}_{n-1}(C) \longrightarrow \mathrm{H}_{n-1}(D)$ induced by f, so there is a long exact sequence

$$\cdots \longrightarrow \mathrm{H}_{n+1}(\mathrm{cone}(f)) \longrightarrow \mathrm{H}_n(C) \xrightarrow{\mathrm{H}_n(f)} \mathrm{H}_n(D) \longrightarrow \mathrm{H}_n(\mathrm{cone}(f)) \longrightarrow \mathrm{H}_{n-1}(C) \longrightarrow \cdots .$$

As a consequence, f is a quasi-isomorphism if and only if $\mathrm{cone}(f)$ is exact.

Remark 14.30. Given a map of complexes $C \xrightarrow{f} D$, we can construct a double complex from f, as follows:

$$X = \quad \begin{tikzcd}
 & \vdots \arrow[d, "\partial"'] & \vdots \arrow[d, "-\partial"] & \\
0 & D_2 \arrow[l] \arrow[d, "\partial"'] & C_2 \arrow[l, "f"'] \arrow[d, "-\partial"] & 0 \arrow[l] \\
0 & D_1 \arrow[l] \arrow[d, "\partial"'] & C_1 \arrow[l, "f"'] \arrow[d, "-\partial"] & 0 \arrow[l] \\
0 & D_0 \arrow[l] \arrow[d, "\partial"'] & C_0 \arrow[l, "f"'] \arrow[d, "-\partial"] & 0 \arrow[l] \\
 & \vdots & \vdots &
\end{tikzcd}$$

Note that $\mathrm{Tot}^{\oplus}(X) = \mathrm{cone}(f)$.

Exercise 56. Given a double complex C with $C_{p,q} = 0$ for all $p < n$, the horizontal differentials $C_{n+1,q} \longrightarrow C_{n,q}$ induce a map of complexes

$$\mathrm{Tot}^{\oplus/\prod}(C_{>n,\bullet}) \longrightarrow C_{n,\bullet}\,,$$

where $C_{>n,\bullet}$ denotes the double complex we obtain from C by excluding the leftmost nonzero column, and $\mathrm{Tot}^{\oplus}(C) \cong \Sigma^{-1}\,\mathrm{cone}(\varphi)$.

Theorem 14.31 (Balancing Tor)**.** *Let R be a ring, M and N be R-modules, and fix a projective resolution P of M and a projective resolution Q of N. For every n, there is an isomorphism*

$$L_n(M \otimes_R -)(N) = \mathrm{H}_n(M \otimes_R Q) \cong \mathrm{H}_n(P \otimes_R N) = L_n(- \otimes_R N)(M).$$

Proof. We have surjections $\pi : P_0 \twoheadrightarrow M$ and $\varepsilon : Q_0 \twoheadrightarrow N$. Consider the double complex $P \otimes Q$, which is a first quadrant double complex:

$$\begin{tikzcd}
\vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \\
P_0 \otimes Q_2 \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_2 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
P_0 \otimes Q_1 \arrow[d, "1\otimes\partial^Q"'] & P_1 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
P_0 \otimes Q_0 & P_1 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] & P_2 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] & \cdots \arrow[l]
\end{tikzcd}$$

Each P_i and Q_i is projective and thus flat, by Theorem 11.44, so $P_i \otimes_R -$ and $- \otimes_R Q_i$ are both exact functors. The rows and columns of our double complex are thus exact everywhere except for the 0th row and column. We can complete our double complex to make a double complex C with both exact rows if we add in a column induced by the surjection π:

$$
\begin{tikzcd}
\vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] \\
M \otimes Q_2 \arrow[d] & P_0 \otimes Q_2 \arrow[l, "\pi\otimes 1"] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_2 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
M \otimes Q_1 \arrow[d] & P_0 \otimes Q_1 \arrow[l, "\pi\otimes 1"] \arrow[d, "1\otimes\partial^Q"'] & P_1 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
M \otimes Q_0 & P_0 \otimes Q_0 \arrow[l, "\pi\otimes 1"] & P_1 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] & P_2 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] & \cdots \arrow[l]
\end{tikzcd} .
$$

Similarly, we can make a double complex D with exact columns by adding in a row induced by ε:

$$
\begin{tikzcd}
\vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] \\
P_0 \otimes Q_2 \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_2 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
P_0 \otimes Q_1 \arrow[d, "1\otimes\partial^Q"'] & P_1 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & P_2 \otimes Q_1 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\partial^Q"'] & \cdots \arrow[l] \\
P_0 \otimes Q_0 \arrow[d, "1\otimes\varepsilon"'] & P_1 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\varepsilon"'] & P_2 \otimes Q_0 \arrow[l, "\partial^P\otimes 1"'] \arrow[d, "1\otimes\varepsilon"'] & \cdots \arrow[l] \\
P_0 \otimes N & P_1 \otimes N \arrow[l] & P_2 \otimes N \arrow[l] & \cdots \arrow[l]
\end{tikzcd} .
$$

By Lemma 14.26, $\mathrm{Tot}^{\oplus}(C)$ and $\mathrm{Tot}^{\oplus}(D)$ are both exact. Notice that $\pi \otimes Q$ is a map of complexes $\mathrm{Tot}^{\oplus}(P \otimes Q) \longrightarrow M \otimes Q$, and $P \otimes \varepsilon$ is a map of complexes $\mathrm{Tot}^{\oplus}(P \otimes Q) \longrightarrow P \otimes N$. The mapping cone of $\pi \otimes Q$ is precisely $\Sigma\,\mathrm{Tot}^{\oplus}(C)$, while the mapping cone of $P \otimes \varepsilon$ is precisely $\Sigma\,\mathrm{Tot}^{\oplus}(D)$. Since these are both exact, Remark 14.29 says that $\mathrm{Tot}^{\oplus}(P \otimes Q) \xrightarrow{\pi\otimes Q} M \otimes Q$ and $\mathrm{Tot}^{\oplus}(P \otimes Q) \xrightarrow{P\otimes\varepsilon} P \otimes N$ are quasi-isomorphisms, so that

$$L_n(M \otimes_R -)(N) = \mathrm{H}_n(M \otimes_R Q) \cong \mathrm{H}_n(P \otimes_R N) = L_n(- \otimes_R N)(M). \qquad \square$$

Theorem 14.32 (Balancing Ext)**.** *Let R be a ring, M and N be R-modules, and fix a projective resolution P of M and an injective resolution E of N. For every n, there is an isomorphism*

$$R^n \operatorname{Hom}_R(M, -)(N) = \mathrm{H}^n(\operatorname{Hom}_R(M, E)) \cong \mathrm{H}^n(\operatorname{Hom}_R(P, N)) = R^n \operatorname{Hom}_R(-, N)(M).$$

Proof. We have a surjection $\pi : P_0 \longrightarrow M$ and an inclusion $\varepsilon : M \longrightarrow E_0$. The double cocomplex $\mathrm{Hom}_R(P,E)$ with $\mathrm{Hom}_R(P,E)_{p,q} = \mathrm{Hom}_R(P_p,E^q)$ and

$$\begin{array}{ccc} \mathrm{Hom}_R(P_p,E^q) \xrightarrow{d^h} \mathrm{Hom}_R(P_{p+1},E^q) & \text{and} & \mathrm{Hom}_R(P_p,E^q) \xrightarrow{d^v} \mathrm{Hom}_R(P_p,D_{q+1})\ . \\ f \longmapsto f\circ\partial^P & & f \longmapsto (-1)^{p+q+1}\partial^E\circ f \end{array}$$

is a first quadrant double cocomplex:

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
\mathrm{Hom}_R(P_0,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^2) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots
\end{tikzcd}$$

We proceed just like in Theorem 14.31, by considering the double cocomplex C

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \vdots & \\
\mathrm{Hom}(M,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_0,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^2) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}(M,E^2) \arrow[r] \arrow[u] & Hom_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}(M,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots
\end{tikzcd}$$

obtained by adding in a column induced by π, and the double cocomplex D

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
\mathrm{Hom}_R(P_0,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^2) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^2) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}_R(P_0,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,E^1) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,E^1) \arrow[r] \arrow[u] & \cdots \\
\mathrm{Hom}_R(P_0,N) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_1,N) \arrow[r] \arrow[u] & \mathrm{Hom}_R(P_2,N) \arrow[r] \arrow[u] & \cdots
\end{tikzcd}$$

obtained by adding in a row induced by ε. Now we notice that the cone of $\operatorname{Hom}_R(P,N) \longrightarrow \operatorname{Tot}^{\oplus}(\operatorname{Hom}(P,E))$ is exactly $\operatorname{Tot}^{\oplus}(C)$, while the cone of $\operatorname{Hom}_R(M,E) \longrightarrow \operatorname{Tot}^{\oplus}(\operatorname{Hom}(P,E))$ is exactly $\operatorname{Tot}^{\oplus}(C)$

The dual of Lemma 14.26 says that $\operatorname{Tot}^{\oplus}(C)$ and $\operatorname{Tot}^{\oplus}(D)$ are both exact, and thus $\operatorname{Hom}_R(P,N) \longrightarrow \operatorname{Tot}^{\oplus}(\operatorname{Hom}(P,E))$ and $\operatorname{Hom}_R(M,E) \longrightarrow \operatorname{Tot}^{\oplus}(\operatorname{Hom}(P,E))$ are both quasi-isomorphisms. We conclude that

$$R^n \operatorname{Hom}_R(M,-)(N) = \mathrm{H}^n(\operatorname{Hom}_R(P,N)) \cong \mathrm{H}^n(\operatorname{Hom}_R(M,E)) = R^n \operatorname{Hom}_R(-,N)(M). \qquad \square$$

Definition 14.33. Let R be a ring and M and N be R-modules. The ith Ext module from M to N is

$$\operatorname{Tor}_i^R(M,N) := L_i(M \otimes_R -)(N) \cong L_i(- \otimes_R N)(M).$$

Notice in particular that the R-module $\operatorname{Tor}_i^R(M,N)$ is defined only up to isomorphism.

Definition 14.34. Let R be a ring and M and N be R-modules. The ith Ext module from M to N is

$$\operatorname{Ext}_R^i(M,N) := R^i \operatorname{Hom}_R(M,-)(N) \cong L^i \operatorname{Hom}_R(-,N)(M).$$

Notice in particular that the R-module $\operatorname{Ext}_R^i(M,N)$ is only defined up to isomorphism.

Theorem 14.19 immediately gives us long exact sequences for Ext and Tor.

Theorem 14.35. *Let R be a ring and M an R-module. Every short exact sequence in* $\operatorname{Ch}(R)$

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

induces a long exact sequence

$$\begin{aligned} \cdots \longrightarrow \operatorname{Tor}_{n+1}^R(M,C) \longrightarrow \operatorname{Tor}_n^R(M,A) \longrightarrow \operatorname{Tor}_n^R(M,B) \longrightarrow \operatorname{Tor}_n^R(M,A) \longrightarrow \cdots \\ \cdots \longrightarrow \operatorname{Tor}_1^R(M,C) \longrightarrow A \otimes_R M \longrightarrow B \otimes_R M \longrightarrow C \otimes_R M \longrightarrow 0. \end{aligned}$$

Theorem 14.36. *Let R be a ring and M an R-module. Every short exact sequence in* $\operatorname{Ch}(R)$

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

induces a natural long exact sequence

$$\begin{aligned} 0 \longrightarrow \operatorname{Hom}_R(M,A) \longrightarrow \operatorname{Hom}_R(M,B) \longrightarrow \operatorname{Hom}_R(M,C) \longrightarrow \operatorname{Ext}_R^1(M,A) \longrightarrow \cdots \\ \cdots \longrightarrow \operatorname{Ext}_R^n(M,B) \longrightarrow \operatorname{Ext}_R^n(M,C) \longrightarrow \operatorname{Ext}_R^{n+1}(M,A) \longrightarrow \cdots . \end{aligned}$$

and

$$\begin{aligned} 0 \longrightarrow \operatorname{Hom}_R(C,M) \longrightarrow \operatorname{Hom}_R(B,M) \longrightarrow \operatorname{Hom}_R(A,M) \longrightarrow \operatorname{Ext}_R^1(C,M) \longrightarrow \cdots \\ \cdots \longrightarrow \operatorname{Ext}_R^n(B,M) \longrightarrow \operatorname{Ext}_R^n(A,M) \longrightarrow \operatorname{Ext}_R^{n+1}(C,M) \longrightarrow \cdots . \end{aligned}$$

Finally, we note that $\operatorname{Tor}_i^R(M,N) \cong \operatorname{Tor}_i^R(M,N)$, just like $M \otimes_R N \cong N \otimes_R M$.

Theorem 14.37. *Let M and N be R-modules. For all i, there are natural isomorphisms*

$$\operatorname{Tor}_i^R(M,N) \cong \operatorname{Tor}_i^R(N,M).$$

Proof. Let P be a projective resolution of M. By Theorem 14.31, $\operatorname{Tor}_i^R(M,N) = \mathrm{H}_i(P \otimes_R N)$ and $\operatorname{Tor}_i^R(N,M) = \mathrm{H}_i(N \otimes_R P)$. By Lemma 10.20, $M \otimes_R N$ and $N \otimes_R M$ are naturally isomorphic. In fact, $m \otimes n \mapsto n \otimes m$ determines an isomorphism. So consider the map

$$\begin{array}{ccc} P_n \otimes_R N \xrightarrow{f_n} N \otimes_R P_n & \qquad & N \otimes_R M \xrightarrow{g_n} P_n \otimes_R N \\ m \otimes n \longmapsto n \otimes m & \qquad & n \otimes m \longmapsto m \otimes n \end{array}$$

which again are isomorphisms for all n. Notice that these f_n assemble into a map of complexes $P \otimes_R N \xrightarrow{f} N \otimes_R P$, since

$$f_n(\partial(m \otimes n)) = f_n(\partial(m) \otimes n) = n \otimes \partial(m) = \partial(n \otimes m) = \partial f_{n+1}(m \otimes n).$$

Since all the f_n are isomorphisms, f is an isomorphism of complexes, and must then induce isomorphisms in homology. We conclude that

$$\operatorname{Tor}_i^R(M,N) = \mathrm{H}_i(P \otimes_R N) \cong \mathrm{H}_i(N \otimes_R P) = \operatorname{Tor}_i^R(N,M).$$

$\square$

14.4 Computing Ext and Tor

When computing explicit Ext and Tor modules, it helps to keep a few simple ideas in mind:

- If P is a projective R-module, then $\operatorname{Tor}_i^R(M,P) = \operatorname{Tor}_i^R(P,M) = 0$ and $\operatorname{Ext}_R^i(P,M) = 0$ for all $i > 0$ and all R-modules M, since $0 \longrightarrow P \longrightarrow 0$ is a projective resolution for M.
- If E is an injective R-module, $\operatorname{Ext}_R^i(M,E) = 0$ for all $i > 0$ and all R-modules M.
- Free resolutions are often easier to compute explicitly, and the best path towards finding $\operatorname{Ext}_R^n(M,N)$.
- Relating one of our modules to other, easier modules via a short exact sequence can often simplify complicated computations.

Let's compute some examples.

Example 14.38. Let's compute $\operatorname{Ext}_{\mathbb{Z}}^i(\mathbb{Z}/(2), \mathbb{Z}/(3))$. Injective resolutions are not so easy to find, so we start from a projective resolution for $\mathbb{Z}/(2)$:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2} \mathbb{Z} \longrightarrow \mathbb{Z}/(2) \longrightarrow 0.$$

Notice that $\operatorname{pdim}_{\mathbb{Z}}(\mathbb{Z}/(2)) \neq 0$, since $\mathbb{Z}/(2)$ is not a projective $\mathbb{Z}$-module. We found a free resolution of length 1 for $\mathbb{Z}/(2)$, so it must be that $\operatorname{pdim}_{\mathbb{Z}}(\mathbb{Z}/(2)) = 1$. This immediately tells us that $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(2), \mathbb{Z}/(3)) = 0$ for all $i \leqslant 2$. Now we apply $\mathrm{Hom}_{\mathbb{Z}}(-, \mathbb{Z}/(3))$ to our free resolutions for $\mathbb{Z}/(2)$, and obtain

$$0 \longrightarrow \underset{0}{\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/(3))} \xrightarrow{2^*} \underset{1}{\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/(3))} \longrightarrow 0.$$

By Exercise 31, $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/(3)) \cong \mathbb{Z}/(3)$, via the isomorphism $f \mapsto f(1)$. Since 2^* was the map $f \mapsto (2 \cdot -) \circ f = 2f(-)$, we can simplify our complex to

$$0 \longrightarrow \mathbb{Z}/(3) \xrightarrow{2} \mathbb{Z}/(3) \longrightarrow 0.$$

Notice that multiplication by 2 is an isomorphism on $\mathbb{Z}/(3)$, so the complex above is exact, and $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(2), \mathbb{Z}/(3)) = 0$ for all i.

Example 14.39. Given an integer $n > 1$,

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/(n) \longrightarrow 0$$

with π the canonical projection is a free resolution for $\mathbb{Z}/(n)$ over $\mathbb{Z}$. Notice that since $\mathbb{Z}/(n)$ is not a free $\mathbb{Z}$-module, there is no shorter free resolution for $\mathbb{Z}/\mathfrak{n}$. Now we can use this resolution to compute $\mathrm{Tor}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M)$ and $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M)$ for any $\mathbb{Z}$-module M. For Tor,

$$\mathrm{Tor}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M) = \mathrm{H}_i(\ 0 \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} M \xrightarrow{n \otimes 1} \mathbb{Z} \otimes_{\mathbb{Z}} M \longrightarrow 0\ ).$$

By Lemma 10.22 $\mathbb{Z} \otimes_{\mathbb{Z}} M \cong M$, via the map $k \otimes m \mapsto km$, and the map $n \otimes 1_M$ corresponds to multiplication by n on M. Therefore,

$$\mathrm{Tor}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M) = \mathrm{H}_i(\ 0 \longrightarrow M \xrightarrow{n} M \longrightarrow 0\ ),$$

so

$$\mathrm{Tor}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M) = \begin{cases} M/nM & \text{for } i = 0 \\ (0 :_M n) & \text{for } i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $\mathrm{Tor}^{\mathbb{Z}}_0(\mathbb{Z}/(n), M) = M/nM = \mathbb{Z}/n\mathbb{Z} \otimes_{\mathbb{Z}} M$, as we already knew from Proposition 14.17.

Similarly, we can compute all the Ext modules from $\mathbb{Z}/(n)$:

$$\mathrm{Ext}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M) = \mathrm{H}_i(\ 0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \xrightarrow{n^*} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \longrightarrow 0\ ).$$

By Exercise 31, $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \cong M$, via the map $f \mapsto f(1)$, and $n^* = \mathrm{Hom}_{\mathbb{Z}}(n, M)$ corresponds to multiplication by n on M. So

$$\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M) = \mathrm{H}^i(\ 0 \longrightarrow M \xrightarrow{n} M \longrightarrow 0\ ).$$

We conclude that

$$\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M) = \begin{cases} M/nM & \text{for } i = 1 \\ (0 :_M n) & \text{for } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/(n), M) = (0 :_M n) = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/(n), M)$, as we already knew from Proposition 14.17.

Alternatively, we can compute $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M)$ and $\operatorname{Tor}^{\mathbb{Z}}_i(\mathbb{Z}/(n), M)$ by looking at some long exact sequences. The long exact sequence for Tor induced by the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \longrightarrow \mathbb{Z}/(n) \longrightarrow 0$$

is

$$\cdots \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_{n+1}(\mathbb{Z}/(n), M) \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_n(\mathbb{Z}, M) \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_n(\mathbb{Z}, M) \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_n(\mathbb{Z}/(n), M) \longrightarrow \cdots$$
$$\cdots \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_1(\mathbb{Z}/(n), M) \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} M \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} M \longrightarrow \mathbb{Z}/(n) \otimes_{\mathbb{Z}} M \longrightarrow 0.$$

Since $\mathbb{Z}$ is a projective $\mathbb{Z}$-module and thus flat, $\operatorname{Tor}^{\mathbb{Z}}_i(\mathbb{Z}, M) = 0$ for all $i > 0$. As a consequence, the long exact sequence above forces $\operatorname{Tor}^{\mathbb{Z}}_2(\mathbb{Z}/(n), M) = 0$. So our long exact sequence really gets reduced to

$$0 \longrightarrow \operatorname{Tor}^{\mathbb{Z}}_1(\mathbb{Z}/(n), M) \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} M \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} M \longrightarrow \mathbb{Z}/(n) \otimes_{\mathbb{Z}} M \longrightarrow 0.$$

Now $\mathbb{Z} \otimes_{\mathbb{Z}} M \cong M$ via $k \otimes m \mapsto km$, and this isomorphism turns $n \otimes 1_M$ into multiplication by n on M, same as above. So $\operatorname{Tor}^{\mathbb{Z}}_1(\mathbb{Z}/(n), M)$ is the kernel of multiplication by n on M, or $(0 :_M n)$.

If we want to compute $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M)$, we should now look at the long exact sequence

$$0 \longrightarrow \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/(n), M) \longrightarrow \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \xrightarrow{n^*} \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \longrightarrow \operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/(n), M) \longrightarrow \cdots$$
$$\cdots \longrightarrow \operatorname{Ext}^n_{\mathbb{Z}}(\mathbb{Z}, M) \longrightarrow \operatorname{Ext}^n_{\mathbb{Z}}(\mathbb{Z}, M) \longrightarrow \operatorname{Ext}^{n+1}_{\mathbb{Z}}(\mathbb{Z}/(n), M) \longrightarrow \cdots.$$

Again, $\mathbb{Z}$ is a free $\mathbb{Z}$-module, so $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}, M) = 0$ for all $i > 0$. Then $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/(n), M) = 0$ for all $i > 1$, and our long exact sequence is actually just

$$0 \longrightarrow \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/(n), M) \longrightarrow \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \xrightarrow{n^*} \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \longrightarrow \operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/(n), M) \longrightarrow 0.$$

So $\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/(n), M)$ is the cokernel of n^*. As before, notice that $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \cong M$ via the map $f \mapsto f(1)$, and n^* corresponds to multiplication by n on M. We conclude that $\operatorname{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/(n), M) \cong M/nM$.

14.5 Other derived functors

Here are some other examples of derived functors you may encounter.

Group homology and group cohomology

Definition 14.40. Let G be a group. A **(left) G-module** is an abelian group A with an action of G by additive maps on the left, meaning that

$$g(a + b) = ga + gb$$

for all $a, b \in A$ and all $g \in G$, where we write ga for the action of $g \in G$ on $a \in A$. Given two G-modules A and B, a morphism of G-modules $A \xrightarrow{f} B$ is a group homomorphism that is also G-equivariant, meaning $f(ga) = gf(a)$ for all $g \in G$ and $a \in A$. The **category of G-modules G-mod** has objects all G-modules and arrows all G-module morphisms. We write $\mathrm{Hom}_G(A, B)$ instead of $\mathrm{Hom}_{G\textbf{-mod}}(A, B)$.

This category can be identified with the category of $\mathbb{Z}G$-modules, of modules over the (noncommutative) ring $\mathbb{Z}G$, the group ring of G. It can also be identified with the functor category $\mathbf{Ab}^G$ of functors from the category with one object G and arrows the elements of G to the category $\mathbf{Ab}$ of abelian groups.

Definition 14.41. The **invariant subgroup** A^G of a G-module A is

$$A^G := \{a \in A \mid ga = a \text{ for all } g \in G\}.$$

The **coinvariant subgroup** A^G of a G-module A is

$$A_G := A/G\text{-submodule generated by } \{ga - a \in A \mid g \in G, a \in A\}.$$

Exercise 57. Given any G-module A, $A_G \cong \mathbb{Z} \otimes_{\mathbb{Z}G} A$ and $A^G \cong \mathrm{Hom}_G(\mathbb{Z}, A)$, where $\mathbb{Z}$ denotes the trivial G-module.

This automatically tells us that taking coinvariants is right exact, and taking invariants is left exact.

Definition 14.42. Let G be a group and A a G-module. The **homology groups of G with coefficients in A** are the G-modules $\mathrm{H}_i(G, A)$ obtained via the left derived functors of the coinvariants functor:

$$\mathrm{H}_i(G; A) := L_i(-_G)(A).$$

Similarly, the **cohomology groups of G with coefficients in A** are the G-modules $\mathrm{H}^i(G, A)$ obtained via the right derived functors of the invariants functor:

$$\mathrm{H}^i(G; A) := R^i(-^G)(A).$$

By Exercise 57,

$$\mathrm{H}_i(G; A) \cong \mathrm{Tor}_i^{\mathbb{Z}G}(\mathbb{Z}, A) \text{ and } \mathrm{H}^i(G; A) \cong \mathrm{Ext}^i_{\mathbb{Z}G}(\mathbb{Z}, A).$$

By Proposition 14.17, $\mathrm{H}_0(G; A) = A_G$ and $\mathrm{H}^0(G; A) = A^G$.

For a detailed treatment of group (co)homology, see Weibel's Homological Algebra [Wei94].

Local Cohomology

Let I be an ideal in a ring R. The **I-torsion** functor $\Gamma_I \colon R\textbf{-mod} \longrightarrow R\textbf{-mod}$ is defined by

$$\Gamma_I(M) := \{m \in M \mid I^n m = 0 \text{ for some } n\}$$

which acts on maps by restriction. This functor is left exact, and it gives rise to local cohomology, the right derived functors H^i_I of Γ_I. The ith local cohomology of M with support on I is then given by

$$\mathrm{H}^i_I(M) := R^i\Gamma_I(M).$$

Local cohomology was introduced by Grothendieck in a series of famous seminars at Harvard in 1961. Grothendieck himself never published any notes on the subject, but Robin Hartshorne's notes of those lectures have been published. For a modern treatment of the subject and its connections, the book *24 hours of local cohomology* [ILL+07] and the very nice notes by Craig Huneke, Mel Hochster, and Jack Jeffries are excellent resources.

Local cohomology modules play a crucial, ubiquitous role in commutative algebra. They measure many important invariants, such as dimension and depth, and are extremely useful tools for studying all sorts of topics; for example, they can be used to detect if a ring is Gorenstein (meaning, if it has finite injective dimension as a module over itself) or Cohen-Macaulay (a nice class we will introduce in the next chapter). However, local cohomology modules are typically not finitely generated, a departure with most of the commutative algebra ideas we have studied so far. One reason for this is that injective modules are also often not finitely generated. Local cohomology is also a major reason why commutative algebraists are interested in studying injective modules.

In fact, local cohomology is almost *never* finitely generated. Here's a very simple example.

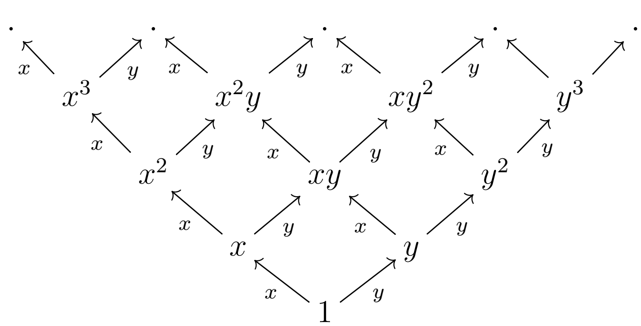
Example 14.43. Let $R = k[x_1, \ldots, x_n]$, k be a field, and $\mathfrak{m} = (x_1, \ldots, x_n)$. Then $\mathrm{H}^n_{\mathfrak{m}}(R)$ has the k-vector space structure

$$\bigoplus_{\text{all } a_i > 0} k \cdot \frac{1}{x_1^{a_1} \cdots x_n^{a_n}},$$

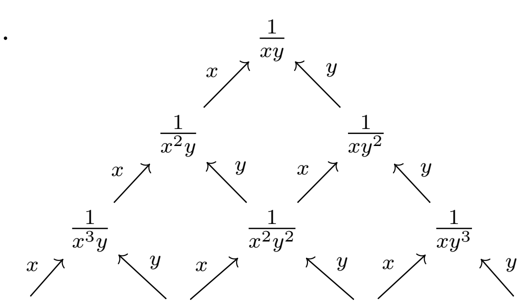
with R-module structure given by

$$x_1^{b_1} \cdots x_n^{b_n} \cdot \frac{z}{x_1^{a_1} \cdots x_n^{a_n}} = \begin{cases} \frac{z}{x_1^{a_1 - b_1} \cdots x_n^{a_n - b_n}} & \text{if all } b_i < a_i \\ 0 & \text{otherwise.} \end{cases}$$

This is not a finitely generated module! Note also that every finitely generated submodule only has terms with bounded negative degree. But this is still a very nice module: it looks like R upside down.



$k[x, y]$



$\mathrm{H}^2_{(x,y)}(k[x, y])$

Despite being infinitely generated, local cohomology modules enjoy many finiteness properties we have gotten used to expecting from finitely generated modules. For example, over a local ring $(R, \mathfrak{m})$, the local cohomology modules $\mathrm{H}^i_{\mathfrak{m}}(M)$ of a finitely generated module M are Artinian — but not Noetherian!

Huneke raised the question of whether local cohomology modules of Noetherian rings always have finitely many associated primes, a problem which has been a very active research are in commutative algebra in the last few decades. While the answer to Huneke's question is no — as famous examples by Katzmann, Singh, and Singh and Swanson show — the local cohomology modules of finitely generated R-modules over a regular ring do have finitely many associated primes.

One very important invariant we can study with local cohomology is the arithmetic rank.

Definition 14.44. Let I be an ideal in a Noetherian ring R. The **arithmetic rank** of I is defined by

$$\operatorname{ara}(I) := \min\{s \mid \text{there exist some } x_1, \ldots, x_s \text{ such that } \sqrt{(x_1, \ldots, x_s)} = \sqrt{I}\}.$$

Given a variety $X = V(I) \subseteq \mathbb{A}^n_k$, the arithmetic rank of its defining ideal $I(X)$ is the minimum number of equations needed to define X. It turns out that this number is difficult to study, and it is best understood via local cohomology, a thought best described by Lyubeznik:

> Part of what makes the problem about the number of defining equations so interesting is that it can be very easily stated, yet a solution, in those rare cases when it is known, usually is highly nontrivial and involves a fascinating interplay of Algebra and Geometry.
>
> (Lyubeznik, in [Lyu92])

The connection to local cohomology begins with the following two elementary facts about local cohomology:

- If $\sqrt{I} = \sqrt{J}$, then $\mathrm{H}^i_I(-) = \mathrm{H}^i_J(-)$.
- Given any ideal I, $\operatorname{ara}(I) \geqslant \min\{i \mid \mathrm{H}^i_I(M) \neq 0 \text{ for some } R\text{-module } M\}$.

It turns out that local cohomology modules can be defined in a few different ways, which are in no way obviously equivalent, and those different points of view are quite helpful. For example, we can define local cohomology via the Čech complex.

Definition 14.45 (Čech complex)**.** Let M be an R-module and $x \in R$. The **Čech complex** of x on R is given by

$$\check{C}^\bullet(x) := \left(0 \longrightarrow \underset{0}{R} \longrightarrow \underset{1}{R_x} \longrightarrow 0 \right)$$

The **Čech complex** of $f_1, \ldots, f_t \in R$ on M is given by

$$\check{C}^\bullet(f_1^n, \ldots, f_t^n; M) := \check{C}^\bullet(f_1) \otimes \cdots \otimes \check{C}^\bullet(f_t) \otimes M.$$

Example 14.46. Let's compute the Čech complex on f and g and an R-module M.

$$\check{C}^\bullet(f,g;M) \quad = \quad \begin{tikzcd} & 0 & 0 & \\ 0 \arrow[r] & M_g \arrow[u] \arrow[r, "-1"'] & M_{fg} \arrow[u] \arrow[r] & 0 \\ 0 \arrow[r] & M \arrow[u, "1"] \arrow[r, "1"'] & M_f \arrow[u, "1"'] \arrow[r] & 0 \\ & 0 \arrow[u] & 0 \arrow[u] & \end{tikzcd} \quad = \quad 0 \longrightarrow M \xrightarrow{\binom{1}{1}} M_f \oplus M_g \xrightarrow{(1\ -1)} M_{fg} \longrightarrow 0$$

Exercise 58.

a) $\check{C}^\bullet(f_1,\dots,f_t;M) \cong \bigoplus_{\{j_1,\dots,j_i\}\subseteq[t]} M_{f_{j_1}\cdots f_{j_i}}$

b) The maps between components corresponding to subsets I, J are zero if $I \not\subseteq J$, and ± 1 if $J = I \cup \{k\}$.

As we mentioned above, it turns out that the cohomology of the Čech complex gives us local cohomology. If $I = (f_1,\dots,f_n)$ is an n-generated ideal, then

$$\begin{aligned} \mathrm{H}^n_I(M) &= \mathrm{H}^i(\check{C}^\bullet(f_1,\dots,f_t;M)) \\ &= \text{ cohomology of } \left(0 \to M \to \cdots \to \bigoplus_i M_{f_i} \to \cdots \bigoplus_{i=1}^{n} M_{f_1\cdots\widehat{f_i}\cdots f_n} \to M_{f_1\cdots f_n} \to 0\right) \end{aligned}$$

so elements in the n-th local cohomology can be realized as equivalence classes of fractions.

Local cohomology is also closely related to Ext.

Definition 14.47. Given a directed system of modules

$$(M_i)_{i\in\mathbb{N}} = \left(\cdots \longrightarrow M_i \longrightarrow M_{i+1} \longrightarrow \cdots\right)$$

its **direct limit** is the module $M = \varinjlim_{i\in\mathbb{N}} M_i$, equipped with maps $M_i \xrightarrow{\iota_i} M$ for all i, satisfying the property that if there are maps $M_i \xrightarrow{\alpha_i} N$ that commute with the maps in the system, then there is a unique map $M \xrightarrow{\varphi} N$ such that $\alpha_i = \varphi \circ \iota_i$ for all i.

This notion is functorial: given two directed systems of complexes, and maps of complexes for each i that commute with the maps in the systems, there is an induced map on the direct limits. In fact, direct limits are a special case of the categorical notion of a colimit.

Direct limits can be realized explicitly as follows:

- Every element is represented by a class (m,i) with $m \in M_i$.
- Two classes $(m,i), (n,j)$ are the same if and only if for some $k \geqslant \max\{i,j\}$, the images of m and n under the composed transition maps agree in M_k.

In particular, an element represents the zero class if and only if it is in the kernel of a large composition of the transition map.

Similar considerations hold for systems indexed by an arbitrary poset P; this consists of a collection of modules M_p for $p \in P$, and commuting maps $M_p \to M_q$ for all $p \leq q$.

It turns out that local cohomology modules also arise as a direct limit of Ext modules:

$$\varinjlim_n \operatorname{Ext}^i_R(R/I^n, M)$$

The equivalence between all these different definitions is a fundamental result in the theory of local cohomology.

Appendix A

Macaulay2

There are several computer algebra systems dedicated to algebraic geometry and commutative algebra computations, such as Singular (more popular among algebraic geometers), CoCoA (which is more popular with european commutative algebraists, having originated in Genova, Italy), and Macaulay2. There are many computations you could run on any of these systems (and others), but we will focus on Macaulay2 since it's the most popular computer algebra system among US based commutative algebraists.

Macaulay2, as the name suggests, is a successor of a previous computer algebra system named Macaulay. Macaulay was first developed in 1983 by Dave Bayer and Mike Stillman, and while some still use it today, the system has not been updated since its final release in 2000. In 1993, Daniel Grayson and Mike Stillman released the first version of Macaulay2, and the current stable version if Macaulay2 1.16.

Macaulay2, or M2 for short, is an open-source project, with many contributors writing packages that are then released with the newest Macaulay2 version. Journals like the *Journal of Software for Algebra and Geometry* publish peer-refereed short articles that describe and explain the functionality of new packages, with the package source code being peer reviewed as well.

The National Science Foundation has funded Macaulay2 since 1992. Besides funding the project through direct grants, the NSF has also funded several Macaulay2 workshops — conferences where Macaulay2 package developers gather to work on new packages, and to share updates to the Macaulay2 core code and recent packages.

A.1 Getting started

A Macaulay2 session often starts with defining some ambient ring we will be doing computations over. Common rings such as the rationals and the integers can be defined using the commands `QQ` and `ZZ`; one can easily take quotients or build polynomial rings (in finitely many variables) over these. For example,

```
i1 : R = ZZ/101[x,y]

o1 = R
```

```
o1 : PolynomialRing
```

and

```
i1 : k = ZZ/101;

i2 : R = k[x,y];
```

both store the ring $\mathbb{Z}/101$ as R, with the small difference that in the second example Macaulay2 has named the coefficient field k. One quirk that might make a difference later is that if we use the first option and later set k to be the field $\mathbb{Z}/101$, our ring R is *not* a polynomial ring over k. Also, in the second example we ended each line with a `;`, which tells Macaulay2 to run the command but not display the result of the computation — which is in this case was simply an assignment, so the result is not relevant. Lines indicated with as in, where n is some integer, are input lines, whereas lines with an on indicate output lines.

We can now do all sorts of computations over our ring R. We can define ideals in R, and use them to either define a quotient ring S of R or an R-module M, as follows:

```
i3 : I = ideal(x^2,y^2,x*y)

               2   2
o3 = ideal (x , y , x*y)

o3 : Ideal of R

i4 : M = R^1/I

o4 = cokernel | x2 y2 xy |

                               1
o4 : R-module, quotient of R

i5 : S = R/I

o5 = S

o5 : QuotientRing
```

It's important to note that while R is a ring, R^1 is the R-module R — this is a very important difference for Macaulay2, since these two objects have different types. So S defined above is a ring, while M is a module. Notice that Macaulay2 stored the module M as the cokernel of the map

$$R^3 \xrightarrow{\begin{bmatrix} x^2 & y^2 & xy \end{bmatrix}} R\,.$$

Note also that there is an alternative syntax to write our ideal I from above, as follows:

```
i15 : I = ideal"x2,xy,y2"

              2       2
o15 = ideal (x , x*y, y )

o15 : Ideal of R
```

When you make a new definition in Macaulay2, you might want to pay attention to what ring your new object is defined over. For example, now that we defined this ring S, Macaulay2 has automatically taken S to be our current ambient ring, and any calculation or definition we run next will be considered over S and not R. If you want to return to the original ring R, you must first run the command `use S`.

If you want to work over a finitely generated algebra over one of the basic rings you can define in Macaulay2, and your ring is not a quotient of a polynomial ring, you want to rewrite this algebra as a quotient of a polynomial ring. For example, suppose you want to work over the 2nd Veronese in 2 variables over our field k from before, meaning the algebra $k[x^2, xy, y^2]$. We need 3 algebra generators, which we will call a, b, c, corresponding to x^2, xy, and y^2:

```
i11 : U = k[a,b,c]

o11 = U

o11 : PolynomialRing

i12 : f = map(R,U,{x^2,x*y,y^2})
                 2       2
o12 = map(R,U,{x , x*y, y })

o12 : RingMap R <--- U

i13 : J = ker f

             2
o13 = ideal(b  - a*c)

o13 : Ideal of U

i14 : T = U/J

o14 = T

o14 : QuotientRing
```

Our ring T at the end is isomorphic to the 2nd Veronese of R, which is the ring we wanted.

A.2 Basic commands

Many Macaulay2 commands are easy to guess, and named exactly what you would expect them to be named. If you are not sure how to use a certain command, you can run `viewHelp` followed by the command you want to ask about; this will open an html file with the documentation for the method you asked about. Often, googling "Macaulay2" followed by descriptive words will easily land you on the documentation for whatever you are trying to do.

Here are some basic commands you will likely use:

- `ideal`$(f_1, \ldots, f_n)$ will return the ideal generated by $f_1, \ldots, f_n$. Here products should be indicated by $*$, and powers with $\hat{}$. If you'd rather not use $\hat{}$ (this might be nice if you have lots of powers), you can write `ideal`$(f_1, \ldots, f_n)$ instead.
- `map`$(S, R, f_1, \ldots, f_n)$ gives a ring map $R \to S$ if R and S are rings, and R is a quotient of $k[x_1, \ldots, x_n]$. The resulting ring map will send $x_i \mapsto f_i$. There are many variations of `map` — for example, you can use it to define R-module homomorphisms — but you should carefully input the information in the required format. Try `viewHelp map` in Macaulay2 for more details
- `ker`(f) returns the kernel of the map f.
- `I + J` and `I*J` return the sum and product of the ideals I and J, respectively.
- `A = matrix`$\{\{a_{1,1}, \ldots, a_{1,n}\}, \ldots, \{a_{m,1}, \ldots, a_{m,n}\}\}$ returns the matrix

$$A = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ & \ddots & \\ a_{m,1} & \ldots & a_{m,n} \end{pmatrix}$$

A.3 Complexes in Macaulay2

There are two different ways to do computations involving complexes in Macaulay2: using `ChainComplexes`, or the new (and still incomplete) `Complexes` package. To use `Complexes`, you must first load the `Complexes` package, while the `ChainComplexes` methods are automatically loaded with Macaulay2.

A.3.1 Chain Complexes

To create a new chain complex by hand, we start by setting up R-module maps.

```
i1 : R = QQ[x,y]/ideal"xy";

i2 : d0 = map(R^1,R^1,matrix{{x}})

o2 = | x |

               1        1
o2 : Matrix R  <--- R
```

To make sure we set up the next map in a way that is composable with d_0, we can use the methods `source` and `target`:

```
i3 : d1 = map(source d0, R^1, matrix{{y}})

o3 = | y |
             1       1
o3 : Matrix R  <--- R
```

We can also double check our maps do indeed map a complex:

```
i4 : d1 ** d0 == 0

o4 = true
```

So now we are ready to set up our new chain complex.

```
i5 : C = new ChainComplex

o5 = 0

o5 : ChainComplex

i6 : C#0 = target d0
      1
o6 = R

o6 : R-module, free

i7 : C#1 = target d1

      1
o7 = R

o7 : R-module, free

i8 : C#2 = source d1
      1
o8 = R

o8 : R-module, free

i9 : C.dd_1 = d0

o9 = | x |
             1       1
o9 : Matrix R  <--- R
```

```
i10 : C.dd_2 = d1

o10 = | y |
                1       1
o10 : Matrix R  <--- R
```

Given a chain complex `C`, we can ask Macaulay2 what our complex is by simply running the name of the complex:

```
i11 : C

       1      1      1
o11 = R  <-- R  <-- R

      0      1      2

o11 : ChainComplex
```

Or we can ask for a better visual description of the maps, using `C.dd`:

```
i12 : C.dd

          1             1
o12 = 0 : R  <--------- R  : 1
                | x |

          1             1
      1 : R  <--------- R  : 2
                | y |

o12 : ChainComplexMap
```

We can also set up the same complex in a more compact way, by simply feeding the maps we want in order. Macaulay2 will automatically place the first map with the target in homological degree 0 and the source in degree 1.

```
i13 : D = chainComplex(d0,d1)

       1      1      1
o13 = R  <-- R  <-- R

      0      1      2

o13 : ChainComplex
```

Notice this is indeed the same complex.

```
i14 : D.dd

           1             1
o14 = 0 : R  <--------- R  : 1
                | x |

           1             1
      1 : R  <--------- R  : 2
                | y |

o14 : ChainComplexMap
```

We can also ask Macaulay2 to compute the homology of our complex:

```
i15 : HH D

o15 = 0 : cokernel | x |

      1 : subquotient (| y |, | y |)

      2 : image | x |

o15 : GradedModule
```

Or we could simply ask for the homology in a specific degree:

```
i16 : HH_2(D)

o16 = image | x |

                                    1
o16 : R-module, submodule of R

i17 : HH_3(D)

o17 = 0

o17 : R-module
```

A.3.2 The Complexes package

To use this functionality, you must first load the `Complexes` package.

```
i18 : needsPackage "Complexes"
--loading configuration for package "FourTiTwo" from file /Users/eloisa/Library/Applicat
--loading configuration for package "Topcom" from file /Users/eloisa/Library/Application
```

```
o18 = Complexes

o18 : Package
```

We can use our maps from above to set up a complex with the same maps. We feed a list of the maps we want to use to the method `complex`.

```
i19 : F = complex({d0,d1})

       1      1      1
o19 = R  <-- R  <-- R

      0      1      2

o19 : Complex
```

We cam read off the maps and the homology in our complex using the same commands as we use with `chainComplexes`, although the information returned gets presented in a slightly different fashion.

```
i20 : HH F

o20 = cokernel | x | <-- subquotient (| y |, | y |) <-- image | x |

      0                  1                                2

o20 : Complex

i21 : F.dd

          1             1
o21 = 0 : R  <--------- R  : 1
                | x |

          1             1
      1 : R  <--------- R  : 2
                | y |

o21 : ComplexMap
```

If we want to set up our complex starting in a different homological degree, we can do the following:

```
i22 : G = complex({d0,d1}, Base => 7)
```

```
       1      1      1
o22 = R  <-- R  <-- R

      7      8      9

o22 : Complex

i23 : H = complex({d0,d1}, Base => -13)

       1       1       1
o23 = R   <-- R   <-- R

      -13     -12     -11

o23 : Complex
```

A.3.3 Maps of complexes

Suppose we are given two complexes `C` and `D` and a map of complexes `f` : `C` $\longrightarrow$ `D`. The routine `map` can be used to define f using `chainComplexes`: it receives the target `D`, the source `D`, and a function `f` that returns f_i when we compute `f(i)`.

```
i1 : R = QQ[x,y,z]

o1 = R

o1 : PolynomialRing

i2 : R = QQ[x,y]/ideal"xy";

i3 : mx = map(R^1,R^1,matrix{{x}})

o3 = | x |

             1       1
o3 : Matrix R  <--- R

i4 : my = map(R^1, R^1, matrix{{y}})

o4 = | y |

             1       1
o4 : Matrix R  <--- R

i5 : C = chainComplex(mx,my)
```

```
      1      1      1
o5 = R  <-- R  <-- R

     0      1      2

o5 : ChainComplex

i6 : D = chainComplex(my,mx)

      1      1      1
o6 = R  <-- R  <-- R

     0      1      2

o6 : ChainComplex

i7 : f = map(D,C,i -> if even(i) then my else mx)

          1             1
o7 = 0 : R  <--------- R  : 0
               | y |
          1             1
     1 : R  <--------- R  : 1
               | x |
          1             1
     2 : R  <--------- R  : 2
               | y |

o7 : ChainComplexMap
```

Here's what we can do if we prefer to write a list with the maps in `f`:

```
i8 : f = map(D,C,i -> {my,mx,my}_i)

          1             1
o8 = 0 : R  <--------- R  : 0
               | y |
          1             1
     1 : R  <--------- R  : 1
               | x |
          1             1
     2 : R  <--------- R  : 2
               | y |

o8 : ChainComplexMap
```

If we prefer to do the same with the `Complexes` package, one advantage is that `map` *does* receive `(target, source, list of maps)`.

```
i2 : R = QQ[x,y,z]

o2 = R

o2 : PolynomialRing

i3 : R = QQ[x,y]/ideal"xy";

i4 : mx = map(R^1,R^1,matrix{{x}});

i5 : my = map(R^1, R^1, matrix{{y}});

i6 : C = complex({mx,my});

i7 : D = complex({my,mx});

i8 : f = map(D,C,{my,mx,my})

          1             1
o8 = 0 : R  <--------- R  : 0
               | y |
          1             1
     1 : R  <--------- R  : 1
               | x |
          1             1
     2 : R  <--------- R  : 2
               | y |

o8 : ComplexMap
```

Index

Bibliography

[AM69] Michael F. Atiyah and Ian G. Macdonald. *Introduction to commutative algebra*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969.

[DF04] David S. Dummit and Richard M. Foote. *Abstract algebra*. Wiley, 3rd ed edition, 2004.

[DSGJ20] Alessandro De Stefani, Eloísa Grifo, and Jack Jeffries. A Zariski-Nagata theorem for smooth $\mathbb{Z}$-algebras. *J. Reine Angew. Math.*, 761:123–140, 2020.

[EH79] David Eisenbud and Melvin Hochster. A Nullstellensatz with nilpotents and Zariski's main lemma on holomorphic functions. *J. Algebra*, 58(1):157–161, 1979.

[Eis95] David Eisenbud. *Commutative algebra with a view toward algebraic geometry*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[Eis05] David Eisenbud. *The geometry of syzygies*, volume 229 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005. A second course in commutative algebra and algebraic geometry.

[FMS14] Christopher A Francisco, Jeffrey Mermin, and Jay Schweig. A survey of stanley–reisner theory. In *Connections Between Algebra, Combinatorics, and Geometry*, pages 209–234. Springer, 2014.

[Fre03] Peter J. Freyd. Abelian categories [mr0166240]. *Repr. Theory Appl. Categ.*, (3):1–190, 2003.

[ILL$^+$07] Srikanth B. Iyengar, Graham J. Leuschke, Anton Leykin, Claudia Miller, Ezra Miller, Anurag K. Singh, and Uli Walther. *Twenty-four hours of local cohomology*, volume 87 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2007.

[Kap58] Irving Kaplansky. Projective modules. *Ann. of Math (2)*, 68:372–377, 1958.

[Kun69] Ernst Kunz. Characterizations of regular local rings for characteristic p. *Amer. J. Math.*, 91:772–784, 1969.

[Las05] Emanuel Lasker. Zur theorie der moduln und ideale. *Mathematische Annalen*, 60:20–116, 1905.

[Lyu92] Gennady Lyubeznik. The number of defining equations of affine algebraic sets. *American Journal of Mathematics*, 114(2):413–463, 1992.

[Mat58] Eben Matlis. Injective modules over Noetherian rings. *Pacific J. Math.*, 8:511–528, 1958.

[Mat80] Hideyuki Matsumura. *Commutative algebra*, volume 56 of *Mathematics Lecture Note Series*. Benjamin/Cummings Publishing Co., Inc., Reading, Mass., second edition, 1980.

[Mat89] Hideyuki Matsumura. *Commutative ring theory*, volume 8 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 1989. Translated from the Japanese by M. Reid.

[ML98] Saunders Mac Lane. *Categories for the working mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.

[Nag62] Masayoshi Nagata. *Local rings*. Interscience, 1962.

[NM91] Luis Narváez-Macarro. A note on the behaviour under a ground field extension of quasicoefficient fields. *J. London Math. Soc. (2)*, 43(1):12–22, 1991.

[Noe21] Emmy Noether. Idealtheorie in ringbereichen. *Mathematische Annalen*, 83(1):24–66, 1921.

[Pee11] Irena Peeva. *Graded syzygies*, volume 14 of *Algebra and Applications*. Springer-Verlag London, Ltd., London, 2011.

[Poo19] Bjorn Poonen. Why all rings should have a 1. *Mathematics Magazine*, 92(1):58–62, 2019.

[Rie17] E. Riehl. *Category Theory in Context*. Aurora: Dover Modern Math Originals. Dover Publications, 2017.

[Rot09] Joseph J. Rotman. *An introduction to homological algebra*. Universitext. Springer, New York, second edition, 2009.

[Wei94] Charles A. Weibel. *An introduction to homological algebra*, volume 38 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1994.

[Zar49] Oscar Zariski. A fundamental lemma from the theory of holomorphic functions on an algebraic variety. *Ann. Mat. Pura Appl. (4)*, 29:187–198, 1949.