# Introduction to commutative algebra

Arne Smeets

2022–2023

# Contents

# Foreword

These are notes for an introductory graduate course on commutative and homological algebra taught at KU Leuven during the first term of the academic year 2020–2021. They should be read in conjunction with parts of Rotman's colossal *Advanced modern algebra*. As opposed to Rotman's book, these notes are meant to be reasonably concise; they give a faithful image of the material covered in this course. At various points, there will be references to Rotman using the symbol 📖; of course the material contained in these references should be known as well. Finally, it is always a good idea to read more than strictly needed. There are lots of good sources around, such as Aluffi's *Algebra: Chapter 0*, Lang's *Algebra* and Knapp's *Basic algebra* and *Advanced algebra*, all three of which cover most of the material in this course.

The goal of these notes is also to make you work hard, because working hard and being hungry and eager is the only good way to learn the material in this course. Therefore the notes will be full of *quick questions*: modest intermediate assignments which the attentive reader should be able to answer within a few minutes. The symbol ⚠, on the other hand, indicates that details (or even full proofs) have been left out deliberately. Both types of assignments are meant to encourage active reading.

These notes will inevitably contain typos and mistakes of all sorts. Any suggestion for improvement of these notes is most welcome; at the end of the term, the student who contributes the highest number of suggestions – typos, additional examples, mathematical corrections, clarifying remarks, . . . – will be rewarded with a nice prize: yet another way to encourage active reading!

**Acknowledgements.** Thanks to Dylan Boyd, Yenthe Deschryver, Myrthe D'Haen, Samira le Grand, Jiayi Kang, Maarten Lathouwers, Ben Valkeners, Lieven Van Baelen, Charlotte Van Isveldt, Alexander Van Werde and Thomas Wolfs for providing corrections and comments in the past.

# Chapter 1

# Rings

We start by rapidly reviewing and complementing previous knowledge on the definitions and elementary properties of rings and their ideals. This includes brief discussions of factorisation behaviour, of various types of ideals, and of the elementary operations which can be performed on ideals.

## 1.1 Basic definitions

Let us recall a few basic notions which are hopefully still familiar to the reader.

**Definition 1.1.1.** A ring $(R, +, \cdot)$ is a set $R$ equipped with two binary operations, called addition and multiplication (denoted by $+$ and $\cdot$ respectively), such that the following properties hold:

- $(R, +)$ is an abelian group; in particular, $R$ has a zero element, denoted by $0$, and every $x \in R$ has an additive inverse denoted by $-x$.
- Multiplication is associative, i.e. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in R$, and distributive over addition, i.e. $x \cdot (y + z) = x \cdot y + x \cdot z$, $(y + z) \cdot x = y \cdot x + z \cdot x$ for all $x, y, z \in R$.

The ring is said to be *commutative* if $x \cdot y = y \cdot x$ for all $x, y \in R$, and is said to have an *identity element* (or to be *unital*) if there exists some element $1 \in R$ such that $x \cdot 1 = x = 1 \cdot x$ for all $x \in R$.

*Remark* 1.1.2. Very often the symbol $\cdot$ will be omitted, and we will write $xy$ instead of $x \cdot y$. To simplify notation, we will almost always say "the ring $R$" rather than "the ring $(R, +, \cdot)$". If a ring admits an identity element, this element is necessarily unique. We do not exclude the possibility that the elements $0$ and $1$ of a given ring $R$ coincide; in this case $R$ is necessarily the *zero ring*, with a unique element denoted $0$ (check this). **In this course, "ring" will mean "commutative, unital ring", unless stated otherwise.**

**Definition 1.1.3.** A homomorphism of rings is a function $f$ from a ring $R_1$ to a ring $R_2$ such that

- $f$ respects addition: $f$ is a homomorphism of abelian groups, i.e. $f(x + y) = f(x) + f(y)$ for all $x, y \in R_1$, and hence also $f(0) = 0$ and $f(-x) = -f(x)$ for all $x \in R_1$.
- $f$ respects multiplication: $f(xy) = f(x)f(y)$ for all $x, y \in R_1$.
- $f$ respects the identity: $f(1) = 1$.

A subset $R'$ of a ring $R$ is a *subring* if $R'$ is closed under addition and multiplication and contains the identity element of $R$. The inclusion $R' \hookrightarrow R$ is then a homomorphism of rings. If $f : R_1 \to R_2$ and $g : R_2 \to R_3$ are homomorphisms of rings, then so is the composition $g \circ f : R_1 \to R_3$.

**Definition 1.1.4.** An *ideal* $I$ of a ring $R$ is a subset of $R$ which is an additive subgroup of $R$ with the property that $RI \subseteq I$, i.e. such that $x \in R$ and $y \in I$ together imply $xy \in I$.

The quotient group $R/I$ inherits a uniquely defined multiplication operation from $R$ which makes it into a ring, called the *quotient ring* $R/I$. The elements of this ring are the cosets of $I$ in $R$, and the map $\pi : R \to R/I$ which maps each $x \in R$ to its coset $x + I$ is a surjective homomorphism of rings.

**Proposition 1.1.5.** *Let $R$ be a ring and let $I$ be an ideal of $R$. There is a bijective, inclusion-preserving correspondence between the set of ideals of $R$ which contain $I$, and the set of ideals of $R/I$.*

**Definition 1.1.6.** If $f : R_1 \to R_2$ is a homomorphism of rings, then we call $\ker f = f^{-1}(0)$ the *kernel* of $f$. The *image* of $f$ is $\mathrm{im}(f) = f(R_1)$.

**Proposition 1.1.7.** *In the setting of Definition 1.1.6,* $\ker f$ *is an ideal of $R_1$. Moreover,* $\mathrm{im}\, f = f(R_1)$ *is a subring of $R_2$, and $f$ induces an isomorphism (i.e. bijective homomorphism) of rings*

$$R_1/\ker f \cong \mathrm{im}\, f.$$

**Definition 1.1.8.** A *zero-divisor* in a ring $R$ is an element $x \in R$ which “divides 0”, in the sense that there exists a non-zero element $y \in R$ such that $xy = 0$. A non-zero ring *without* zero-divisors different from 0 (i.e., a ring in which the product of two non-zero elements is never 0) is called an *integral domain*.

**Definition 1.1.9.** A *nilpotent element* in a ring $R$ is an element $x \in R$ with the property that $x^n = 0$ for some positive integer $n$. A *unit* in $R$ is an element $x \in R$ which “divides 1”, in the sense that there exists $y \in R$ such that $xy = 1$. Such an element $y$ is uniquely determined by $x$ (why?), and is denoted by $x^{-1}$.

*Quick question* 1.1.10. Check that a (non-zero) nilpotent element is a zero-divisor, but not conversely in general. Check that the set of nilpotent elements is closed under multiplication, and that the set of units even becomes an abelian group under multiplication, denoted by $R^\times$. Finally, check that the sum of a unit and a nilpotent element is again a unit (this is slightly harder).

In a ring $R$, the set of multiples $xy$ of a given element $x \in R$ form a so-called *principal ideal*, denoted by $(x)$; it is not hard to see that $x$ is a unit if and only if $(x) = R$. A *field* is a ring in which $0 \neq 1$ and every non-zero element is a unit (in particular, a field is an integral domain).

**Proposition 1.1.11.** *Let $R$ be a ring different from the zero ring* $0$. *Then the following are equivalent:*

*(1) $R$ is a field;*

*(2) $R$ has only two ideals, namely* $0$ *and $R$ itself;*

*(3) if $f : R \to R'$ is a homomorphism of rings, then either $f$ is injective, or the zero map.*
*(Note that the latter case forces $R' = 0$ as well, because of the last condition in Definition 1.1.3.)*

**Definition 1.1.12.** Let $R$ be a ring. A *prime ideal* in $R$ is an ideal $\mathfrak{p} \neq R$ with the property that if $xy \in \mathfrak{p}$ for some elements $x, y \in R$, then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. A *maximal ideal* in $R$ is an ideal $\mathfrak{m} \neq R$ with the property that there does not exist an ideal $I$ such that $\mathfrak{m} \subsetneq I \subsetneq R$.

**Proposition 1.1.13.** *Let $R$ be a ring. An ideal $\mathfrak{p}$ in $R$ is prime if and only if $R/\mathfrak{p}$ is an integral domain; in particular, the zero ideal is prime if and only if $R$ is an integral domain. An ideal $\mathfrak{m}$ in $R$ is maximal if and only if $R/\mathfrak{m}$ is a field; in particular, maximal ideals are prime.*

*Quick question* 1.1.14. Show that the inverse image of a prime ideal under a homomorphism of rings is again a prime ideal. Show (using an example) that this statement fails if one replaces "prime" by "maximal": all we can say for sure is that the inverse image of a maximal ideal is prime.

The following theorem yields the existence of a sufficient supply of maximal ideals:

**Theorem 1.1.15.** *Every non-zero ring has at least one maximal ideal.*

*Proof.* We use Zorn's lemma, see [Rotman, §6.4] (we refer to [Rotman, Appendix] for background). If $R$ is a non-zero ring, denote by $\Sigma$ the set of proper ideals (i.e. ideals not equal to $R$), ordered by inclusion. Then $\Sigma$ is non-empty since $0 \in \Sigma$. Moreover, every chain in $\Sigma$ has an upper bound in $\Sigma$: if $(I_\alpha)_{\alpha\in A}$ is a chain, then one easily checks that $\bigcup_{\alpha\in A} I_\alpha$ is such a bound in $\Sigma$. Therefore Zorn's lemma yields the existence of a maximal element, as desired. (See [Rotman, Theorem 6.46] for more details.) □

*Quick question* 1.1.16. In fact, every proper ideal $I$ of a ring $R$ is contained in a maximal ideal; in particular, every non-unit is contained in a maximal ideal. Check this in two different ways: first by applying Theorem 1.1.15 directly to $R/I$, then by (slightly) modifying the proof of Theorem 1.1.15.

**Definition 1.1.17.** A *local ring* is a ring with exactly one maximal ideal. If $R$ is a local ring with maximal ideal $\mathfrak{m}$ then $\kappa = R/\mathfrak{m}$ is called the *residue field* of $R$.

**Proposition 1.1.18.** *If $R$ is a ring and if $\mathfrak{m}$ is a proper ideal such that $R \setminus \mathfrak{m}$ consists entirely of units, then $R$ is local and $\mathfrak{m}$ is its maximal ideal. Moreover, if $R$ is a ring and if $\mathfrak{m}$ is a maximal ideal such that $1 + \mathfrak{m} = \{1 + x : x \in \mathfrak{m}\}$ consists entirely of units, then the same conclusion holds.*

*Proof.* The first part follows from the fact that every proper ideal consists entirely of non-units, and hence must be contained in $\mathfrak{m}$; therefore $\mathfrak{m}$ is the unique maximal ideal. For the second part, observe that if $x \notin \mathfrak{m}$, then $(\mathfrak{m}, x) = R$. Hence there exist $y \in R$ and $z \in \mathfrak{m}$ such that $xy + z = 1$; therefore $xy = 1 - z \in 1 + \mathfrak{m}$ is a unit, and so is $x$. The result now follows from the first statement. □

## 1.2 Operations on ideals

We will now study a number of ways to build new ideals in a ring starting from old ones.

A very classical operation is the *sum* of two ideals $I$ and $J$ in a ring $R$: this is defined to be the set $I + J = \{x + y : x \in I, y \in J\}$ of all sums of elements of $I$ and $J$. One easily checks that this is indeed an ideal: the smallest ideal which contains both $I$ and $J$. More generally, given a (possibly infinite) family $(I_\alpha)_{\alpha\in A}$ of ideals, one defines the sum $\sum_{\alpha\in A} I_\alpha$ as the set of sums $\sum_{\alpha\in A} x_\alpha$, where $x_\alpha \in I_\alpha$ for all $\alpha \in A$ and *where almost all $x_\alpha$ are equal to* 0*, i.e. $x_\alpha \neq 0$ for at most finitely many $\alpha \in A$.* As before, this is an ideal, which can be characterised as the smallest ideal of $R$ containing all $I_\alpha$ for $\alpha \in A$.

A second classical operation is the *intersection* of an arbitrary family of ideals $(I_\alpha)_{\alpha\in A}$ in a ring $R$: it is entirely trivial that this is again an ideal.

A third construction is the *product* of two ideals $I$ and $J$ in a ring $R$: this is the ideal (denoted by $IJ$) generated by all products $xy$, with $x \in I$ and $y \in J$. The elements of this ideal are exactly the finite sums of the form $\sum_{k=1}^{n} x_k y_k$ where $x_k \in I$ and $y_k \in J$ for $k = 1, \cdots, n$.

*Quick question* 1.2.1. Check that $IJ \subseteq I \cap J$, and that this inclusion is not an equality in general.

Similarly, one defines the product of any *finite* family of ideals; in particular, one can define the $n$-th power $I^n$ of an ideal $I$ for any positive integer $n$. This is precisely the ideal generated by all products of the form $x_1 \cdots x_n$, where $x_1, \cdots, x_n \in I$.

*Quick question* 1.2.2. The ring **Z** is a principal ideal domain: every ideal is generated by a single integer. If $I = (m)$ and $J = (n)$ are two ideals, describe $I + J$, $IJ$ and $I \cap J$ in terms of $m$ and $n$.

*Quick question* 1.2.3. Let $k$ be a field, let $R = k[X_1, \cdots, X_n]$ and let $\mathfrak{m} = (X_1, \cdots, X_n)$. Check that $\mathfrak{m}$ is a maximal ideal, and give an explicit description for the various powers of $\mathfrak{m}$.

The operations defined so far are both commutative and associative. Moreover, there is a distributive law in the sense that $I_1(I_2 + I_3) = I_1I_2 + I_1I_3$ holds for arbitrary ideals $I_1$, $I_2$ and $I_3$ in a ring $R$.

*Quick question* 1.2.4. Do we also have $I_1 \cap (I_2 + I_3) = I_1 \cap I_2 + I_1 \cap I_3$ for arbitrary $I_1$, $I_2$ and $I_3$?

**Definition 1.2.5.** Two ideals $I$ and $J$ in a ring $R$ are said to be *coprime* if $I + J = R$.

Given ideals $I$ and $J$ in a ring $R$, we always have $IJ \subseteq I \cap J$; even though $IJ \neq I \cap J$ in general (Question 1.2.1), the equality $IJ = I \cap J$ *does* hold provided that $I$ and $J$ are coprime, since in that case

$$I \cap J = (I + J)(I \cap J) = I(I \cap J) + J(I \cap J) \subseteq IJ.$$

If $R_1, , \cdots , R_n$ are rings, then the *direct product* $R = \prod_{i=1}^n R_i$ is defined to be the set of all $n$-tuples of the form $(x_1, \cdots, x_n)$, where $x_i \in R_i$ for $i = 1, \cdots, n$, equipped with componentwise addition and multiplication. This is a commutative ring with identity element $(1, \cdots, 1)$. The *projection on the $i$-th component* is the homomorphism of rings $\pi_i : R \to R_i$ given by $(x_1, \cdots, x_n) \mapsto x_i$.

**Lemma 1.2.6.** *Let $I_1, \cdots, I_n$ be pairwise coprime ideals in a ring $R$. Then $I_1 \cdots I_n = \bigcap_{j=1}^n I_j$.*

*Proof.* The case $n = 2$ has already been dealt with above. The proof for general $n$ proceeds by induction, and is left to the reader. □

We are now ready to state the "Chinese remainder theorem" for rings.

**Theorem 1.2.7** (Chinese remainder theorem)**.** *Let $R$ be a ring and let $I_1, \cdots, I_n$ be ideals in $R$. Then*

$$\varphi : R \to \prod_{j=1}^n R/I_j : x \mapsto (x + I_1, \cdots, x + I_n)$$

*is a homomorphism of rings. Moreover $\varphi$ is surjective if and only if the ideals $I_1, \cdots, I_n$ are pairwise coprime, and* $\ker \varphi = I_1 \cap \cdots \cap I_n$ *(so that $\varphi$ is injective if and only if $\bigcap_{j=1}^n I_j = 0$).*

(Convince yourself that this indeed generalises the classical Chinese remainder theorem!)

*Proof.* Since the first and last statements are trivial, we focus on the surjectivity of $\varphi$.

Assume that $\varphi$ is surjective. Let us check that $I_1$ and $I_2$ are coprime – the same argument then works for arbitrary pairs of indices. By assumption, there exists $x \in R$ such that $\varphi(x) = (1, 0, \cdots, 0)$; hence $x \equiv 1 \pmod{I_1}$ and $x \equiv 0 \pmod{I_2}$, so that $1 = (1 - x) + x \in I_1 + I_2$, as desired.

Conversely, assume that $I_1, \cdots, I_n$ are pairwise coprime. To show that $\varphi$ is surjective, it suffices to show that $(1, 0, \cdots, 0) \in \operatorname{im} \varphi$ (why?). For all $j \geq 2$, the fact that $I_1 + I_j = R$ implies the existence of $x_j \in I_1$ and $y_j \in I_j$ such that $x_j + y_j = 1$. Take $y = y_2 \cdots y_n$, then $y = (1 - x_2) \cdots (1 - x_n) \equiv 1 \pmod{I_1}$. Since moreover $y \equiv 0 \pmod{I_j}$ for all $j \geq 2$, we have $\varphi(y) = (1, 0, \cdots, 0)$, as desired. □

Let us continue with some (presumably new) operations on ideals.

**Definition 1.2.8.** Given two ideals $I$ and $J$ in a ring $R$, the *ideal quotient* is $(I : J) = \{x \in R : xJ \subseteq I\}$. In particular, the quotient $(0 : J)$ is called the *annihilator* of $J$, denoted by $\operatorname{Ann} J$.

It is clear that the ideal quotient of two ideals is again an ideal. If $J = (x)$ is a principal ideal, we write $(I : x)$ rather than $(I : (x))$, and $\operatorname{Ann} x$ rather than $\operatorname{Ann}(x)$, to simplify notation.

**Proposition 1.2.9.** *We have the following equalities for arbitrary ideals in a ring $R$:*

$$I \subseteq (I : J),\ (I : J)J \subseteq I,\ ((I : J) : K) = (I : JK) = ((I : K) : J).$$

*For arbitrary families, we have $\left(\bigcap_{\alpha\in A} I_\alpha : J\right) = \bigcap_{\alpha\in A}(I_\alpha : J)$ and $\left(I : \sum_{\beta\in B} J_\beta\right) = \bigcap_{\beta\in B}(I : J_\beta)$.*

*Quick question* 1.2.10. Let $I = (m)$ and $J = (n)$ be two ideals in $\mathbf{Z}$. Describe $(I : J)$.

The final operation on ideals which we introduce here is the construction of the radical:

**Definition 1.2.11.** Let $R$ be a ring. If $I$ is an ideal, the *radical* of $I$ is the set

$$\sqrt{I} = \{x \in R : x^n \in I \text{ for some } n > 0\}.$$

A *radical ideal* is an ideal which is equal to its own radical.

*Quick question* 1.2.12. What are the radical ideals in $\mathbf{Z}$? Describe the radical of an arbitrary ideal in $\mathbf{Z}$.

*Quick question* 1.2.13. Let $\mathfrak{p}$ be a prime ideal in a ring $R$, and let $n$ be a positive integer. Describe $\sqrt{\mathfrak{p}^n}$.

**Proposition 1.2.14.** *If $R$ is a ring and if $I$ is an arbitrary ideal, then $\sqrt{I}$ is a radical ideal.*

*Proof.* If $x \in \sqrt{I}$ and $r \in R$, then $x^m \in I$ for some $m > 0$; it follows that $(rx)^m \in I$ and therefore $rx \in \sqrt{I}$. If $x, y \in \sqrt{I}$, then there exist $m, n > 0$ such that $x^m \in I$ and $y^n \in I$. The binomial theorem (valid in any commutative ring!) now implies that $(x + y)^{m+n-1} \in I$, and therefore $x + y \in \sqrt{I}$. This proves that $\sqrt{I}$ is an ideal; the proof of the fact that $\sqrt{I}$ is radical is left to the reader. □

*Quick question* 1.2.15. Let $I$ be an ideal in a ring $R$. Show that $I$ is radical if and only if $R/I$ does not have any non-zero nilpotent elements.

**Definition 1.2.16.** Let $R$ be a ring. The *nilradical* of $R$ is the ideal $\mathfrak{n}_R = \sqrt{0}$.

In other words, the nilradical $\mathfrak{n}_R$ of $R$ is nothing but the set of nilpotents of $R$; by Proposition 1.2.14, this is an ideal of $R$, and it follows from Question 1.2.15 that $R/\mathfrak{n}_R$ does not have any non-zero nilpotents.

**Proposition 1.2.17.** *Let $I$ be an ideal in a ring $R$. The radical $\sqrt{I}$ is the intersection of all prime ideals of $R$ which contain $I$. In particular, the nilradical $\mathfrak{n}_R$ is the intersection of all prime ideals of $R$.*

*Proof.* Let $J$ be the intersection of all prime ideals of $R$ which contain $I$. It is not hard to see that $\sqrt{I} \subseteq J$: indeed, if $x \in R$ has the property that $x^n \in I$ for some $n > 0$, and if $\mathfrak{p}$ is a prime ideal which contains $I$, then $x^n \in \mathfrak{p}$ and therefore $x \in \mathfrak{p}$. Since this holds for all primes $\mathfrak{p}$ containing $I$, we obtain that $x \in J$.

Conversely, assume that $x \notin \sqrt{I}$; we wish to prove that $x \notin J$. Let $\Sigma$ be the set of ideals $K$ of $R$ which contain $I$ and which have the property that $x^n \notin K$ for all $n > 0$. Like in the proof of Theorem 1.1.15, Zorn's lemma can be applied to $\Sigma$ (ordered by inclusion), and therefore $\Sigma$ has a maximal element $\mathfrak{p}$. We claim that $\mathfrak{p}$ is prime; since $x \notin \mathfrak{p}$ by construction of $\Sigma$, this claim clearly implies the desired result.

To prove the claim, let $y, z \notin \mathfrak{p}$; we want to show that $yz \notin \mathfrak{p}$. The ideals $\mathfrak{p} + (y)$ and $\mathfrak{p} + (z)$ strictly contain $\mathfrak{p}$ and therefore do not belong to $\Sigma$. Hence there exist $m, n > 0$ such that $x^m \in \mathfrak{p} + (y)$ and $x^n \in \mathfrak{p} + (z)$. It follows that $x^{m+n} \in \mathfrak{p} + (yz)$, whence $\mathfrak{p} + (yz) \notin \Sigma$ and $yz \notin \mathfrak{p}$, as desired. □

**Definition 1.2.18.** The *Jacobson radical* $\mathfrak{j}_R$ of a ring $R$ is the intersection of all maximal ideals in $R$.

It follows immediately from Proposition 1.2.17 that the Jacobson radical of an arbitrary ring contains the nilradical. We have the following alternative characterisation for the Jacobson radical:

**Proposition 1.2.19.** *Let $R$ be a ring. If $x \in R$, then $x \in \mathfrak{j}_R$ if and only if $1 + xy$ is a unit for all $y \in R$.*

## 1.3 Factorisation

We end this chapter with a discussion of *factorisation properties*. Recall the following notion:

**Definition 1.3.1.** Let $x$ and $y$ be elements of a commutative ring $R$. Then $x$ *divides* $y$ if there exists $a \in R$ such that $ax = y$; this is denoted by $x \mid y$. Moreover, $x$ and $y$ are *associates* if $x \mid y$ and $y \mid x$.

It is clear that if $x$ and $y$ are associates in an integral domain $R$, then they differ (multiplicatively) by units, i.e. there exists $u \in R^\times$ such that $ux = y$. Statements about divisibility can be translated into statements about ideals: $x \mid y$ if and only if $(y) \subseteq (x)$, and $x$ and $y$ are associates if and only if $(x) = (y)$.

**Definition 1.3.2.** Let $R$ be a ring. A non-zero element $x \in R$ is said to be *irreducible* if $x$ is not itself a unit and whenever $x = yz$ for some $y, z \in R$, then one of $y$ and $z$ must be a unit. Moreover, a non-zero element $x \in R$ is said to be *prime* if $x$ is not a unit and whenever $x$ divides $yz$ for some $y, z \in R$, then $x$ divides $y$ or $x$ divides $z$; in other words, $x$ is prime if and only if $(x)$ is a non-zero prime ideal.

*Quick question* 1.3.3. Show that if $R$ is an integral domain, then any prime element in $R$ is irreducible. (This result no longer holds for arbitrary rings $R$ – can you give an example which illustrates this?)

The following notion is very classical and has appeared already in these notes.

**Definition 1.3.4.** A *principal ideal domain* (or PID in short) is an integral domain in which every ideal is principal, i.e., generated by a single element.

More generally, a principal ideal ring is a ring in which all ideals are principal.

*Quick question* 1.3.5. Let $R$ be a PID. If $p \in R$ is irreducible, show that $(p)$ is maximal, and $p$ is prime.

A prominent class of PID's is the class of Euclidean domains:

**Definition 1.3.6.** A *Euclidean domain* is an integral domain $R$ which can be equipped with a so-called *degree function* $\varphi : R \setminus \{0\} \to \mathbf{Z}_{\geq 0}$ satisfying the following conditions:

- if $x, y \in R$ are non-zero and $x \mid y$, then $\varphi(x) \leq \varphi(y)$;
- if $x, y \in R$ are non-zero, there exist $q, r \in R$ such that $y = qx + r$ and either $r = 0$ or $\varphi(r) < \varphi(x)$.

The second condition should be thought of as a generalisation of "division with remainder" in $\mathbf{Z}$; indeed, the ring of integers $\mathbf{Z}$ is a Euclidean domain with degree function $\varphi$ given by $\varphi(m) = |m|$. If $k$ is a field, then $k[X]$ is a euclidean domain with degree function the usual degree of a polynomial.

*Quick question* 1.3.7. Check that if $\varphi$ is identically $0$ in Definition 1.3.6, then $R$ must be a field.

**Example 1.3.8.** The ring $\mathbf{Z}[i]$ of *Gaussian integers* is a Euclidean domain with degree function $\varphi$ given by $\varphi(a + bi) = a^2 + b^2$; see [Rotman, Example 3.59.(iii)] for a detailed argument.

**Theorem 1.3.9.** *A Euclidean domain is a PID.*

*Proof.* Let $R$ be a Euclidean domain, and let $I$ be a non-zero ideal. Choose $x \in I$ non-zero such that $\varphi(x)$ is minimal among all images of non-zero elements of $I$ under $\varphi$. We claim that $I = (x)$. Indeed, let $y \in I$ be non-zero, and choose $q, r \in R$ as in the second condition in Definition 1.3.6. Clearly $r = y - qx \in I$; since $\varphi(x)$ was chosen minimal, this forces $r = 0$. Therefore $x \mid y$, as desired. □

*Remark* 1.3.10. One can show that $\mathbf{Z}\left[\frac{1}{2}(1 + \sqrt{-19})\right]$ is a PID, but not a Euclidean domain.

We continue with another very classical notion:

**Definition 1.3.11.** A *unique factorisation domain* (or UFD in short) is an integral domain $R$ which has the following property. Let $x \in R$ be a non-zero element which is not a unit. Then $x$ can be written as a product of irreducible elements of $R$. Moreover, given two factorisations

$$x = up_1 \cdots p_r = vq_1 \cdots q_s$$

where $u$ and $v$ are units and $p_1, \cdots, p_r, q_1, \cdots, q_s$ are irreducible elements, then $r = s$, and there is a permutation $\sigma : \{1, \cdots, r\} \to \{1, \cdots, s\}$ such that $p_i$ and $q_{\sigma(i)}$ are associates, for $i = 1, \cdots, r$.

This definition implies in particular that the notions of prime and irreducible elements coincide in a UFD (check this). The following result is covered in any undergraduate course on abstract algebra:

**Theorem 1.3.12.** *Any PID is a UFD.*

*Proof.* The first part of the statement (existence of factorisations) is proven using *ascending chains*, see [Rotman, Lemma 6.18] for details. The second part (uniqueness) can be shown by adapting the proof of the "fundamental theorem of arithmetic", see [Rotman, Proposition 6.17, Theorem 6.19]. □

Using the unique factorisation property, one can show that finitely many elements of a UFD admit a *greatest common divisor* (gcd) and a *least common multiple* (lcm); see [Rotman, Proposition 6.20] for the existence of the gcd (a similar argument works for the lcm). It is important to note that the gcd and lcm are only well-defined up to multiplication by a unit.

On the other hand, Bézout's theorem fails for general UFD's: it is not true in general that the gcd of a number of elements of a UFD can be written as a linear combination of these elements. For example, the polynomial ring $\mathbf{C}[X, Y]$ happens to be a UFD, the elements $X$ and $Y$ are coprime (in other words, $\gcd(X, Y) = 1$), but clearly $1 \notin (X, Y)$. However, the natural generalisation of Bézout's theorem does hold in an arbitrary PID (see [Rotman, Theorem 3.57]).

Let us give an important example of a (non-pathological!) ring in which unique factorisation fails.

**Example 1.3.13.** Let $R = \mathbf{Z}[\sqrt{-5}]$. Given an element $x = a + b\sqrt{-5} \in R$, denote $\varphi(x) = a^2 + 5b^2$. Then $\varphi$ is multiplicative, i.e., $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in R$. Consider the equalities

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

One can check that 2, 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are all irreducible, using the function $\varphi$. For example, assume that $2 = xy$ for certain elements $x, y \in R$. Writing $x = a + b\sqrt{-5}$ and $y = c + d\sqrt{-5}$, we see that $4 = \varphi(2) = \varphi(x)\varphi(y) = (a^2 + 5b^2)(c^2 + 5d^2)$ in $\mathbf{Z}$. Since the factors in the right hand side cannot be equal to 2 (why?), we see that one of them must be equal to 1. Hence one of $x$ and $y$ must be a unit, proving that 2 is irreducible. A similar argument shows that 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are all irreducible.

On the other hand, it is not hard to see that 2 and 3 are not associates of $1 \pm \sqrt{-5}$, which means that the second condition in Definition 1.3.11 is not satisfied, in other words, that $R$ is not a UFD.

To end this chapter, we will now work our way towards a proof of the following fundamental result. For a slightly different (and more detailed) treatment, we refer to [Rotman, Lemma 6.23–Theorem 6.25]

**Theorem 1.3.14.** *If $R$ is a UFD, then so is $R[X]$.*

To prove this, we will need the following notions.

**Definition 1.3.15.** Let $R$ be a UFD, and let $f \in R[X]$ be non-zero. The *content* of $f$, denoted by $c_f$, is the gcd of the coefficients of $f$. A polynomial $f \in R[X]$ is *primitive* if $c_f$ is a unit. Given $f \in R[X]$, we denote by $\tilde{f} \in R[X]$ a polynomial such that $f = c_f \tilde{f}$; this is the *primitive polynomial associated to $f$*.

It is clear that $c_f$ and $\tilde{f}$ are well-defined up to multiplication by a unit.

**Lemma 1.3.16** (Gauss' lemma)**.** *Let $R$ be a UFD. If $f, g \in R[X]$ are primitive, then so is $fg$.*

*Proof.* Assume that $fg$ is *not* primitive. Let $p \in R$ be a prime element which divides $c_{fg}$, then $fg \equiv 0 \pmod{p}$ in $R[X]$, and therefore the image of $fg$ in $(R/p)[X]$ vanishes. Since $R/p$ is an integral domain, so is $(R/p)[X]$; this implies that $f \equiv 0 \pmod{p}$ or $g \equiv 0 \pmod{p}$, in other words: $p$ divides $c_f$ or $c_g$. This contradicts our assumption that $f$ and $g$ both be primitive. □

The previous lemma has the following useful application.

**Lemma 1.3.17.** *Let $R$ be a UFD, with field of fractions $F$. Let $f \in R[X]$ be a non-constant polynomial. Then $f$ is irreducible in $R[x]$ if and only if $f$ is both primitive in $R[X]$ and irreducible in $F[X]$.*

*Proof.* We sketch the proof and leave the details to the reader. Since the "if" part is trivial, we focus on the "only if". Let $f \in R[X]$ be irreducible, then $f$ is certainly primitive. Assume that $f$ is reducible in $F[X]$, i.e., $f = gh$ where $g, h \in F[x]$ are non-constant. There exist $c, d \in F$ such that $g = c\tilde{g}$ and $h = d\tilde{h}$ where $\tilde{g}, \tilde{h} \in R[X]$ are primitive. Writing $cd = a/b$ for coprime $a, b \in R$, we get $bf = a\tilde{g}\tilde{h}$. Since $f$ and $\tilde{g}\tilde{h}$ are primitive, we have $a/b \in R^\times$. Now $f = (a/b)\tilde{g}\tilde{h}$ shows that $f$ is reducible in $R[X]$. □

*Quick question* 1.3.18. Let $R = \mathbf{Z}[\sqrt{5}]$, with fraction field $F = \mathbf{Q}(\sqrt{5})$. Check that $x^2 - x - 1$ is irreducible in $R[x]$, but reducible in $F[x]$; deduce that $R$ is not a UFD. (Compare with Example 1.3.13.)

*Proof of Theorem 1.3.14.* We need to check the two conditions in Definition 1.3.11.

For the first one, we proceed by induction on the degree of $f$, the base case of "constant" (degree 0) polynomials being trivial. Let $f \in R[X]$ be non-constant, and write $f = c_f\tilde{f}$. Since $R$ is a UFD, $c_f$ is either a unit or a product of irreducible elements of $R$ (hence also of $R[X]$). If $\tilde{f}$ is irreducible in $R[X]$, there is nothing to prove; otherwise, $\tilde{f} = gh$ for certain $g, h \in R[X]$ which are primitive non-units by Lemma 1.3.16 (and therefore non-constant). Since both $g$ and $h$ have smaller degree than $f$, the inductive hypothesis now says that $g$ and $h$ both factor as a product of irreducible elements in $R[X]$. Therefore the same is true for $f = c_f gh$, which proves that the first condition is satisfied.

For the second condition (uniqueness), observe that any factorisation of $f \in R[X]$ into irreducible elements is obtained from such factorisations of $c_f$ and $\tilde{f}$, where the latter one is a factorisation into *primitive* irreducible elements of $R[X]$. Since $R$ is a UFD, it suffices to prove that the factorisation of a primitive element of $R[X]$ into primitive irreducible elements is unique.

By Lemma 1.3.17, irreducible elements of $R[X]$ remain irreducible in $F[X]$, where $F$ denotes the field of fractions of $R$. Therefore any representation of a primitive element $f \in R[X]$ as a product of primitive irreducible elements is a representation of $f$ as a product of irreducible elements of $F[X]$. Since $F[X]$ is a unique factorisation domain, it is now sufficient to show that if two primitive polynomials $g, h \in R[x]$ are associates in $F[X]$, then $g$ and $h$ are associates in $R[X]$ as well. To prove this, observe that there exist $a, b \in R$ such that $ag = bh$. Since $b$ divides $c_{ag} = ac_g$, and $c_g$ is a unit, it follows that $b$ divides $a$. Similarly, $a$ divides $b$, which means that $a$ and $b$ are associates in $R$. It follows that $u = a/b$ is a unit such that $h = ug$, in other words: $g$ and $h$ are associates in $R[X]$, as desired. □

**Corollary 1.3.19.** *If $R$ is a UFD, then $R[X_1, \cdots, X_n]$ is a UFD for arbitrary $n \geq 1$.*

## Exercises

*Easy exercise* 1.1. Let $R$ be the subring of the PID $\mathbf{Q}[X]$ which consists of all polynomials without linear term. Show that $X^2$ is irreducible in $R$, but not prime; deduce from this that $R$ is *not* a PID.

*Easy exercise* 1.2. Let $R$ be a ring. A polynomial $f \in R[X]$ is said to be *very primitive* if for every prime ideal $\mathfrak{p}$ of $R$, we have $f \notin \mathfrak{p}[X]$ (the ideal of $R[X]$ generated by $\mathfrak{p}$). Prove that if $g, h \in R[X]$, then $gh$ is very primitive if and only if both $g$ and $h$ are very primitive. (Compare with Lemma 1.3.16.)

*Easy exercise* 1.3. A *gcd domain* is an integral domain in which every two elements have a greatest common divisor. Prove that in a gcd domain, every irreducible element is prime.

*Easy exercise* 1.4. A *Bézout domain* is an integral domain in which every two elements have a greatest common divisor that is a linear combination of these elements. Prove that an integral domain $R$ is a Bézout domain if and only if every ideal of $R$ generated by finitely many elements is principal.

*Exercise* 1.5. Let $R$ be a ring and assume that for each $x \in R$, there exists an integer $n \geq 2$, which may depend on $x$, such that $x^n = x$. Prove that all prime ideals in $R$ are maximal.

*Exercise* 1.6. Let $k$ be a field. Consider the ring $k[\![t]\!] = \left\{\sum_{i=0}^{\infty} a_i t^i \mid a_i \in k\right\}$ of formal power series. Define $v : k[\![t]\!] \setminus \{0\} \to \mathbf{Z}_{\geq 0}$ by

$$v\left(\sum_{i=0}^{\infty} a_i t^i\right) = \min\{i : a_i \neq 0\}.$$

(a) Check that if $f, g \in k[\![t]\!] \setminus \{0\}$ then $v(fg) = v(f) + v(g)$. Prove furthermore that if $f + g \neq 0$, then $v(f+g) \geq \min\{v(f), v(g)\}$. (The map $v$ is called a *discrete valuation*.)

(b) Show that $k[\![t]\!]$ is a PID and describe its ideals. How many prime ideals does $k[\![t]\!]$ have?

*Exercise* 1.7. Let $R$ be an integral domain with field of fractions $K$, such that $x \in R$ or $x^{-1} \in R$ for all $x \in K^{\times}$. Show that $R$ is a local ring. (The ring $R$ is said to be a *valuation ring* of $K$.)

*Exercise* 1.8. Let $R$ be a ring which is not a principal ideal ring. Show that $R$ has at least 6 ideals. Show also that $\mathbf{F}_2[X,Y]/(X^2, XY, Y^2)$ has exactly 6 ideals, and is not a principal ideal ring.

*Exercise* 1.9. Consider a potentially non-commutative unital ring $R$ such that $x^2 = x$ holds for all $x \in R$. (Such a ring is called a *Boolean ring*.)

(a) Show that $2x = 0$ for all $x \in R$, and that $R$ is commutative. Determine the nilradical of $R$.

(b) Let $\mathfrak{p}$ be a prime ideal of $R$. Show that $\mathfrak{p}$ is maximal and determine $R/\mathfrak{p}$.

(c) If $R$ is local, show that $R \cong \mathbf{F}_2$.

(d) Let $\mathfrak{m}$ be an ideal of $R$. Show that $\mathfrak{m}$ is maximal iff for each $x \in R$, either $x \in \mathfrak{m}$ or $1 - x \in \mathfrak{m}$.

(e) Show that every ideal of $R$ generated by finitely many elements is a principal ideal. Furthermore, give an example of such a ring in which not every ideal is principal. (This is slightly harder.)

*Exercise* 1.10. Let $R$ be a ring. Let $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$. Prove that $f$ is nilpotent if and only if $a_0, a_1, \cdots, a_n$ are all nilpotent. Prove that $f$ is a unit if and only if $a_0$ is a unit and $a_1, \cdots, a_n$ are nilpotent. Finally, prove that the Jacobson radical of $R[X]$ is equal to the nilradical.

*Hard exercise* 1.11. Let $R$ be a ring. Consider the ring of formal power series $R[\![X]\!]$ over $R$. Say as much as you can about the nilpotent elements and units of $R[\![X]\!]$, and describe the Jacobson radical.

# Chapter 2

# Modules

This chapter introduces the notion of *modules* over arbitrary rings, a wide generalisation of the notion of vector space over a field. After having introduced the basic notions, we will discuss ways to describe modules using *generators* and to construct new modules from old ones. Next, we will see that endomorphisms of finitely generated modules satisfy a Cayley-Hamilton type theorem, just like classical vector spaces. Finally, we will take a look at commutative diagrams, exact sequences and "diagram chasing".

## 2.1 Basic definitions

**Definition 2.1.1.** Let $R$ be a ring. An $R$-*module* is an abelian group $(M, +)$ equipped with a binary operation $R \times M \to M : (r, x) \mapsto r \cdot x$ such that for all $r, s \in R$ and $x, y \in M$, we have

$$r \cdot (x + y) = r \cdot x + r \cdot y, \ (r + s) \cdot x = r \cdot x + s \cdot x, \ (rs) \cdot x = r \cdot (s \cdot x), \ 1 \cdot x = x.$$

The operation $\cdot$ is sometimes called *scalar multiplication*; in practice, one often writes $rx$ rather than $r \cdot x$.

*Remark* 2.1.2. If $R$ is a non-commutative ring, then we have in fact two competing notions of $R$-modules: *left* and *right* $R$-modules, with scalar multiplication written "on the left" and "on the right" respectively. For a commutative ring $R$, these two notions are essentially the same and yield the definition given above; see [Rotman, pp. 525–526] (and also [Rotman, p. 524] for the definition of left and right ideals).

*Quick question* 2.1.3. Check that Definition 2.1.1 implies the equalities

$$r0 = 0, \ 0x = 0 \ \text{ and } \ (-r)x = r(-x) = -rx$$

for all $r \in R$ and $x \in M$.

*Remark* 2.1.4. Let us give an alternative description of $R$-modules. If $A$ is an abelian group, then the set $\operatorname{End}(A)$ of *endomorphisms* of $A$ can be turned into a (typically non-commutative) ring. Indeed, one can define addition and multiplication of $f, g \in \operatorname{End}(A)$ pointwise, as follows:

$$(f + g)(a) = f(a) + g(a), \ (fg)(a) = f(g(a)).$$

(What are the zero and identity elements of this ring?) An $R$-module can now be defined as a pair $(M, \varphi)$ consisting of an abelian group $M$ and a homomorphism of rings $\varphi : R \to \operatorname{End}(M)$. We leave it to the reader to verify in detail that both definitions are indeed equivalent.

**Example 2.1.5.** Let $R$ be a ring. The *zero module*, denoted by $0$, is the trivial group $(0, +)$ equipped with the obvious scalar multiplication, given by $r \cdot 0 = 0$ for all $r \in R$.

**Example 2.1.6.** Let $R$ be a ring and let $n \geq 1$ be an integer. Then $M = R^n$ is an $R$-module with addition defined componentwise, and scalar multiplication given by $r \cdot (x_1, \cdots, x_n) = (rx_1, \cdots, rx_n)$.

**Example 2.1.7.** Given an abelian group $A$, there exists a unique homomorphism of rings $\mathbf{Z} \to \mathrm{End}(A)$ (see Remark 2.1.4). It follows that a **Z**-module is nothing but an abelian group (check the details!).

**Example 2.1.8.** If $k$ is a field, then a $k$-module is simply a $k$-vector space.

**Example 2.1.9.** Let $I$ be an ideal in a ring $R$. Then $I$ is an $R$-module. More generally, if $R$ is a non-commutative ring and if $I$ is a left (resp. right) ideal of $R$, then $I$ is a left (resp. right) $R$-module.

**Example 2.1.10.** If $R$ is a subring of a ring $S$, then $S$ is an $R$-module.

**Example 2.1.11.** Let $k$ be a field. Then a $k[X]$-module is the same as a $k$-vector space equipped with an endomorphism. Indeed, given some $k[X]$-module $M$, consider the underlying $k$-vector space $M$, obtained by restricting the scalar multiplication from $k[X]$ to $k$; then $\varphi : M \to M : m \mapsto X \cdot m$ is clearly $k$-linear. Conversely, given a $k$-vector space $M$ equipped with an endomorphism $\varphi : M \to M$, one constructs a structure of $k[X]$-module on $M$ as follows: given $f = a_0 + a_1 X + \cdots + a_n X^n \in k[X]$, define

$$f \cdot m = a_0 m + a_1 \varphi(m) + \cdots + a_n \varphi^n(m)$$

where $\varphi^n = \varphi \circ \cdots \circ \varphi$ denotes $n$-fold composition. It is clear that this yields a $k[X]$-module structure.

*Quick question* 2.1.12. Reinterpret this example in the language of Remark 2.1.4.

**Example 2.1.13.** Let $k$ be a field. Let $G$ be a finite group. The *group algebra* $k[G]$ consists of all formal sums $\sum_{g \in G} a_g\, g$ where $a_g \in k$ for all $g \in G$. There is a natural ring structure on $k[G]$: addition is defined in the obvious way, multiplication is defined by linearly extending the group law on $G$. The ring $k[G]$ is typically non-commutative. The reader should check that a left $k[G]$-module is the same as a $k$-vector space $M$, equipped with a group homomorphism $G \to \mathrm{GL}_k(M)$ (i.e., a *$k$-linear representation of* $G$).

**Definition 2.1.14.** Let $R$ be a ring and let $M, N$ be $R$-modules. An *$R$-module homomorphism* from $M$ to $N$ is a map $f : M \to N$ which is both $R$-linear (i.e., $r \cdot f(x) = f(r \cdot x)$ for all $r \in R$ and $x \in M$) and a homomorphism of abelian groups (i.e., $f(x + y) = f(x) + f(y)$ for all $x, y \in M$). An *$R$-module isomorphism* is a bijective $R$-module homomorphism (yielding the notion of *isomorphic $R$-modules*).

Clearly the composition of two $R$-module homomorphisms is again an $R$-module homomorphism. The set of all $R$-module homomorphisms from $M$ to $N$ will be denoted by $\mathrm{Hom}_R(M, N)$. If $M = N$, we write $\mathrm{End}_R(M)$ rather than $\mathrm{Hom}_R(M, M)$ for the set of $R$-module endomorphisms of $M$.

**Proposition 2.1.15.** *Let $R$ be a ring and let $M, N$ be given $R$-modules. Given two homomorphisms $f, g \in \mathrm{Hom}_R(M, N)$ and a scalar $r \in R$, define $f + g$ and $r \cdot f$ pointwise via*

$$(f + g)(x) = f(x) + g(x), \ (r \cdot f)(x) = r \cdot f(x).$$

*Then $\mathrm{Hom}_R(M, N)$ becomes an $R$-module under these operations.*

*Quick question* 2.1.16. Let $R$ be a ring, and let $M$ be any $R$-module. Show that $\mathrm{Hom}_R(R, M) \cong M$.

**Definition 2.1.17.** Let $R$ be a ring. An *$R$-submodule* $N$ of an $R$-module $M$ is a subgroup of $M$ which is closed under multiplications by scalars (elements of $R$). The *quotient* of $M$ by $N$ is the abelian group $M/N$ equipped with the $R$-module structure given by $r(x + N) = rx + N$ for all $r \in R$ and $x \in M$.

The quotient map $M \to M/N : x \mapsto x + N$ is a surjective homomorphism of $R$-modules. There is an inclusion-preserving bijection between submodules of $M$ which contain $N$ on the one hand, and submodules of $M/N$ on the other hand. This generalises Proposition 1.1.5 (see [Rotman, Theorem 7.11]).

Given an $R$-module $M$ and a collection of $R$-submodules $(N_\alpha)_{\alpha\in A}$ of $M$, the *sum* $\sum_{\alpha\in A} N_\alpha$ is the set of all sums $\sum_{\alpha\in A} x_\alpha$, where $x_\alpha \in N_\alpha$ for all $\alpha \in A$, and where only finitely many $x_\alpha$ are non-zero; this is again an $R$-submodule. Similarly, the *intersection* $\bigcap_{\alpha\in A} N_\alpha$ is again an $R$-submodule of $M$.

**Example 2.1.18.** Let $R$ be a ring and let $M$ be an $R$-module. If $I$ is an ideal in $R$, then the subset $IM$ of $M$ which consists of all finite sums of the form $\sum_k i_k x_k$, where $i_k \in I$ and $x_k \in M$ for all $k$, is an $R$-submodule of $M$. Moreover, the quotient $M/IM$ is not only an $R$-module, but also $R/I$-module if one defines scalar multiplication by the rule $(r + I) \cdot (x + IM) = rx + IM$.

**Definition 2.1.19.** Let $R$ be a ring, and let $f : M \to N$ be a homomorphism of $R$-modules. The *kernel* of $f$ is $\ker f = f^{-1}(0)$; this is easily seen to be an $R$-submodule of $M$. The *image* $\operatorname{im} f$ of $f$ is $f(M)$; this is an $R$-submodule of $N$. Finally, the *cokernel* of $f$ is $\operatorname{coker} f = N/\operatorname{im} f$.

As with groups or vector spaces, the map $M/\ker f \to \operatorname{im} f : m + \ker f \mapsto f(m)$ is an isomorphism. This is the "first isomorphism theorem" for $R$-modules, the proof of which we leave to the reader.

*Remark* 2.1.20. There are also natural analogues of the "second and third isomorphism theorems" for $R$-modules; for the statements and proofs, we refer to [Rotman, Theorem 7.9 and Theorem 7.10].

We end this section with the notion of *torsion element* of an $R$-module.

**Definition 2.1.21.** Let $R$ be a ring and let $M$ be an $R$-module. Then $x \in M$ is said to be *a torsion element* or simply *torsion* if there exists a non-zero $r \in R$ such that $rx = 0$. One says that $M$ *is torsion (resp. torsion free)* if every (resp. no non-zero) element of $M$ is torsion. The set $\operatorname{Ann}_R(x) = \{r \in R : rx = 0\}$ is an ideal of $R$, called the *annihilator* of $x$; then $x$ is torsion if and only if $\operatorname{Ann}_R(x) \neq 0$. Similarly, the *annihilator of* $M$ is $\operatorname{Ann}_R(M) = \{r \in R : rM = 0\}$ (this is again an ideal of $R$).

The set of torsion elements of $M$ is denoted by $T(M)$.

*Quick question* 2.1.22. It may be tempting to guess that an $R$-module $M$ is torsion if and only if $\operatorname{Ann}_R(M) \neq 0$, but this is not true in general. Find an example which illustrates this, e.g. for $R = \mathbf{Z}$.

*Quick question* 2.1.23. Let $R$ be a ring, let $M$ be an $R$-module, and let $I$ be an ideal of $R$ contained in $\operatorname{Ann}_R(M)$. Show that $M$ is also an $R/I$-module for the scalar multiplication given by $(r + I) \cdot x = rx$.

**Lemma 2.1.24.** *Let $R$ be an integral domain, and let $M$ be an arbitrary $R$-module. Then $T(M)$ is an $R$-submodule of $M$, called the* torsion submodule *of $M$.*

If $R$ is not an integral domain, and if $M$ is an $R$-module, then $T(M)$ need not be an $R$-submodule. For example, if $R = M = \mathbf{Z}/6$ with the "traditional" addition and multiplication, then both $\overline{2}$ and $\overline{3}$ are torsion elements (they are annihilated by $\overline{3}$ and $\overline{2}$ respectively), but their sum is not.

*Quick question* 2.1.25. Alternatively, one could define an element $x$ of an $R$-module $M$ to be torsion if there exists a non-zero $r \in R$ *which is not a zero divisor* such that $rx = 0$. Show that with this alternative definition, the torsion subset $T(M)$ of an $R$-module $M$ is always an $R$-submodule. However, Definition 2.1.21 seems to be the preferred one in the literature for a number of (mathematical and historical) reasons.

## 2.2 Constructing modules

Let us now take a look at a number of ways to construct modules, either by building new modules from old ones, or by describing them via *generating sets* (generalising the idea of a basis for a vector space).

**Definition 2.2.1.** Let $R$ be a ring, and let $(M_\alpha)_{\alpha\in A}$ be a collection of $R$-modules. The Cartesian product $\prod_{\alpha\in A} M_\alpha$ becomes an $R$-module when equipped with componentwise addition, and scalar multiplication given by $r\cdot(x_\alpha)_{\alpha\in A} = (rx_\alpha)_{\alpha\in A}$. This is the *direct product* of the collection $(M_\alpha)_{\alpha\in A}$.

The *direct sum* $\bigoplus_{\alpha\in A} M_\alpha$ is the $R$-submodule of $\prod_{\alpha\in A} M_\alpha$ defined as follows:

$$\bigoplus_{\alpha\in A} M_\alpha = \left\{ (x_\alpha)_{\alpha\in A} \in \prod_{\alpha\in A} M_\alpha : x_\alpha = 0 \text{ for almost all } \alpha \right\}$$

where as usual "for almost all" means that only finitely many exceptions are allowed. In particular, if $A$ is finite, then the direct sum $\bigoplus_{\alpha\in A} M_\alpha$ is equal to the direct product $\prod_{\alpha\in A} M_\alpha$. In the special case where $M_\alpha = M$ for all $\alpha\in A$, we denote $\prod_{\alpha\in A} M_\alpha$ and $\bigoplus_{\alpha\in A} M_\alpha$ by $M^A$ and $M^{(A)}$ respectively.

In the notation of this definition, we have, for all $\beta\in A$, natural inclusion homomorphisms

$$\iota_\beta : M_\beta \to \bigoplus_{\alpha\in A} M_\alpha : x \mapsto \iota_\beta(x), \quad \begin{cases} \iota_\beta(x)_\alpha = x & \text{if } \alpha = \beta \\ \iota_\beta(x)_\alpha = 0 & \text{if } \alpha \neq \beta \end{cases}$$

and projection homomorphisms $\pi_\beta : \prod_{\alpha\in A} M_\alpha \to M_\beta : (x_\alpha)_{\alpha\in A} \mapsto x_\beta$.

**Definition 2.2.2.** Let $R$ be a ring and let $M$ be an $R$-module. Given a subset $S\subseteq M$, the $R$-submodule of $M$ *generated (or spanned) by* $S$ is the smallest $R$-submodule of $M$ which contains $S$ or, equivalently, the intersection of all $R$-submodules of $M$ which contain $S$. The set $S$ *generates* $M$ if $M = \langle S\rangle$.

The $R$-submodule from the definition is usually denoted by $\langle S\rangle$. It is straightforward to see that $\langle S\rangle$ consists of all finite $R$-linear combinations of elements of $S$.

**Definition 2.2.3.** Let $R$ be a ring. An $R$-module $M$ is said to be *finitely generated* if there exists a finite subset $S = \{x_1,\cdots,x_n\}$ of $M$ such that $M = \langle S\rangle = \langle x_1,\cdots,x_n\rangle$. An $R$-module is *cyclic* if it can be generated by a single element. Given $x\in M$, we call $\langle x\rangle$ the *cyclic submodule generated by* $x$.

**Lemma 2.2.4.** *Let $R$ be a ring, and let $M$ be an $R$-module. Then $M$ is a cyclic $R$-module if and only if $M$ is isomorphic (as an $R$-module) to $R/I$, for some ideal $I$ in $R$.*

*Proof.* Let $x\in M$ be such that $M = \langle x\rangle$. Consider the $R$-module homomorphism $\varphi : R\to M : r\mapsto rx$. Then $\varphi$ is surjective since $M = \langle x\rangle$, and therefore $M \cong R/\ker\varphi$ by the first isomorphism theorem.

Conversely, given an ideal $I$ in $R$, the $R$-module $R/I$ is cyclic with generator $1+I$. □

Given a ring $R$ and an arbitrary set $I$, the $R$-module $R^{(I)}$ consists of all $I$-tuples of elements of $R$, with only finitely many components non-zero. Given $i\in I$, let us consider the "standard basis element" $e_i = (\cdots,0,1,0,\cdots)$ with $1$ at position $i$. It is easy to see that an arbitrary element $x = (x_i)_{i\in I}$ of $R^{(I)}$ can be written as a finite linear combination of the basis elements $e_i$, in other words: $(e_i)_{i\in I}$ generates $R^{(I)}$. Note that if $I$ is infinite, then $(e_i)_{i\in I}$ does *not* generate $R^I$.

*Quick question* 2.2.5. Let $R$ be a ring and let $S, T$ be arbitrary sets. Show that $(R^{(S)})^{(T)} \cong R^{(S\times T)}$.

**Lemma 2.2.6.** *Let $R$ be a ring and let $M$ be an $R$-module. Given a collection $(x_i)_{i\in I}$ of elements of $M$, there exists a unique $R$-module homomorphism $\varphi : R^{(I)} \to M$ such that $\varphi(e_i) = x_i$ for all $i\in I$.*

*Furthermore, we have $M = \langle x_i\rangle_{i\in I}$ if and only if $\varphi$ is surjective.*

**Definition 2.2.7.** In the setting of the previous lemma, we say that $(x_i)_{i\in I}$ is a *basis* for $M$ if $\varphi$ is an isomorphism, and we say that $M$ is a *free $R$-module* if there exists a basis for $M$ as an $R$-module. If there exists a basis of cardinality $n$, then we say that $M$ is *free of rank* $n$; in that case, we have $M\cong R^n$.

In contrast with the case of vector spaces (modules over a field), an $R$-module may very well not be free, even if it is finitely generated. For example, the $\mathbf{Z}$-module $\mathbf{Z}/2$ does not have a basis (why?).

*Quick question* 2.2.8. Show that $\mathbf{Q}$ is neither free, nor finitely generated as a $\mathbf{Z}$-module.

(Hint for the first part: show first that any two distinct elements of $\mathbf{Q}$ are "$\mathbf{Z}$-linearly dependent" and hence cannot belong to a basis together – make this precise! Then show that $\mathbf{Q}$ is not a cyclic $\mathbf{Z}$-module.)

**Proposition 2.2.9.** *Let $R$ be a ring and let $M$ be an $R$-module with basis $(x_i)_{i\in I}$. For every $x \in M$, there exists a unique $I$-tuple $(r_i)_{i\in I}$, with $r_i = 0$ for almost all $i \in I$, such that $x = \sum_{i\in I} r_i x_i$. In other words: every element of $M$ can be written uniquely as a finite $R$-linear combination of basis elements.*

For a vector space over a field, we know that any two bases have the same cardinality. The following proposition says that this result remains true for free modules of finite rank over a commutative ring:

**Proposition 2.2.10.** *Let $R$ be a non-zero ring. If $R^m \cong R^n$ for some positive integers $m, n$, then $m = n$.*

*Proof.* Let $M = R^m$, $N = R^n$ and let $\varphi : M \to N$ be an $R$-module isomorphism. Let $I$ be a maximal ideal. Then $\varphi$ induces an $R/I$-module isomorphism $\overline{\varphi} : M/IM \to N/IN$ (see Example 2.1.18; why?). It is not hard to see that $M/IM \cong (R/I)^m$ and $N/IN \cong (R/I)^n$. Since $R/I$ is a field and since a vector space over a field has a well-defined dimension, we deduce that $m = n$, as desired. □

*Remark* 2.2.11. It may be rather tempting to believe that Proposition 2.2.10 must be true in greater generality, but caution is advised. For example, an $R$-module may have minimal generating sets of different cardinality: the $\mathbf{Z}$-module $\mathbf{Z}$ has both $\{1\}$ and $\{2, 3\}$ as minimal generating sets.

If one allows the ring $R$ to be non-commutative, then even the cardinality of a basis of free $R$-module is no longer well-defined! In fact there exist examples of non-commutative rings $R$ such that $R \cong R^2$ as left $R$-modules, see Exercise 2.17.

Even if an $R$-module is very often not free, any $R$-module can be written as a quotient of a free $R$-module. Finitely generated $R$-modules are exactly those which are isomorphic to a quotient of $R^n$, for some positive integer $n$, by a submodule: if $M = \langle x_1, \cdots, x_n\rangle$, then $\varphi : R^n \to M$ given by $\varphi(e_i) = x_i$ for $i = 1, \cdots, n$ is a surjective homomorphism of $R$-modules, and hence $M \cong R^n/\ker\varphi$. Conversely, if $N$ is an $R$-submodule of $R^n$, then $R^n/N = \langle e_1 + N, \cdots, e_n + N\rangle$ is certainly finitely generated.

## 2.3 Cayley-Hamilton

The classic Cayley-Hamilton theorem from undergraduate linear algebra says that any endomorphism of a finite-dimensional vector space satisfies its own characteristic polynomial. We will see that a more general statement holds for finitely generated modules, and we will discuss some of the consequences.

**Theorem 2.3.1.** *Let $R$ be a ring, let $M$ be an $R$-module generated by $n$ elements, and let $\varphi \in \mathrm{End}_R(M)$. Let $I$ be an ideal of $R$ such that $\varphi(M) \subseteq IM$. Then $\varphi$ satisfies an equation of the form*

$$\varphi^n + c_1\varphi^{n-1} + \cdots + c_{n-1}\varphi + c_n\mathrm{Id}_M = 0$$

*where $c_k \in \overline{I}^k$ for $1 \leq k \leq n$.*

This equation should be understood as an equality of elements of the (typically non-commutative) ring $\mathrm{End}_R(M)$ or, even better, of the commutative subring $\overline{R}[\varphi]$ of $\mathrm{End}_R(M)$. Here $\overline{R}$ denotes the image of $R$ in $\mathrm{End}_R(M)$ under the map which sends $r \in R$ to the endomorphism $M \to M : x \mapsto rx$. Since two different elements $r, r' \in R$ yield the same element of $\mathrm{End}_R(M)$ if and only if $r - r' \in \mathrm{Ann}_R(M)$, we have $\overline{R} \cong R/\mathrm{Ann}_R(M)$. Finally, $\overline{I}$ denotes the image of $I$ in $\overline{R}$.

*Proof.* First, observe that $M$ becomes an $\overline{R}[\varphi]$-module in the obvious way. Let $x_1, \cdots, x_n$ generate $M$ as an $R$-module. Since $\varphi(x_i) \in IM$ for $i = 1, \cdots, n$, there exists a $(n \times n)$-matrix $(a_{ij})_{1 \leq i,j \leq n}$ with entries in $I$ such that $\varphi(x_i) = \sum_{j=1}^{n} a_{ij} x_j$ for all $i$ (when $M$ is seen as an $R$-module) or, equivalently, $\varphi x_i = \sum_{j=1}^{n} \overline{a_{ij}} x_j$ (when $M$ is seen as an $\overline{R}[\varphi]$-module). In other words, we have

$$\sum_{j=1}^{n} (\delta_{ij}\varphi - \overline{a_{ij}})\, x_j = 0$$

for all $i$, where $\delta_{ij}$ is the Kronecker delta (equal to 1 when $i = j$ and 0 otherwise).

From here it is not hard to conclude (we leave the details to the reader): let $A = (\delta_{ij}\varphi - \overline{a_{ij}})_{1 \leq i,j \leq n}$. Multiplying the equation $A(x_1, \cdots, x_n)^T = 0$ on the left by $\operatorname{adj} A$ (which again has entries in $\overline{R}[\varphi]$) shows that $\det A \in \overline{R}[\varphi]$ annihilates each $x_i$, hence must be the zero element of the ring $\operatorname{End}_R(M)$. Expanding out $\det A$ yields the result. □

*Quick question* 2.3.2. Explain why Theorem 2.3.1 and its proof do indeed provide a wide generalisation of the Cayley-Hamilton theorem, classically stated for finite-dimensional vector spaces over a field.

**Corollary 2.3.3.** *Let $R$ be a ring and let $M$ be a finitely generated $R$-module. Assume that there exists an ideal $I$ of $R$ such that $IM = M$. Then there exists an element $x \equiv 1 \pmod{I}$ such that $xM = 0$, in other words: $x \in \operatorname{Ann}_R(M)$. In particular, if $I$ is contained in the Jacobson radical $\mathfrak{j}_R$, then $M = 0$.*

*Proof.* The first statement follows from Theorem 2.3.1: it suffices to take $\varphi = \operatorname{Id}_M$ in the statement. The second statement is a consequence of the first one and Proposition 1.2.19: if $I \subseteq \mathfrak{j}_R$, then $1 + I \subseteq R^\times$. □

The last statement of Corollary 2.3.3 is known as *Nakayama's lemma* and will turn out to be very useful. Therefore we state the following special case separately for future use.

**Corollary 2.3.4.** *Let $R$ be a local ring with maximal ideal $\mathfrak{m}$, and let $M$ be a finitely generated $R$-module. If $M = \mathfrak{m}M$, then $M = 0$.*

*Quick question* 2.3.5. Check that this is indeed a special case of Corollary 2.3.3.

**Proposition 2.3.6.** *Let $R$ be a local ring with maximal ideal $\mathfrak{m}$ and residue field $\kappa = R/\mathfrak{m}$, and let $M$ be a finitely generated $R$-module. Then $M/\mathfrak{m}M$ is a finite-dimensional vector space over $\kappa$. Let $x_1, \cdots, x_n$ be elements of $M$ whose images in $M/\mathfrak{m}M$ form a basis for this vector space. Then $M = \langle x_1, \cdots, x_n \rangle$.*

*Proof.* We sketch the proof and leave the details to the reader. Let $N = \langle x_1, \cdots, x_n \rangle$. The composition $N \hookrightarrow M \to M/\mathfrak{m}M$ is a surjective homomorphism of $R$-modules. It follows that $M = N + \mathfrak{m}M$, or equivalently $M/N = \mathfrak{m}(M/N)$. An application of Corollary 2.3.4 now yields the result. □

## 2.4 Exact sequences

Exact sequences are a useful and ubiquitous tool in commutative algebra; they allow to replace tedious, verbose arguments involving submodules and quotients by quick, transparent "diagram chases".

**Definition 2.4.1.** Let $R$ be a ring. A sequence of homomorphisms of $R$-modules

$$\cdots \longrightarrow M_{i-1} \xrightarrow{f_{i-1}} M_i \xrightarrow{f_i} M_{i+1} \longrightarrow \cdots \tag{2.1}$$

is *exact at* $M_i$ if $\ker f_i = \operatorname{im} f_{i-1}$ as $R$-submodules of $M_i$.

The sequence is *exact* if it is exact at every $R$-module appearing in the sequence.

**Example 2.4.2.** A sequence of the form $0 \to M \xrightarrow{f} N$ is exact if and only if $f : M \to N$ is injective. A sequence of the form $M \xrightarrow{f} N \to 0$ is exact if and only if $f : M \to N$ is surjective.

**Definition 2.4.3.** Let $R$ be a ring. A *short exact sequence* of $R$-modules (and $R$-module homomorphisms) is an exact sequence of the form $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$.

In this definition, exactness of the sequence is equivalent to $f$ being injective, $g$ being surjective and $\ker g = \operatorname{im} f$. This means that $f$ induces an isomorphism $M_1 \cong \ker g$ and that $g$ induces an isomorphism $M_3 \cong \operatorname{coker} f$. In other words: $f$ can be thought of as the inclusion of the submodule $M_1$ into $M_2$, and $g$ can be thought of as the quotient of $M_2$ by the submodule $M_1$.

In this situation of Definition 2.4.3, $M_2$ is said to be an *extension* of $M_3$ by $M_1$.

*Quick question* 2.4.4. Let $R$ be a ring and let $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$ be an exact sequence of $R$-module homomorphisms. Show that $\operatorname{coker} f \cong \ker h$.

Any long exact sequences such as (2.1) can be split up into short exact sequences:

**Lemma 2.4.5.** *With the notation of Definition 2.4.1, define* $N_i = \operatorname{im} f_{i-1} = \ker f_i$. *Then we have short exact sequences* $0 \longrightarrow N_i \longrightarrow M_i \longrightarrow N_{i+1} \longrightarrow 0$ *for all* $i$*, where* $N_i \to M_i$ *is the inclusion map and* $M_i \to N_{i+1}$ *is the obvious map induced by* $f_i$.

Let $R$ be a ring and let $u : M_1 \to M_2$ be an $R$-module homomorphism. Then we have a map

$$\operatorname{Hom}_R(u, N) : \operatorname{Hom}_R(M_2, N) \to \operatorname{Hom}_R(M_1, N)$$

(for any $R$-module $N$) induced by $u$, defined simply by sending $f : M_2 \to N$ to $f \circ u : M_1 \to N$. We leave it to the reader to check that $\operatorname{Hom}_R(u, N)$ is indeed an $R$-module homomorphism. Similarly, if we are given a homomorphism $v : N_1 \to N_2$, then we obtain an induced homomorphism of $R$-modules

$$\operatorname{Hom}_R(M, v) : \operatorname{Hom}_R(M, N_1) \to \operatorname{Hom}_R(M, N_2)$$

by mapping $f : M \to N_1$ to $v \circ f : M \to N_2$.

*Quick question* 2.4.6. Let $R$ be a ring, let $u_1 : M_1 \to M_2$ and $u_2 : M_2 \to M_3$ be $R$-module homomorphisms. Show that for any $R$-module $N$, $\operatorname{Hom}_R(u_1, N) \circ \operatorname{Hom}_R(u_2, N) = \operatorname{Hom}_R(u_2 \circ u_1, N)$.

State and prove a similar result in the “dual” situation.

**Proposition 2.4.7.** *Let* $R$ *be a ring. Let* $u_1 : M_1 \to M_2$, $u_2 : M_2 \to M_3$ *be* $R$*-module homomorphisms. Then*

$$M_1 \xrightarrow{u_1} M_2 \xrightarrow{u_2} M_3 \longrightarrow 0$$

*is exact if and only if for every* $R$*-module* $N$*, the induced sequence*

$$0 \longrightarrow \operatorname{Hom}_R(M_3, N) \longrightarrow \operatorname{Hom}_R(M_2, N) \longrightarrow \operatorname{Hom}_R(M_1, N)$$

*involving* $\operatorname{Hom}_R(u_2, N)$ *and* $\operatorname{Hom}_R(u_1, N)$ *is exact.*

*Similarly, given* $R$*-module homomorphisms* $v_1 : N_1 \to N_2$ *and* $v_2 : N_2 \to N_3$*, then*

$$0 \longrightarrow N_1 \xrightarrow{v_1} N_2 \xrightarrow{v_2} N_3$$

*is exact if and only if for every* $R$*-module* $M$*, the induced sequence*

$$0 \longrightarrow \operatorname{Hom}_R(M, N_1) \longrightarrow \operatorname{Hom}_R(M, N_2) \longrightarrow \operatorname{Hom}_R(M, N_3)$$

*involving* $\operatorname{Hom}_R(M, v_1)$ *and* $\operatorname{Hom}_R(M, v_2)$ *is exact.*

*Proof.* For the proof of half of this theorem (the "only if" part), see [Rotman, Theorems 7.44 and 7.46]. The proof of the other half is left to the reader. □

Given a ring $R$ and two $R$-modules $M$ and $N$, we have an exact sequence

$$0 \longrightarrow M \xrightarrow{\iota} M \oplus N \xrightarrow{\pi} N \longrightarrow 0$$

where $\iota : M \to M \oplus N : m \mapsto (m, 0)$ is the inclusion and $\pi : M \oplus N \to N : (m, n) \mapsto n$ is the projection. It is often useful to recognize whether or not a given exact sequence is of this particular form.

**Proposition 2.4.8.** *Let $R$ be a ring and let $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ be a short exact sequence of $R$-modules. The following conditions are equivalent:*

*(1) there exists an $R$-module homomorphism $r : M_2 \to M_1$ such that $r \circ f = \mathrm{id}_{M_1}$,*

*(2) there exists an $R$-module homomorphism $s : M_3 \to M_2$ such that $g \circ s = \mathrm{id}_{M_3}$,*

*(3) there exists an isomorphism $h : M_2 \to M_1 \oplus M_3$ such that the diagram*

$$\begin{tikzcd}
0 \arrow[r] & M_1 \arrow[r, "f"] \arrow[d, "\mathrm{id}_{M_1}"] & M_2 \arrow[r, "g"] \arrow[d, "h"] & M_3 \arrow[r] \arrow[d, "\mathrm{id}_{M_3}"] & 0 \\
0 \arrow[r] & M_1 \arrow[r, "\iota"] & M_1 \oplus M_3 \arrow[r, "\pi"] & M_3 \arrow[r] & 0
\end{tikzcd} \tag{2.2}$$

*commutes.*

*If these conditions are satisfied, then the exact sequence is said to be* split *or* split exact, *and $M_2$ is said to be a* split extension *of $M_3$ by $M_1$. A map $r$ as described in (1) is called a* retraction *of $f$, and a map $s$ as described in (2) is called a* section *of $g$.*

*Proof.* It is easy to see that (3) implies (1) and (2).

To show that (2) implies (1), assume that $s$ as in (2) exists. Given $y \in M_2$, take $\tilde{y} = y - s(g(y)) \in M_2$. Then $\tilde{y} \in \ker g = \operatorname{im} f$, and hence there exists a unique $x \in M_1$ such that $f(x) = \tilde{y}$. Defining $r(y) = x$ now yields the desired retraction $r : M_2 \to M_1$. (Why is this indeed a retraction?)

To show that (1) implies (3), take $r$ as in (1), and consider $h : M_2 \to M_1 \oplus M_3 : y \mapsto (r(y), g(y))$. An easy computation then shows that $h$ makes (2.2) commute.

To show that $h$ is injective, let $y \in \ker h$, then certainly $g(y) = 0$, so that $y \in \ker g = \operatorname{im} f$. However, $h(y) = 0$ forces $r(y) = 0$ as well; since $r$ is injective on $\operatorname{im} f$ (indeed, $r \circ f = \mathrm{id}_{M_1}$), we get $y = 0$.

To show that $h$ is surjective, let $(x, z) \in M_1 \oplus M_3$. Pick $y_0 \in M_2$ such that $g(y_0) = z$; this is certainly possible since $g$ is surjective. Then $y = y_0 - f(r(y_0)) + f(x)$ satisfies $h(y) = (x, z)$, as desired. □

**Example 2.4.9.** The short exact sequences

$$0 \longrightarrow \mathbf{Z} \xrightarrow{\cdot 2} \mathbf{Z} \longrightarrow \mathbf{Z}/2 \longrightarrow 0$$

$$0 \longrightarrow \mathbf{Z}/2 \xrightarrow{\cdot \bar{2}} \mathbf{Z}/4 \longrightarrow \mathbf{Z}/2 \longrightarrow 0$$

are *not* split.

Let us finally state two useful lemma's involving exact sequences and commutative diagrams, the *five lemma* and the *snake lemma*, proven via *diagram chasing*: the art of pointing fingers at different parts of a diagram, pushing elements around and stating their fate. Such proofs are mostly free of content and creativity, but are sometimes headache-provokingly resistant to being rendered or read in prose.

**Proposition 2.4.10** ("Five lemma")**.** *Let $R$ be a ring and consider a commutative diagram*

$$\begin{tikzcd}
M_1 \arrow[r] \arrow[d, "f_1"] & M_2 \arrow[r] \arrow[d, "f_2"] & M_3 \arrow[r] \arrow[d, "f_3"] & M_4 \arrow[r] \arrow[d, "f_4"] & M_5 \arrow[d, "f_5"] \\
N_1 \arrow[r] & N_2 \arrow[r] & N_3 \arrow[r] & N_4 \arrow[r] & N_5
\end{tikzcd}$$

*of $R$-modules. Assume that the rows are exact. Then, if $f_1$, $f_2$, $f_4$ and $f_5$ are isomorphisms, so is $f_3$. More precisely:*

*(1) If $f_1$ is surjective and if $f_2$ and $f_4$ are injective, then $f_3$ is injective.*

*(2) If $f_5$ is injective and if $f_2$ and $f_4$ are surjective, then $f_3$ is surjective.*

*Proof.* We will only prove the first claim, and leave the second claim as an exercise for the reader.

Let $x_3 \in \ker f_3$; our goal is to show that $x_3 = 0$. The image $x_4$ of $x_3$ in $M_4$ gets mapped to 0 by $f_4$, by commutativity of the diagram. Since $f_4$ is injective, we know that $x_4 = 0$. By exactness of the upper row, we know that $x_3$ is the image in $M_3$ of some $x_2 \in M_2$. Let $y_2 = f_2(x_2)$.

Again by commutativity of the diagram, we see that $y_2$ gets mapped to 0 in $N_3$. Exactness of the lower row implies that $y_2$ is the image in $N_2$ of some $y_1 \in N_1$. Since $f_1$ is surjective, there exists $x_1 \in M_1$ such that $f_1(x_1) = y_1$. We claim that $x_2$ is the image in $M_2$ of $x_1$. Indeed, denoting this image by $x_2'$, we see that $f_2(x_2') = y_2$ by commutativity of the diagram, hence $f_2(x_2') = f_2(x_2)$; since $f_2$ is injective, this yields $x_2 = x_2'$. But now we are done: $x_3$ comes all the way from $x_1 \in M_1$, and hence is zero. □

**Proposition 2.4.11** ("Snake lemma")**.** *Let $R$ be ring. Assume that in the diagram below, the maps $\varphi_1$, $\varphi_2$, $\varphi_3$, $f_1$, $f_2$, $g_1$ and $g_2$ are $R$-module homomorphisms which make all solid arrows commute. Assume moreover that the rows consisting of solid arrows are exact. Then there exist maps $\overline{f}_1$, $\overline{f}_2$, $\partial$, $\overline{g}_1$ and $\overline{g}_2$, as indicated below by dashed arrows, such that the six-term sequence formed by these arrows is exact.*

$$\begin{tikzcd}
& \ker\varphi_1 \arrow[r, dashed, "\overline{f}_1"] \arrow[d] & \ker\varphi_2 \arrow[r, dashed, "\overline{f}_2"] \arrow[d] & \ker\varphi_3 \arrow[d] \arrow[dddll, dashed, "\partial"] & \\
& M_1 \arrow[r, "f_1"] \arrow[d, "\varphi_1"] & M_2 \arrow[r, "f_2"] \arrow[d, "\varphi_2"] & M_3 \arrow[r] \arrow[d, "\varphi_3"] & 0 \\
0 \arrow[r] & N_1 \arrow[r, "g_1"] \arrow[d] & N_2 \arrow[r, "g_2"] \arrow[d] & N_3 \arrow[d] & \\
& \operatorname{coker}\varphi_1 \arrow[r, dashed, "\overline{g}_1"] & \operatorname{coker}\varphi_2 \arrow[r, dashed, "\overline{g}_2"] & \operatorname{coker}\varphi_3 &
\end{tikzcd}$$

*Proof.* The maps $\overline{f}_1$, $\overline{f}_2$, $\overline{g}_1$ and $\overline{g}_2$ are easy to construct, and proving exactness at $\ker\varphi_2$ and $\operatorname{coker}\varphi_2$ is not hard either; this part of the proof is left as an exercise for the reader.

The construction of the so-called *connecting homomorphism* or *boundary map* $\partial : \ker \varphi_3 \to \operatorname{coker} \varphi_1$ (sometimes also called "snake map") is more intricate, so we will explain it in detail. Pick $x_3 \in \ker \varphi_3$.

Since $f_2$ is surjective, there exists $x_2 \in M_2$ such that $f_2(x_2) = x_3$. Let $y_2 = \varphi_2(x_2)$, then since $x_3 \in \ker \varphi_3$ and $\varphi_3 \circ f_2 = g_2 \circ \varphi_2$, we have $g_2(y_2) = 0$. Since $\ker g_2 = \operatorname{im} g_1$, there exists $y_1 \in N_1$ such that $g_1(y_1) = y_2$. We now propose the image of $y_1 \in N_1$ in $\operatorname{coker} \varphi_1$ as our candidate for $\partial(x_3)$.

Is this assignment well-defined? We have picked a preimage $x_2$ for $x_3$, and a preimage $y_1$ for $y_2$. In the latter case, there was no choice to be made ($g_1$ is injective). However, the choice of $x_2$ was potentially ambiguous. Therefore we have to check that choosing a different $x_2'$ leads to the same value of $\partial(x_3)$.

If $f_2$ maps both $x_2$ and $x_2'$ to $x_3$, then $\Delta_2 = x_2 - x_2' \in \ker f_2 = \operatorname{im} f_1$. Choose $\Delta_1 \in M_1$ such that $f_1(\Delta_1) = \Delta_2$. Then $y_2$ and $y_2' = \varphi_2(x_2')$ differ by $\varphi_2(\Delta_2) = g_1(\varphi_1(\Delta_1))$. Since $g_1$ is injective, we see that $y_1$ and $y_1'$ (the preimage of $y_2'$ under $g_1$) differ by $\varphi_1(\Delta_1)$. However, $\varphi_1(\Delta_1)$ dies in $\operatorname{coker} \varphi_1$, which confirms that $y_1$ and $y_1'$ have the same image in $\operatorname{coker} \varphi_1$, as desired.

It follows easily from the construction that $\partial$ is an $R$-module homomorphism. Proving exactness at $\ker \varphi_3$ and $\operatorname{coker} \varphi_1$ requires more work, but is left once more to the reader as an exercise. □

*Remark* 2.4.12. The snake lemma figured in the Hollywood movie "It's my turn", see this link.

## Exercises

*Easy exercise* 2.1. Let $R$ be an integral domain and let $I$ be an ideal in $R$. Show that the $R$-module $I$ is isomorphic to $R$ (regarded as an $R$-module over itself in the obvious way) if and only if $I$ is principal.

*Easy exercise* 2.2. Let $R$ be a ring. A *simple* $R$-module is an $R$-module with exactly two submodules, namely $0$ and itself. (In particular, the zero module is *not* simple.)

(a) If $M$ is a simple $R$-module, show that $M \cong R/\mathfrak{m}$ for some maximal ideal $\mathfrak{m}$ of $R$.

(b) Give an example of a ring $R$ and an $R$-module $M$ which does not have any simple submodule.

(c) Let $M$ and $N$ be simple $R$-modules, and let $f : M \to N$ be an $R$-module homomorphism. Show that $f$ is either the zero homomorphism, or an isomorphism. (This is *Schur's lemma*.)

*Easy exercise* 2.3. Let $R$ be a ring with ideals $I$ and $J$. Show that $I(R/J) \cong (I + J)/J$ as $R$-modules.

*Easy exercise* 2.4. Let $R$ be a ring and let $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$ be an exact sequence of $R$-modules. Show that $\ker h \cong \operatorname{coker} f$ as $R$-modules.

*Easy exercise* 2.5. Give an example of a ring $R$, an $R$-module $M$ which is finitely generated, and a submodule $N$ of $M$ which is *not* finitely generated as an $R$-module.

*Exercise* 2.6. Construct a ring $R$ and a non-zero $R$-module $M$ such that $M \cong M \oplus M$ as $R$-modules.

*Exercise* 2.7. Let $R$ be a ring and let $M$ be an $R$-module. The *dual* of $M$ is $M^\vee = \operatorname{Hom}_R(M, R)$. Construct a natural homomorphism of $R$-modules $\varphi : M \to (M^\vee)^\vee$, and show that $\varphi$ need not be an isomorphism in general, even if $M$ is finitely generated. Finally, show that $(M_1 \oplus M_2)^\vee = M_1^\vee \oplus M_2^\vee$ for arbitrary $R$-modules $M_1$ and $M_2$, and deduce that $\varphi$ is an isomorphism if $M$ is free of finite rank.

*Exercise* 2.8. Recall the statement of *Nakayama's lemma*: if $R$ is a ring and if $M$ is a finitely generated $R$-module such that $IM = M$ for some ideal $I$ of $R$ contained in the Jacobson radical $\mathfrak{j}_R$, then $M = 0$.

Below we sketch an alternative proof (which does not use the Cayley-Hamilton theorem for modules) of Nakayama's lemma. Provide full details for all steps of the proof.

If $M \neq 0$, let $\{x_1, \cdots, x_n\}$ be a generating set for $M$ which is minimal in the sense that no proper subset generates $M$. Then $x_1 = \sum_{k=1}^{n} i_k x_k$ for certain $i_k \in I$. Since $1 - i_1$ is a unit, we see that $M = \langle x_2, \cdots, x_n \rangle$, contradicting the minimality assumption. Hence $M = 0$. □

*Exercise* 2.9. Let $R$ be a ring, let $M$ be a finitely generated $R$-module and let $f : M \to R^n$ be a surjective homomorphism of $R$-modules, where $n \geq 1$ is an integer. Show that $\ker f$ is finitely generated.

*Exercise* 2.10. Let $R$ be a ring. Assume that $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ is a short exact sequence of $R$-modules. Let $I_k = \operatorname{Ann}_R(M_k)$ for $k \in \{1, 2, 3\}$. Show that $I_1 I_3 \subseteq I_2 \subseteq I_1 \cap I_3$.

*Exercise* 2.11. Let $R$ be a ring. Assume that $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ is a short exact sequence of $R$-modules. If $M_1$ and $M_2$ are finitely generated, is the same true for $M_3$? If $M_1$ and $M_3$ are finitely generated, is the same true for $M_2$? If $M_2$ and $M_3$ are finitely generated, is the same true for $M_1$?

*Exercise* 2.12. Let $R$ be a ring and let $M$ and $(N_\alpha)_{\alpha \in A}$ be $R$-modules. Show that

$$\operatorname{Hom}_R\left(M, \prod_{\alpha \in A} N_\alpha\right) \cong \prod_{\alpha \in A} \operatorname{Hom}_R(M, N_\alpha).$$

Similarly, if $(M_\alpha)_{\alpha \in A}$ and $N$ are $R$-modules, show that

$$\operatorname{Hom}_R\left(\bigoplus_{\alpha \in A} M_\alpha, N\right) \cong \prod_{\alpha \in A} \operatorname{Hom}_R(M_\alpha, N).$$

*Exercise* 2.13. Let $R$ be a commutative ring. Let $I$ be a nilpotent ideal of $R$, i.e., an ideal such that $I^k = 0$ for some $k \geq 1$. Let $M$ and $N$ be $R$-modules and let $\varphi : M \to N$ be an $R$-module homomorphism. Prove that if the induced homomorphism $\overline{\varphi} : M/IM \to N/IN$ is surjective, then so is $\varphi$.

*Exercise* 2.14. Let $R$ be a ring. If $I, J$ are coprime ideals of $R$, show that $I \oplus J \cong R \oplus IJ$ as $R$-modules.

*Exercise* 2.15. Let $R$ be a ring and let $M$ be an $R$-module.

(a) Assume that $R$ is an integral domain. Show that the quotient $M/T(M)$ is a torsion free $R$-module. (Recall that $T(M)$ is an $R$-submodule of $M$ if $R$ is an integral domain.)

(b) Give an example of a ring $R$ such that its set of torsion elements $T(R) \subseteq R$ is a submodule of $R$ (seen as an $R$-module over itself), and yet the quotient $R/T$ is *not* torsion free.

*Exercise* 2.16. Let $R$ be a ring and let $M$ be a finitely generated $R$-module.

(a) Let $f : M \to M$ be a surjective $R$-module endomorphism. Show that $f$ is injective. (Hint: turn $M$ into an $R[X]$-module by letting $X$ act as $f$, and use Nakayama's lemma.)

(b) Let $f : M \to M$ be an injective $R$-module endomorphism. Is $f$ necessarily surjective?

(c) Does the statement in (a) remain true without the condition that $M$ be finitely generated?

*Exercise* 2.17. Let $k$ be a field. Let $V$ be a $k$-vector space with a countably infinite basis $(e_1, e_2, \cdots)$. Let $R = \operatorname{End}_k(V)$ be the (non-commutative) ring of $k$-linear endomorphisms of $V$. Define $\phi_1, \phi_2 \in R$ by $\phi_1(e_{2i}) = e_i$ and $\phi_1(e_{2i-1}) = 0$ for all $i \geq 1$, and $\phi_2(e_{2i}) = 0$ and $\phi_2(e_{2i-1}) = e_i$ for all $i \geq 1$. Show that $\{\phi_1, \phi_2\}$ is a basis for $R$ as a left $R$-module, whence an isomorphism of left $R$-modules $R \cong R^2$.

*Exercise* 2.18. Let $R$ be a ring and let $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ be an exact sequence of $R$-modules. If there exists an $R$-module homomorphism $M_2 \longrightarrow M_1 \oplus M_3$ such that

$$\begin{tikzcd}
0 \arrow[r] & M_1 \arrow[r] \arrow[d, equal] & M_2 \arrow[r] \arrow[d, dashed] & M_3 \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & M_1 \arrow[r, "\iota"] & M_1 \oplus M_3 \arrow[r, "\pi"] & M_3 \arrow[r] & 0
\end{tikzcd} \tag{2.3}$$

commutes, where the bottom line is the standard direct sum sequence, show that the top line must be split.

*Exercise* 2.19. Let $R$ be a ring, and let $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ be an exact sequence of $R$-modules. Show that if $M_3$ is a free $R$-module, then this sequence is split; in other words: any extension of a free $R$-module by an arbitrary $R$-module must be split.

*Exercise* 2.20. Let $R$ be a ring. Consider a commutative diagram of $R$-modules of the form

$$\begin{tikzcd}
 & & & 0 \arrow[d] \\
 & A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[d] \\
0 \arrow[r] & A' \arrow[r] \arrow[d] & B' \arrow[r] \arrow[d] & C' \\
 & A'' \arrow[r] \arrow[d] & B'' & \\
 & 0 & &
\end{tikzcd}$$

with exact rows and columns. Show that $A'' \to B''$ is injective.

*Exercise* 2.21. Let $R$ be a ring. Consider a commutative diagram of $R$-modules with exact columns:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A' \arrow[r] \arrow[d] & B' \arrow[r] \arrow[d] & C' \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A'' \arrow[r] \arrow[d] & B'' \arrow[r] \arrow[d] & C'' \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd}$$

Prove that if the first two (resp. the last two) rows are exact, then so is the last (resp. the first) row.

*Hard exercise* 2.22. Let $R$ be a ring. An $R$-module $M$ is *finitely presented* if there exists an exact sequence of the form $F_1 \to F_2 \to M \to 0$, where $F_1$ and $F_2$ are free $R$-modules of finite rank. Let $M$ be a finitely presented $R$-module and let $f : R^n \to M$ be a surjective homomorphism of $R$-modules, where $n$ is a nonnegative integer. Show that $\ker f$ is finitely generated.

# Chapter 3

# Finiteness properties

So far we have mostly considered arbitrary (commutative) rings, and modules over these. To obtain deeper and more striking results, it will often be necessary to impose finiteness conditions on the objects under consideration, mostly using so-called “chain conditions”. This will lead to the notions of Noetherian and Artinian rings and modules; perhaps surprisingly, it is more convenient to treat modules first.

## 3.1 Modules

Let $(\Sigma, \leq)$ be a partially ordered set (or *poset* in short), which means that $\leq$ is a relation on $\Sigma$ which is both reflexive and transitive, and such that $x \leq y$ and $y \leq x$ together imply $x = y$.

**Lemma 3.1.1.** *The following conditions are equivalent for the poset* $(\Sigma, \leq)$*:*

*(1) every increasing sequence* $x_1 \leq x_2 \leq \cdots$ *in* $\Sigma$ *becomes stationary (eventually constant);*

*(2) every non-empty subset of* $\Sigma$ *has a maximal element.*

**Definition 3.1.2.** Let $R$ be a ring, let $M$ be an $R$-module and let $\Sigma$ be the set of submodules of $M$.

Assume first that $\Sigma$ is ordered by the inclusion relation $\subseteq$. Then condition (1) above is called the *ascending chain condition* (acc). If $M$ satisfies the acc, then $M$ is said to be a *Noetherian $R$-module*.

If, on the other hand, $\Sigma$ is ordered by the containment relation $\supseteq$, then condition (1) above is called the *descending chain condition* (dcc). If $M$ satisfies the dcc, then $M$ is said to be an *Artinian $R$-module*.

**Example 3.1.3.** Any finite abelian group is both Noetherian and Artinian as a $\mathbf{Z}$-module, simply because it has only finitely many submodules. On the other hand, $\mathbf{Z}$ (seen as a $\mathbf{Z}$-module) is Noetherian, but not Artinian – indeed, the descending chain of ideals $(2) \supsetneq (4) \supsetneq (8) \supsetneq \cdots$ never becomes stationary.

**Example 3.1.4.** Let $p$ be a prime, and let $G \subseteq \mathbf{Q}/\mathbf{Z}$ be the subgroup consisting of all elements whose order is a power of $p$. Then $G$ has exactly one subgroup $G_n$ of order $p^n$, for each $n \geq 0$; these subgroups satisfy $G_0 \subsetneq G_1 \subsetneq G_2 \subsetneq \cdots$. Hence $G$ (seen as a $\mathbf{Z}$-module) is Artinian, but not Noetherian.

**Example 3.1.5.** The ring $\mathbf{C}[X]$, seen as a module over itself, is Noetherian, but not Artinian.

Let $R = \mathbf{C}[X_1, X_2, \cdots]$ be the ring of polynomial expressions with complex coefficients in countably infinitely many variables. Then $R$, seen as an $R$-module over itself, is neither Noetherian, nor Artinian: the increasing sequence $(X_1) \subsetneq (X_1, X_2) \subsetneq (X_1, X_2, X_3) \subsetneq \cdots$ shows that the acc does not hold, whereas the decreasing sequence $(X_1) \supsetneq (X_1^2) \supsetneq (X_1^3) \supsetneq \cdots$ shows that the dcc does not hold.

**Proposition 3.1.6.** *Let $R$ be a ring and let $M$ be an $R$-module. Then $M$ is a Noetherian $R$-module if and only if every submodule of $M$ is finitely generated.*

*Proof.* Assume that $M$ is Noetherian. Let $N$ be a submodule of $M$ and let $\Sigma$ be the set of all finitely generated submodules of $N$. Let $N_0$ be a maximal element of $\Sigma$. If $x \in N \setminus N_0$, then $N_0 + Rx$ is an element of $\Sigma$ which strictly contains $N_0$. This contradicts the maximality assumption, whence $N_0 = N$.

Conversely, assume that every submodule of $M$ is finitely generated. Let $M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots$ be an ascending chain of submodules. Then $N = \bigcup_{k \geq 0} M_k$ is again a submodule, hence finitely generated by assumption; it follows that $N = M_k$ for $k$ sufficiently large, which is exactly what we want. □

**Proposition 3.1.7.** *Let $R$ be a ring, and let $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ be an exact sequence of $R$-modules. Then $M_2$ is a Noetherian $R$-module if and only if so are both $M_1$ and $M_3$.*

*The same result remains true if "Noetherian" is replaced by "Artinian".*

In other words: the class of Noetherian (resp. Artinian) $R$-modules is closed under taking submodules, quotients and extensions (for the latter notion, see the discussion following Definition 2.4.3). In particular, direct sums of finitely many Noetherian (resp. Artinian) $R$-modules remain Noetherian (resp. Artinian).

*Proof.* If $M_2$ is Noetherian, any ascending chain of submodules of $M_1$ (resp. $M_3$) yields such a chain in $M_2$, simply by taking images under $f$ (resp. inverse images under $g$), and hence becomes stationary.

Conversely, if $M_1$ and $M_3$ are Noetherian, let us consider an ascending chain $(N_i)_{i \geq 1}$ of submodules in $M_2$. Then $(f^{-1}(N_i))_{i \geq 1}$ is such a chain in $M_1$, and $(g(N_i))_{i \geq 1}$ is such a chain in $M_3$. Since both of these chains eventually become stationary, so must $(N_i)_{i \geq 1}$, as desired.

The statement about Artinian $R$-modules can be proven in a completely similar way. □

**Definition 3.1.8.** Let $R$ be a ring. Then $R$ is said to be *Noetherian* (resp. *Artinian*) if $R$ is Noetherian (resp. Artinian) when considered as an $R$-module over itself.

Of course a ring $R$ is Noetherian if and only if every ideal in $R$ is finitely generated. For example, every PID is Noetherian, simply because in a PID, every ideal is generated by a single element; on the other hand, the ring $\mathbf{C}[X_1, X_2, \cdots]$ from Example 3.1.5 is not Noetherian.

The following results are straightforward consequences of Proposition 3.1.7:

**Corollary 3.1.9.** *Let $R$ be a Noetherian (resp. Artinian) ring, and let $M$ be a finitely generated $R$-module. Then $M$ is a Noetherian (resp. Artinian) $R$-module.*

**Corollary 3.1.10.** *Any quotient of a Noetherian (resp. Artinian) ring is again Noetherian (resp. Artinian).*

*Proof.* Let $R$ be a Noetherian (resp. Artinian) ring, and let $I$ be an ideal in $R$. Then $R/I$ is Noetherian (resp. Artinian) as an $R$-module by Proposition 3.1.7, and therefore also as an $R/I$-module (why?). □

Recall from Exercise 2.2 that an $R$-module $M$ is said to be *simple* if $M$ has exactly two submodules, namely 0 and itself.

**Definition 3.1.11.** Let $R$ be a ring. A *composition series* of an $R$-module $M$ is a chain of submodules

$$M = M_0 \supsetneq M_1 \supsetneq \cdots \supsetneq M_n = 0$$

such that $M_{i-1}/M_i$ is simple for $i = 1, \cdots, n$. We call $n$ the *length* of the composition series.

We will cite the following result as a "black box", i.e., we will not cover the proof; those who want to know more are welcome to consult [Rotman, §7.5] for background and detailed proofs.

**Theorem 3.1.12.** *Let $R$ be a ring. An $R$-module $M$ has a composition series if and only if $M$ is both Noetherian and Artinian. If this is the case, we say that $M$ is an $R$-*module of finite length*; any two composition series then have the same length, simply called the* length of $M$*, denoted by $\ell(M)$.*

*Moreover, the* Jordan–Hölder theorem *holds: if $(M_i)_{0\leq i\leq n}$ and $(M_i')_{0\leq i\leq n}$ are composition series, there is a bijective correspondence between the sets of quotients $(M_{i-1}/M_i)_{1\leq i\leq n}$ and $(M_{i-1}'/M_i')_{1\leq i\leq n}$ such that corresponding quotients are isomorphic as $R$-modules.*

In other words: for a module which is Noetherian and Artinian, the length is a basic invariant, but the collection of "simple pieces" (quotients appearing in a composition series) gives much finer information.

*Quick question* 3.1.13. It is not too difficult to see that if an $R$-module $M$ is both Noetherian and Artinian, then $M$ must have a composition series; prove this yourself. (The rest of Theorem 3.1.12 is harder.)

**Definition 3.1.14.** Let $R$ be a ring and let $\lambda$ be a $\mathbf{Z}$-valued function defined on a class $\mathcal{C}$ of $R$-modules. Then $\lambda$ is said to be an *additive* function on $\mathcal{C}$ if whenever $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow 0$ is a short exact sequence of $R$-modules with terms in $\mathcal{C}$, the equality $\lambda(M_1) - \lambda(M_2) + \lambda(M_3) = 0$ holds.

*Remark* 3.1.15. In fact the condition in the above definition is equivalent to the following one: given any exact sequence $0 \to M_1 \to M_2 \to \cdots \to M_n \to 0$ with terms in $\mathcal{C}$, we have $\sum_{i=1}^{n}(-1)^i\lambda(M_i) = 0$. This follows rather easily from the fact that any exact sequence can be split up into a number of short exact sequences, in the sense of Lemma 2.4.5; we leave the details to the reader.

**Example 3.1.16.** Let $k$ be a field, and let $\mathcal{C}$ be the class of all finite dimensional vector spaces over $k$. Then undergraduate linear algebra tells us that the function $\mathcal{C} \to \mathbf{Z} : V \mapsto \dim V$ is additive on $\mathcal{C}$.

Since dimension and length coincide for vector spaces (why?), we can generalise Example 3.1.16:

**Proposition 3.1.17.** *Let $R$ be a ring, and let us denote by $\mathcal{C}$ the class of all $R$-modules of finite length. Then the "length function" $\mathcal{C} \to \mathbf{Z} : M \mapsto \ell(M)$ is additive on $\mathcal{C}$.*

*Proof.* Let $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ be a short exact sequence. Given composition series of both $M_1$ and $M_3$, consider the image under $f$ and the inverse image under $g$ of these series; these fit together to give a composition series of $M_2$. In particular, $\ell(M_2) = \ell(M_1) + \ell(M_3)$. □

## 3.2 Rings

In the previous paragraph, we studied chain conditions for modules over rings, and in a sense the situation was rather "symmetric": Noetherian and Artinian modules behaved similarly in many ways. Strikingly, this symmetry disappears in the case of rings.

Noetherian rings are an extremely important and wide class of rings in commutative algebra, algebraic geometry, et cetera: they are very general and moreover they tend to "reproduce" themselves under various familiar operations (as we shall see). Artinian rings, on the other hand, while important, are a much more restrictive class; in a way they are the simplest kind of rings after fields. In particular, they are necessarily... Noetherian!

The following theorem, known as *Hilbert's basis theorem*, is a deep and central result in commutative algebra, even if the proof is remarkably simple and elementary:

**Theorem 3.2.1.** *If $R$ is a Noetherian ring, then so is $R[X]$.*

Let us introduce some notation needed in the proof. Given $f \in R[X]$, we denote by $\mathrm{lc}(f)$ the *leading coefficient* of $f$: this is simply the coefficient of the highest degree term appearing in $f$. Given an arbitrary subset $S \subseteq R[X]$, we define $\mathrm{LC}(S) = \{\mathrm{lc}(f) : f \in S\} \cup \{0\}$.

*Quick question* 3.2.2. Check that if $I$ is an ideal in $R[X]$, then $\mathrm{LC}(I)$ is an ideal in $R$.

*Proof of Theorem 3.2.1.* Let $I$ be an ideal in $R[X]$. Since $R$ is Noetherian, the ideal $\mathrm{LC}(I)$ is finitely generated. This means that there exist $f_1, \cdots, f_r \in I$ such that $\mathrm{LC}(I) = (\mathrm{lc}(f_1), \cdots, \mathrm{lc}(f_r))$.

Let $d_i = \deg f_i$ and set $d = \max_{1 \leq i \leq r} d_i$. Let $M = \langle 1, X, \cdots, X^{d-1} \rangle$ be the $R$-submodule of $R[X]$ consisting of all polynomials of degree strictly smaller than $d$. Since $M$ is finitely generated by construction, it is Noetherian as an $R$-module by Corollary 3.1.9. Therefore $M \cap I$ is finitely generated as an $R$-module, say by elements $g_1, \cdots, g_s \in I$. We claim that $I = (f_1, \cdots, f_r, g_1, \cdots, g_s)$ in $R[X]$; this claim clearly implies the desired result (namely that $I$ must be finitely generated).

It suffices to prove that $I \subseteq (f_1, \cdots, f_r, g_1, \cdots, g_s)$ (the other inclusion being obvious). Let $h \in I$, and set $e = \deg h$. If $e < d$, then clearly $h \in \langle g_1, \cdots, g_s \rangle$; hence there is nothing to prove. Otherwise, we know that there exist $c_1, \cdots, c_r \in R$ such that $\mathrm{lc}(h) = c_1 \mathrm{lc}(f_1) + \cdots + c_r \mathrm{lc}(f_r)$. It is then clear that $h - \sum_{j=1}^{r} c_j X^{e-d_j} f_j$ has degree strictly smaller than $e$ and still belongs to $I$. Repeating this procedure, we obtain $\gamma_1, \cdots, \gamma_r \in R[X]$ such that $h - \sum_{j=1}^{r} \gamma_j f_j$ has degree strictly smaller than $d$, hence belongs to $M \cap I \subseteq \langle g_1, \cdots, g_s \rangle$. This proves that $h \in (f_1, \cdots, f_r, g_1, \cdots, g_s)$, as required. □

The following result is an easy corollary of Theorem 3.2.1 and Corollary 3.1.10:

**Corollary 3.2.3.** *Let $R$ be a Noetherian ring, and let $I$ be an ideal of the polynomial ring $R[X_1, \cdots, X_n]$. Then $R[X_1, \cdots, X_n]/I$ is again a Noetherian ring.*

As we will see in Chapter 6, rings of the form $R[X_1, \cdots, X_n]/I$ are called *$R$-algebras of finite type*. Such rings are the essential building blocks of much of algebraic geometry, and they are all Noetherian thanks to Theorem 3.2.1 (at least if the "base ring" $R$ is Noetherian).

While Noetherian domains can fail the unique factorisation property – indeed, Example 1.3.13 shows that $\mathbf{Z}[X]/(X^2+5)$ is not a UFD, and this ring is Noetherian by Corollary 3.2.3 – they do satisfy the first half of Definition 1.3.11, namely the existence of factorisations in irreducible elements:

**Proposition 3.2.4.** *Let $R$ be a Noetherian domain. Every non-zero element $r \in R$ which is not a unit factors as a product of irreducibles: there exist irreducible elements $p_1, \cdots, p_m$ such that $r = p_1 \cdots p_m$.*

*Proof.* Let $r \in R$ be a non-zero element which is not a unit. If $r$ does not factor, then $r$ is certainly not itself irreducible; hence there exist $r_1, s_1 \in R$ such that $r = r_1 s_1$, where $(r) \subsetneq (r_1)$ and $(r) \subsetneq (s_1)$. If both $r_1$ and $s_1$ factor into irreducibles, then so does $r$; hence we may assume that $r_1$ does not factor. Repeating the argument yields an infinitely increasing chain $(r) \subsetneq (r_1) \subsetneq (r_2) \subsetneq \cdots$, contradiction! □

**Example 3.2.5.** In the ring $\overline{\mathbf{Z}}$ of algebraic integers (roots of monic polynomials with integral coefficients), not a single non-zero element which is not a unit admits a factorisation into irreducibles; in particular, this ring is certainly not Noetherian. On the other hand, the ring $\mathbf{Z}[X_1, X_2, \cdots]$ is not Noetherian either, and yet every non-zero non-unit element admits a factorisation into irreducibles in this ring!

Let us now turn to the study of Artinian rings. We will see that in a way, there are only "very few" Artinian rings. For example, the following result implies (why?) that Artinian domains must be fields:

**Proposition 3.2.6.** *In an Artinian ring, every prime ideal is maximal.*

*Proof.* Let $R$ be an Artinian ring, and let $\mathfrak{p}$ be a prime ideal in $R$. Our goal is to show that if $f \notin \mathfrak{p}$, then $(\mathfrak{p}, f) = R$. Since $R$ is Artinian, the descending chain $(\mathfrak{p}, f) \supseteq (\mathfrak{p}, f^2) \supseteq \cdots$ must be stationary; in particular, there exists some $n \geq 1$ such that $f^n = f^{n+1} g + h$ for some $g \in R$ and $h \in \mathfrak{p}$. Since $f^n(1 - fg) \in \mathfrak{p}$ and $f \notin \mathfrak{p}$, we have $1 - fg \in \mathfrak{p}$, in other words: $(\mathfrak{p}, f) = R$, as required. □

We will need the following lemma, the proof of which is left to the reader:

**Lemma 3.2.7.** *Let $R$ be a ring, and let $\mathfrak{m}_1, \mathfrak{m}_2, \cdots, \mathfrak{m}_n$ be a collection of distinct maximal ideals in $R$. Then $\mathfrak{m}_1\mathfrak{m}_2 \cdots \mathfrak{m}_n$ is a* proper *submodule of the $R$-module $\mathfrak{m}_1\mathfrak{m}_2 \cdots \mathfrak{m}_{n-1}$.*

As an easy corollary, we obtain the following intermediate result:

**Corollary 3.2.8.** *An Artinian ring has only finitely many maximal ideals.*

*Proof.* Indeed, assume that $\mathfrak{m}_1, \mathfrak{m}_2, \cdots$ is an infinite sequence of distinct maximal ideals in an Artinian ring $R$. Then $\mathfrak{m}_1 \supsetneq \mathfrak{m}_1\mathfrak{m}_2 \supsetneq \mathfrak{m}_1\mathfrak{m}_2\mathfrak{m}_3 \supsetneq \cdots$ is a non-stationary descending chain – contradiction! □

**Proposition 3.2.9.** *In an Artinian ring $R$, the nilradical $\mathfrak{n}_R$ is nilpotent: $\mathfrak{n}_R^s = 0$ for some $s \geq 1$.*

Note that it is not hard to prove that the nilradical is nilpotent in a Noetherian ring (Exercise 3.4), but of course we cannot use this fact here: we have not proven yet that Artinian rings are Noetherian!

*Proof.* We sketch the proof and leave the details to the reader. The descending chain $\mathfrak{n}_R \supseteq \mathfrak{n}_R^2 \supseteq \cdots$ must stabilise, so let $I = \mathfrak{n}_R^s = \mathfrak{n}_R^{s+1} = \cdots$ for $s \gg 0$. We claim that $I = 0$. If $I \neq 0$, let $\Sigma$ be the set of ideals $I'$ such that $II' \neq 0$. Then $\Sigma$ is non-empty since $I \in \Sigma$. Let $J$ be a minimal element of $\Sigma$.

Choose $x \in J$ such that $xI \neq 0$, then $J = (x)$ by minimality of $J$. Since $(xI)I = xI^2 = xI \neq 0$, we have $xI \subseteq (x)$, and therefore $xI = (x)$, again by minimality. Hence $x = xy$ for some $y \in I$, and therefore $x = 0$ since $y \in \mathfrak{n}_R$. This contradicts the choice of $x$, and therefore $I = 0$. □

The following technical result is the last bit of information needed for the proof of Theorem 3.2.11.

**Lemma 3.2.10.** *Let $R$ be a ring in which the zero ideal can be written as a product of maximal ideals (with repetitions allowed). Then $R$ is Noetherian if and only if $R$ is Artinian.*

*Proof.* Let $\mathfrak{m}_1, \cdots, \mathfrak{m}_n$ be maximal ideals in $R$ (with repetitions allowed!) such that $\mathfrak{m}_1 \cdots \mathfrak{m}_n = 0$. For each $i = 1, \cdots, n$, the quotient $M_i = \mathfrak{m}_1 \cdots \mathfrak{m}_{i-1}/\mathfrak{m}_1 \cdots \mathfrak{m}_i$ is a vector space over the field $R/\mathfrak{m}_i$; hence $M_i$ is Noetherian as an $R$-module if and only if $M_i$ is Artinian as an $R$-module (see Exercise 3.2).

For each $i = 1, \cdots, n$, Proposition 3.1.7 applied to the short exact sequence

$$0 \longrightarrow \mathfrak{m}_1 \cdots \mathfrak{m}_i \longrightarrow \mathfrak{m}_1 \cdots \mathfrak{m}_{i-1} \longrightarrow M_i \longrightarrow 0$$

says that the $R$-modules $\mathfrak{m}_1 \cdots \mathfrak{m}_i$ and $M_i$ are Noetherian (respectively Artinian) if and only if the $R$-module $\mathfrak{m}_1 \cdots \mathfrak{m}_{i-1}$ is Noetherian (respectively Artinian). Now a descending induction on $i$, starting with $M_n = \mathfrak{m}_1 \cdots \mathfrak{m}_{n-1}$, then shows that for each $i$, the $R$-module $\mathfrak{m}_1 \cdots \mathfrak{m}_i$ is Noetherian if and only if it is Artinian; when $i = 0$, this statement says precisely that $R$ is Noetherian if and only if it is Artinian. □

The following important result is due to Akizuki, Hopkins and Levitzki (independently of each other). The proof sketched here may seem short, but of course we have done a lot of preparatory work by now.

**Theorem 3.2.11.** *An Artinian ring is also Noetherian.*

*Proof.* Let $R$ be an Artinian ring. By Lemma 3.2.8, $R$ has only finitely many maximal ideals $\mathfrak{m}_1, \cdots, \mathfrak{m}_n$. Since prime ideals in $R$ are maximal (Proposition 3.2.6), we have $\mathfrak{m}_1 \cdots \mathfrak{m}_n \subseteq \mathfrak{n}_R$ by Proposition 1.2.17. Since $\mathfrak{n}_R$ is nilpotent by Proposition 3.2.9, the result now follows immediately from Lemma 3.2.10. □

*Remark* 3.2.12. The statement of Theorem 3.2.11 remains true for non-commutative rings, but the proof is quite a bit harder; we refer the interested reader to [Rotman, §8.3] for (non-compulsory) details.

*Quick question* 3.2.13. Let $R$ be an Artinian ring and let $M$ be a finitely generated $R$-module. Deduce from Theorem 3.2.11 that $M$ is also Noetherian as an $R$-module (i.e. $M$ has finite length).

**Example 3.2.14.** The rings $\mathbf{C}[t]/(t^n)$ are Artinian for any $n \geq 1$.

# Exercises

*Easy exercise* 3.1. Let $R$ be a ring such that $R[X]$ is Noetherian. Must $R$ necessarily be Noetherian?

*Easy exercise* 3.2. Let $k$ be a field and let $V$ be a $k$-vector space ($k$-module). Show that the four following conditions are equivalent: $V$ is finite-dimensional; $V$ has finite length; $V$ is Artinian; $V$ is Noetherian.

*Easy exercise* 3.3. Prove that in an Artinian local ring, every element is either a unit or nilpotent.

*Exercise* 3.4. Prove that in a Noetherian ring $R$, the nilradical $\mathfrak{n}_R$ is nilpotent: $\mathfrak{n}_R^s = 0$ for some $s \geq 1$.

*Exercise* 3.5. Let $R$ be a ring and let $f : M \to M$ be an $R$-module endomorphism.

(a) Assume that $M$ is Noetherian and that $f$ is surjective. Show that $f$ is an isomorphism.
*For a more general result, see Exercise 2.16. However, the Noetherianity assumption allows for a simpler proof avoiding Nakayama's lemma. Hint: consider the chain of submodules* $(\ker f^n)_{n\geq 1}$.

(b) Assume that $M$ is Artinian and that $f$ is injective. Show that $f$ is an isomorphism.

*Exercise* 3.6. Let $R$ be a Noetherian ring, and let $f = \sum_{n=0}^{\infty} a_n X^n \in R[\![X]\!]$. Prove that $f$ is nilpotent if and only if for each $n \geq 0$, the coefficient $a_n$ is nilpotent. (See also Exercise 1.11.)

*Exercise* 3.7. Let $R$ be a ring. Recall from Exercise 2.22 that an $R$-module $M$ is said to be *finitely presented* if there exists an exact sequence of the form $F_1 \to F_2 \to M \to 0$, where $F_1$ and $F_2$ are free $R$-modules of finite rank. Prove that if $R$ is Noetherian, then every finitely generated $R$-module is finitely presented, and show also that this conclusion fails for every ring $R$ which is not Noetherian.

*Exercise* 3.8. Let $R$ be a Noetherian local ring with maximal ideal $\mathfrak{m}$. Show that either $\mathfrak{m}^n \neq \mathfrak{m}^{n+1}$ for all $n \geq 1$, in which case $R$ is not Artinian, or $\mathfrak{m}^n = 0$ for some $n \geq 1$, in which case $R$ is Artinian. Also, give an example of an Artinian local ring in which not every ideal is principal.

*Exercise* 3.9. Let $R$ be a ring and let $M$ be an $R$-module. Prove $M$ is Noetherian as an $R$-module if and only if $M$ is a finitely generated $R$-module, and $R/\operatorname{Ann}_R M$ is a Noetherian ring.

*Hard exercise* 3.10. Let $R$ be a Noetherian ring. Prove that $R[\![X]\!]$ is Noetherian as well.

*Hard exercise* 3.11. Prove the *structure theorem* for Artinian rings: every Artinian ring can be written uniquely (up to isomorphism) as a finite direct product of Artinian local rings.

# Chapter 4

# Linear algebra over a PID

One theme of this course so far has been the generalisation of known concepts from linear algebra to the (much) wider setting of modules of general rings. It has become clear by now that phenomena such as torsion and non-free modules make life much harder in general than in your undergraduate linear algebra course. The point of this chapter is that one can still do linear algebra in a very satisfactory way over a PID, and that this even yields new insights on "traditional" linear algebra over fields!

## 4.1 Finitely generated modules over a PID

If $R$ is a PID, then every non-zero ideal (or equivalently, every $R$-submodule of $R$) is isomorphic to $R$ itself as an $R$-module. This observation may seem rather innocent, but in fact characterises principal ideal domains among arbitrary (commutative) rings, in the following sense.

*Quick question* 4.1.1. Prove that an arbitrary (commutative) ring $R$ is a PID *if and only if* every non-zero ideal of $R$ is free of rank 1 (i.e., isomorphic to $R$ itself) as an $R$-module.

The following result generalises our observation to all free modules of finite rank over a PID:

**Proposition 4.1.2.** *Let $R$ be a PID. Let $F$ be a free $R$-module of rank $n$. Let $M$ be a submodule of $F$. Then $M$ is free, of rank at most $n$.*

*More precisely, there exists a basis $(x_1, \cdots, x_n)$ of $F$ over $R$ and non-zero elements $a_1, \cdots, a_m \in R$, where $m \leq n$, such that $a_1 \mid a_2 \mid \cdots \mid a_m$ and such that $(a_1x_1, \cdots, a_mx_m)$ is a basis of $M$ over $R$.*

*Remark* 4.1.3. This result can be generalised to free $R$-modules of arbitrary (potentially infinite) rank; we leave the challenge of stating (and proving) such a generalisation to the motivated reader.

We will prove Proposition 4.1.2 using the following lemma.

**Lemma 4.1.4.** *Let $R$ be a PID and let $F$ be a free $R$-module of finite rank. If $M \subseteq F$ is a non-zero submodule, there exist $a \in R$ and $x \in F$, together with submodules $F_0 \subseteq F$ and $M_0 \subseteq M$, such that*

$$F = \langle x \rangle \oplus F_0, \quad M = \langle ax \rangle \oplus M_0.$$

Recall that $\langle x \rangle$ denotes the cyclic $R$-module generated by $x$; in this case, $\langle x \rangle$ is free of rank 1. The lemma therefore says that one can split off copies of $R$ from both $F$ and $M$ in a "compatible" way.

*Proof.* For every $\varphi \in F^\vee = \mathrm{Hom}_R(F, R)$ (see Exercise 2.7), the image $\varphi(M)$ is an ideal in $R$. The collection of all ideals of this form must have a maximal element (since $R$ is Noetherian), obtained from some specific $\alpha \in F^\vee$; since $M$ is non-zero, so is the principal ideal $\alpha(M)$. Let $a \in R$ be a generator of this ideal and let $y \in M$ be such that $\alpha(y) = a$.

We claim that $a$ divides $\varphi(y)$ for *all* $\varphi \in F^\vee$. Indeed, given some $\varphi \in F^\vee$, let $b$ be a gcd of $a$ and $\varphi(y)$, and let $r, s \in R$ such that $b = ra + s\varphi(y)$. Taking $\psi = r\alpha + s\varphi \in F^\vee$, we have $b \in \psi(M)$ and therefore $\alpha(M) = \psi(M)$ by maximality of $\alpha(M)$. In particular, $a$ divides $b$ and therefore also $\varphi(y)$.

Let us now identify $F$ with $R^n$ (for some $n \geq 1$) via some choice of basis, and let $y = (s_1, \cdots, s_n)$ with respect to this basis. Each $s_i$ is the image of $y$ under one of the coordinate projections $F \to R$ induced by the choice of basis. Hence $a$ divides $s_i$ for all $i$, i.e., there exists $r_i \in R$ such that $s_i = ar_i$ for $i = 1, \cdots, n$. Let $x = (r_1, \cdots, r_n)$, then clearly $y = ax$ by construction, and $\alpha(x) = 1$.

We now take $F_0 = \ker \alpha$ and $M_0 = M \cap F_0$. Any $z \in F$ may be written as $z = \alpha(z)x + (z - \alpha(z)x)$, where clearly $\alpha(z)x \in \langle x \rangle$ and $z - \alpha(z)x \in F_0$. Since moreover $\langle x \rangle \cap F_0 = 0$, we obtain $F = F_0 \oplus \langle x \rangle$. Similarly, if $z \in M$, it is easy to see that the decomposition $z = \alpha(z)x + (z - \alpha(z)x)$ provides a way to write $z$ as the sum of $\alpha(z)x \in \langle ax \rangle$ and $z - \alpha(z)x \in M_0$. As before, this leads to $M = \langle ax \rangle \oplus M_0$. □

*Proof of Proposition 4.1.2.* If $M = 0$, the result is trivial; otherwise, Lemma 4.1.4 yields $y_1 \in M$ and a submodule $M_1 \subseteq M$ such that $M = \langle y_1 \rangle \oplus M_1$. If $M_1 = 0$, we are done; otherwise, Lemma 4.1.4 applied to $M_1 \subseteq F$ yields $y_2 \in M$ and $M_2 \subseteq M$ such that $M = \langle y_1 \rangle \oplus \langle y_2 \rangle \oplus M_2$. If $M_2 \neq 0$, we continue as before; at each step, we obtain a decomposition $M = \langle y_1 \rangle \oplus \cdots \oplus \langle y_m \rangle \oplus M_m$. However, for any such decomposition, we must have $m \leq n$ (why?). Therefore the process must stop at some point with $M_m = 0$. This means that $M \cong R^m$, proving the first statement.

The second statement is trivial if $n = 1$. If $n \geq 2$, let us use the method from the proof of Lemma 4.1.4 to pick $x_1 \in F$ and $a_1 \in R$ such that $F = \langle x_1 \rangle \oplus F_1$ and $M = \langle a_1 x_1 \rangle \oplus M_1$ for certain submodules $F_1 \subseteq F$ and $M_1 \subseteq M$. Since we have just proven that $F_1$ is free as well (!) we can apply the same method to $M_1$ and $F_1$ to construct $a_2 \in R$ and $x_2 \in F_1$ such that $F_1 = \langle x_2 \rangle \oplus F_2$ and $M_1 = \langle a_2 x_2 \rangle \oplus M_2$. All we need to check is that $a_1 \mid a_2$, for an easy induction will then do the rest. Recall that $(a_1)$ is maximal among all ideals of the form $\varphi(M)$, where $\varphi \in F^\vee$. If we choose $\psi \in F^\vee$ such that $\psi(x_1) = \psi(x_2) = 1$ (why does such a $\psi$ exist?), we obtain $a_1 = \psi(a_1 x_1) \in \psi(M)$, hence $(a_1) = \psi(M)$ by the maximality assumption. Since $a_2 = \psi(a_2 x_2) \in \psi(M) = (a_1)$, we conclude that $a_1 \mid a_2$, as desired. □

The main result of this section is the following.

**Theorem 4.1.5.** *Let $R$ be a PID and let $M$ be a finitely generated $R$-module. There exist a unique integer $\rho \geq 0$ and a descending sequence of proper, non-zero ideals $(c_1) \supseteq \cdots \supseteq (c_k)$ of $R$ such that*

$$M \cong R^\rho \oplus R/(c_1) \oplus \cdots \oplus R/(c_k). \tag{4.1}$$

The integer $\rho$ in the above equality is called the *rank* of the $R$-module $M$. The generators $c_1, \cdots, c_k$, which are unique up to multiplication by a unit, are the *invariant factors* of $M$.

In the proof of Theorem 4.1.5, we will use the following elementary lemma (left to the reader).

**Lemma 4.1.6.** *Let $R$ be a PID and let $M$ be a finitely generated $R$-module. There exists a short exact sequence $0 \longrightarrow F_1 \longrightarrow F_2 \longrightarrow M \longrightarrow 0$, where $F_1$ and $F_2$ are free $R$-modules of finite rank.*

*Proof of Theorem 4.1.5.* Choose an exact sequence of the form $0 \longrightarrow F_1 \xrightarrow{\iota} F_2 \longrightarrow M \longrightarrow 0$ where $F_1$ and $F_2$ are free $R$-modules of finite rank. Let us apply Proposition 4.1.2 to the free $R$-module $F_2$ and its submodule $\iota(F_1)$: we know that there exist a basis $(x_1, \cdots, x_n)$ for $F_2$ and elements $a_1, \cdots, a_m \in R$, for some $m \leq n$, such that $(a_1 x_1, \cdots, a_m x_m)$ is a basis for $\iota(F_1)$. It is then clear that $M \cong F_2/\iota(F_1)$ is of the form described in the theorem. More precisely, if $a_1, \cdots, a_\ell$ are units while $a_{\ell+1}, \cdots, a_m$ are not, we take $k = m - \ell$, $c_j = a_{j+\ell}$ for $1 \leq j \leq k$, and $\rho = n - m$ to obtain the isomorphism (4.1). □

We can refine Theorem 4.1.5 a bit, as follows.

**Corollary 4.1.7.** *Let $R$ be a PID. Let $M$ be a finitely generated $R$-module. There exist a unique $\rho \in \mathbf{Z}_{\geq 0}$, prime elements $p_1, \cdots, p_\ell \in R$ and integers $a_1, \cdots, a_\ell > 0$ such that*

$$M \cong R^\rho \oplus R/(p_1^{a_1}) \oplus \cdots \oplus R/(p_\ell^{a_\ell}).$$

*Proof.* This follows immediately from Theorem 4.1.5, together with the Chinese remainder theorem. □

Note that the prime elements $p_1, \cdots, p_\ell$ of $R$ appearing in this statement need not be distinct or non-associated! The ideals $(p_i^{a_i})$ for $i = 1, \cdots, \ell$ are called the *elementary divisors* of the $R$-module $M$.

If $p \in R$ is a prime element, then the *p-primary component* of $M$, denoted by $M_p$, is defined to be the $R$-submodule of $M$ consisting of elements annihilated by some power of $p$. By Corollary 4.1.7, we know that $M_p$ is a direct sum of $R$-modules of the form $R/(p^n)$.

*Remark* 4.1.8. Corollary 4.1.7 applied to the case $R = \mathbf{Z}$ yields the fact that any finitely generated abelian group is isomorphic to an additive group of the form $\mathbf{Z}^\rho \oplus \mathbf{Z}/(p_1^{a_1}) \oplus \cdots \oplus \mathbf{Z}/(p_\ell^{a_\ell})$ for an integer $\rho \geq 0$, certain primes $p_1, \cdots, p_\ell$ and integers $a_1, \cdots, a_\ell > 0$. In particular, any *finite* abelian group is then necessarily isomorphic to a group of this type with $\rho = 0$.

Theorem 4.1.5 tells us in particular that any finitely generated module over a PID "splits" as the direct sum of a free part and a torsion part. This is certainly not true in more general situations, e.g. when $M$ is no longer finitely generated (see Exercise 4.16) or when $R$ is no longer a PID: if $R = \mathbf{C}[X, Y]$, then the maximal ideal $(X, Y)$ is a finitely generated $R$-module which is torsion free, but not free.

We have seen a concrete description of the free part using what we decided to call *the* rank of $M$, and of the torsion part using *the* invariant factors and elementary divisors of $M$. This was a sloppy choice of terminology, since we have not discussed uniqueness of these invariants – but we will do so now. That the rank of a finitely generated module over a PID is well-defined is easy to see [Rotman, Corollary 9.5.(ii)] (but see also Proposition 2.2.10). For the invariant factors and elementary divisors, things are more subtle; it is clear that it suffices to look at the case of torsion modules. We start with the following result.

**Lemma 4.1.9.** *Let $R$ be a PID and let $M_1$ and $M_2$ be finitely generated $R$-modules which are torsion. Then $M_1 \cong M_2$ if and only if for every prime element $p \in R$, we have $(M_1)_p \cong (M_2)_p$.*

This statement is proven in greater generality, without the assumption that $M_1$ and $M_2$ be finitely generated, in [Rotman, Proposition 9.11] (building on [Rotman, Theorem 9.10]).

**Proposition 4.1.10.** *Let $R$ be a PID and let $M_1$ and $M_2$ be finitely generated $R$-modules which are torsion. Then $M_1 \cong M_2$ if and only if $M_1$ and $M_2$ have the same list of elementary divisors.*

*Proof.* By Lemma 4.1.9, it suffices to prove the statement in the special case where both $M_1$ and $M_2$ are annihilated by some power of a prime element $p$, say by $p^N$. We know from Corollary 4.1.7 that both $M_1$ and $M_2$ can be written as direct sums of $R$-modules of the form $R/(p^k)$, where $0 < k \leq N$.

Let us first check that the number of direct summands is the same for $M_1$ and $M_2$. A computation to be done by the reader shows that the number of summands for $M_1$ is precisely the dimension of $M_1/pM_1$ as an $R/(p)$-vector space, and similarly for $M_2$. Since $M_1/pM_1 \cong M_2/pM_2$ are isomorphic as $R/(p)$-vector spaces, the numbers of summands in both direct sum decompositions must coincide.

Let us now check that the number of summands isomorphic to $R/(p)$ is the same for $M_1$ and $M_2$. By the previous observation, this is equivalent to saying that $M_1$ and $M_2$ have the same number of summands of the form $R/(p^n)$ for some $n \geq 2$. The reader will check that the latter number for $M_1$ is the dimension of $pM_1/p^2M_1$ as an $R/(p)$-vector space, and similarly for $M_2$. Since clearly $pM_1/p^2M_1 \cong pM_2/p^2M_1$ as $R/(p)$-vector spaces, this proves the claim about summands of type $R/(p)$.

For summands annihilated by higher powers of $p$, the result follows inductively via similar arguments, using the isomorphisms of $R/(p)$-vector spaces $p^n M_1/p^{n+1}M_1 \cong p^n M_2/p^{n+1}M_2$ for various $n$. □

**Proposition 4.1.11.** *Let $R$ be a PID. Let $M_1$ and $M_2$ be finitely generated $R$-modules which are torsion. Then $M_1 \cong M_2$ if and only if $M_1$ and $M_2$ have the same list of invariant factors.*

*Proof.* We sketch the argument and leave the details to the reader. If $(c_1) \supseteq (c_2) \supseteq \cdots \supseteq (c_k)$ is a sequence of invariant factors of $M_1$, we obtain the set of elementary divisors – which we already know to be unique, by the previous proposition – simply by taking the prime power factors of $c_1, c_2, \cdots, c_k$.

Conversely, $c_1 \mid c_2 \mid \cdots \mid c_k$ implies that we obtain $c_k$ (up to units) as the product of the largest prime powers among the elementary divisors, $c_{k-1}$ as the product of the largest prime powers among the remaining elementary divisors, and so on. Since the exact same recipe yields the invariant factors of $M_2$ starting from its elementary divisors, the result now follows from Proposition 4.1.10. □

The conclusion is that we have obtained a very satisfactory way to classify finitely generated modules over a PID: any such module is characterised completely up to isomorphism by its rank and its list of invariant factors, or equivalently, by its rank and its list of elementary divisors.

## 4.2 Matrices over a PID

In the previous section, we have reached a complete conceptual understanding of the structure of finitely generated modules over a PID. The goal of this section is to make all this more explicit using concrete computations with matrices. For a more extensive presentation, we refer to [Rotman, §9.4].

Let $R$ be an integral domain. An $(m \times n)$*-matrix over* $R$ is simply a matrix of size $(m \times n)$ with entries in $R$. An $(m \times m)$-matrix $A$ over $R$ is invertible if there exists an $(m \times m)$-matrix $B$ over $R$ such that $AB = BA = I_m$, where $I_m$ denotes the identity matrix of size $(m \times m)$. The traditional arguments from undergraduate linear algebra show that $A$ is invertible if and only if $\det A \in R^\times$.

The elementary $(m \times m)$-matrices over $R$ are the following:

- $F_{ij}$, the matrix obtained from $I_m$ by exchanging rows $i$ and $j$;
- $G_i(u)$, the matrix obtained from $I_m$ by multiplying row $i$ with $u \in R^\times$;
- $H_{ij}(r)$, the matrix obtained from $I_m$ by adding $r$ times row $j$ to row $i$, where $r \in R$.

These matrices are invertible, because their determinants are units; note that $\det G_i(u) = u$ is a unit by assumption. The elementary row operations are those which transform an $(m \times n)$-matrix $A$ into

- $F_{ij}A$, the matrix obtained from $A$ by exchanging rows $i$ and $j$;
- $G_i(u)A$, the matrix obtained from $A$ by multiplying row $i$ with $u \in R^\times$;
- $H_{ij}(r)A$, the matrix obtained from $A$ by adding $r$ times row $j$ to row $i$, where $r \in R$.

Similarly, the elementary column operations are those which transform an $(n \times m)$-matrix $A$ into

- $AF_{ij}$, the matrix obtained from $A$ by exchanging columns $i$ and $j$;
- $AG_i(u)$, the matrix obtained from $A$ by multiplying column $i$ with $u \in R^\times$;
- $AH_{ij}(r)$, the matrix obtained from $A$ by adding $r$ times column $i$ to column $j$, where $r \in R$.

**Definition 4.2.1.** Let $R$ be an integral domain. Two $(m \times n)$-matrices $A$ and $B$ over $R$ are *$R$-equivalent* if there exist an invertible $(m \times m)$-matrix $P$ and an invertible $(n \times n)$-matrix $Q$ such that $B = PAQ$. Moreover, $A$ and $B$ are *Gaussian $R$-equivalent* if $A$ can be transformed into $B$ via elementary operations.

*Quick question* 4.2.2. Check that $R$-equivalence and Gaussian $R$-equivalence define equivalence relations on the set of $(m \times n)$-matrices over $R$, and that Gaussian $R$-equivalent matrices are also $R$-equivalent.

**Theorem 4.2.3.** *Let $R$ be a PID, and let $A$ be an $(m \times n)$-matrix over $R$.*

*Then $A$ is $R$-equivalent to a matrix of the form*

$$\begin{pmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_\ell & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{4.2}$$

*where $\ell \le \min(m, n)$, and where $d_1, d_2, \cdots, d_\ell$ are non-zero elements of $R$ satisfying $d_1 \mid d_2 \mid \cdots \mid d_\ell$. Moreover $\ell$ is uniquely determined, and $d_1, d_2, \cdots, d_\ell$ are unique up to multiplication by units.*

*A matrix of type (4.2) is said to be in* Smith normal form.

In fact we have already proven this theorem in §4.1 – we sketch the argument and leave the details to the reader. Let $F = R^m$, and let $M$ be the image of the $R$-module homomorphism $m_A : R^n \to R^m$ which maps a column vector $v \in R^n$ to $Av \in R^m$. Then Proposition 4.1.2, applied to the free $R$-module $F$ and the submodule $M$, says that after a suitable change of basis, the matrix of $m_A$ with respect to the new bases becomes of type (4.2). The uniqueness part of Theorem 4.2.3 follows from Proposition 4.1.11.

We will now reprove the first part of Theorem 4.2.3 using an argument of algorithmic nature. We start with a technical lemma which allows us to approach the problem inductively (proof left to the reader).

**Lemma 4.2.4.** *In order to prove the first part (existence) of Theorem 4.2.3, it is sufficient to prove that $A$ is $R$-equivalent to an $(m \times n)$-matrix $B$ over $R$ of the form*

$$B = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{pmatrix} \tag{4.3}$$

*where $d_1 \mid b_{ij}$ whenever $2 \le i \le m$ and $2 \le j \le n$.*

Let us therefore focus on proving that any given matrix $A$ over $R$ is $R$-equivalent to a matrix $B$ of type 4.3. We will first prove this claim in the special case where $R$ is a Euclidean domain; in this case, we can actually prove the stronger statement that $A$ is *Gaussian* $R$-equivalent to such a matrix $B$.

*Proof in the case where $R$ is a Euclidean domain.* Let $\varphi : R \setminus \{0\} \to \mathbf{Z}_{\ge 0}$ be a degree function as in Definition 1.3.6. Let $A = (a_{ij})_{1 \le i \le m,\, 1 \le j \le n}$ be an $(m \times n)$-matrix over $R$. If $A = 0$, there is nothing to prove; otherwise, after some exchanges of rows and columns if necessary, we may safely assume that $a_{11} \neq 0$. It suffices to show that $A$ is Gaussian $R$-equivalent to either a matrix of type (4.3), or a matrix

$$A' = \begin{pmatrix} a'_{11} & \cdots & a'_{1n} \\ \vdots & \ddots & \vdots \\ a'_{m1} & \cdots & a'_{mn} \end{pmatrix} \quad \text{with } a'_{11} \neq 0 \text{ and } \varphi(a'_{11}) < \varphi(a_{11}). \tag{4.4}$$

Indeed, we cannot have an infinite decreasing sequence $\varphi(a_{11}) > \varphi(a'_{11}) > \varphi(a''_{11}) > \cdots$ so at some point we will then have to reach a matrix of type (4.3). To prove this claim, we distinguish three cases.

(a) Assume that the first row contains an element $a_{1j}$ which is not divisible by $a_{11}$. Let $a_{1j} = qa_{11} + r$ where $r \neq 0$ and $\varphi(r) < \varphi(a_{11})$. Then the matrix $AH_{1j}(-q)F_{1j}$ is easily seen to be of type (4.4).

(b) Assume that the first column contains an element $a_{i1}$ not divisible by $a_{11}$, and let $a_{i1} = qa_{11} + r$ where $r \neq 0$ and $\varphi(r) < \varphi(a_{11})$. Then the matrix $F_{i1}H_{i1}(-q)A$ is again of type (4.4).

(c) If $a_{11}$ divides all elements in the first row and the column, we subtract suitable multiples of the first row (resp. column) from the other rows (resp. columns) to obtain a matrix of the form

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

If $a_{11}$ divides all other entries of this matrix, then we have reached a matrix of type (4.3) and hence we are done. If not, we add a suitable row of this matrix to the first row in order to obtain a matrix covered by case (a). (Of course we could also use a column operation and resort to case (b).)

This finishes the proof in the case where $R$ is a Euclidean domain. □

In the more general case when $R$ is only assumed to be a PID and not necessarily a Euclidean domain, it is not necessarily true that $A$ is *Gaussian* $R$-equivalent to a matrix of type (4.3). This is reflected by the fact that in our modified "algorithm" below used to show that $A$ is $R$-equivalent to a matrix of type (4.3), we use matrix operations which are no longer elementary row and column operations.

*Proof in the case where $R$ is a general PID.* Let us consider the map $\varphi : R \setminus \{0\} \to \mathbf{Z}_{\geq 0}$ which sends a non-zero $r \in R$ to the number of irreducible factors (counted with multiplicities) in the prime factorisation of $r$. (Note that if $r$ is a unit, then $\varphi(r) = 0$.) The map $\varphi$ is *not* a Euclidean degree function in general, since the second condition in Definition 1.3.6 need not hold. However, it is still possible to use $\varphi$ in an "algorithm" as above; let us explain how the recipe used for Euclidean domains needs to be modified.

The second condition in Definition 1.3.6 has only been used in (a) and (b) above, so it suffices to show that any matrix $A$ satisfying the assumptions of (a) or (b) is $R$-equivalent to a matrix of type (4.4).

Assume first that $a_{11} \nmid a_{12}$. Let $d$ be a greatest common divisor of $a_{11}$ and $a_{12}$, then $\varphi(d) < \varphi(a_{11})$. Let $x, y \in R$ such that $a_{11}x + a_{12}y = d$. Then the equality

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x & \frac{-a_{12}}{d} & 0 & \cdots & 0 \\ y & \frac{a_{11}}{d} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} d & a'_{12} & a'_{13} & \cdots & a'_{1n} \\ a'_{21} & a'_{22} & a'_{23} & \cdots & a'_{2n} \\ a'_{31} & a'_{32} & a'_{33} & \cdots & a'_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a'_{m1} & a'_{m2} & a'_{m3} & \cdots & a'_{mn} \end{pmatrix}$$

shows that $A$ is indeed $R$-equivalent to a matrix of type (4.4) – note that we are indeed multiplying $A$ on the right by an $(n \times n)$-matrix which is *invertible*, since its determinant is equal to 1.

Since we can reduce any matrix satisfying the assumptions of cases (a) and (b) to a matrix $A$ for which $a_{11} \nmid a_{12}$ simply by exchanging some rows and columns, we have now covered all possible cases. □

*For the curious...* Where does this proof fail when $R$ is a UFD, and no longer a PID?

Why should one even bother to (re)prove Theorem 4.2.3, given the fact that this result is a pretty straightforward consequence of the already proven Theorem 4.1.5? The point is that the method used to prove Theorem 4.2.3 above yields an explicit approach (amenable to computation) which is very useful in practice, for example to study the structure of abelian groups given by generators and relations.

**Example 4.2.5.** Consider the following matrix over the Euclidean domain $\mathbf{Z}$:

$$A = \begin{pmatrix} 3 & 6 & 3 \\ 3 & 4 & 5 \\ 6 & 5 & 6 \end{pmatrix}$$

Let us determine a $(3 \times 3)$-matrix over $\mathbf{Z}$ in Smith normal form which is Gaussian $\mathbf{Z}$-equivalent to this matrix. We abbreviate the algorithm used above a little bit, using some (entirely legal) shortcuts:

$$\begin{pmatrix} 3 & 6 & 3 \\ 3 & 4 & 5 \\ 6 & 5 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & 6 & 3 \\ 3 & 4 & 5 \\ 3 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 6 & 3 & 3 \\ 4 & 3 & 5 \\ 1 & 3 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 1 \\ 4 & 3 & 5 \\ 6 & 3 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 4 & -9 & 1 \\ 6 & -15 & -3 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 9 & 1 \\ 0 & 15 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & -3 & 15 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & 0 & 42 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 42 \end{pmatrix}$$

This computation can be interpreted in terms of the structure theory of abelian groups, as follows. Let $G$ be the abelian group with generators $a$, $b$ and $c$, and relations $3a + 6b + 3c = 0$, $3a + 4b + 5c = 0$ and $6a + 5b + 6c = 0$, in other words: $G$ is the cokernel of the homomorphism $\mathbf{Z}^3 \to \mathbf{Z}^3 : v \mapsto Av$. The Smith normal form of $A$ obtained above then shows that $G \cong \mathbf{Z}/42$.

For another example of this type of computation, we refer to [Rotman, Example 9.67].

*Quick question* 4.2.6. In Example 4.2.5, find an explicit element of order 42 in $G$ (in terms of $a$, $b$ and $c$).

## 4.3 Applications to endomorphisms of vector spaces

We will now cover some applications of our theory to "traditional" linear algebra involving vector spaces over fields, for example to the theory of the Jordan canonical form of a matrix. Our presentation will be somewhat brief; for a (much) more extensive treatment, see [Rotman, §9.2 – §9.3].

**Notation 4.3.1.** Let $k$ be a field, and let $V$ be a finite-dimensional vector space over $k$. Let $n = \dim V$, and let $\varphi : V \to V$ be a $k$-linear endomorphism of $V$. Let $R = k[X]$ and let $M$ be the abelian group $V$, seen as an $R$-module, where $X$ acts as $\varphi$ (in the sense of Example 2.1.11).

By the Cayley–Hamilton theorem (see Theorem 2.3.1), we know that $M$ is a finitely generated *torsion* module over $R$. Since $R$ is a PID, it follows from Theorem 4.1.5 that there exist a sequence of non-zero elements $c_1 \mid c_2 \mid \cdots \mid c_k$ in $R$ such that we have an isomorphism of $R$-modules

$$M \cong R/(c_1) \oplus \cdots \oplus R/(c_k).$$

Let $V_i$ be the subspace of $V$ which corresponds to the direct summand $R/(c_i)$, and let $\varphi_i = \varphi|_{V_i}$.

*Quick question* 4.3.2. Check that $V_i$ is an *invariant subspace* of $\varphi$, i.e., $\varphi(V_i) \subseteq V_i$ for $1 \leq i \leq k$.

By Quick question 4.3.2, we may consider $\varphi_i$ as an endomorphism of $V_i$, for $1 \leq i \leq k$. If we choose bases for the subspaces $V_1, \cdots, V_k$ of $V$, then it is clear that these bases will together yield a basis for $V$. The matrix of $\varphi$ with respect to this basis is of the form

$$\begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_k \end{pmatrix}, \tag{4.5}$$

to be understood as a block matrix, where $M_i$ is the matrix of $\varphi_i$ with respect to the basis for $V_i$ chosen previously, and where each $0$ represents a block of zeroes of the "correct" size.

**Proposition 4.3.3.** *With notation as above, if $c_i = X^m + a_{m-1}X^{m-1} + \cdots + a_1X + a_0 \in R$, and if $v \in V_i$ is a generator for the cyclic $R$-module $R/(c_i)$, then $\{v, \varphi(v), \cdots, \varphi^{m-1}(v)\}$ is a basis for $V_i$ as a $k$-vector space. Furthermore the matrix of $\varphi_i$ with respect to this specific basis is*

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{m-1} \end{pmatrix}. \tag{4.6}$$

**Definition 4.3.4.** With notation as above, the *rational canonical matrix* of $\varphi$ is the block matrix (4.5) which is obtained by putting the $k$ blocks of the type described in Proposition 4.3.3 together. We will see in a minute (see Remark 4.3.6 and Proposition 4.3.7) that this matrix is unique.

The *rational canonical form* of an $(n \times n)$-matrix $A$ over a field $k$ is the unique rational canonical matrix of the $k$-linear endomorphism $m_A : k^n \to k^n : v \mapsto Av$; this matrix is similar to $A$ by definition.

*Proof of Proposition 4.3.3.* We will only sketch the proof. If some $k$-linear combination of $v, \varphi(v), \cdots, \varphi^{m-1}(v)$ vanishes, then $R/(c_i)$ is annihilated by an element of $R$ of degree at most $m - 1$. This is a contradiction since $\deg c_i = m$. Hence $v, \varphi(v), \cdots, \varphi^{m-1}(v)$ are linearly independent.

To show that $v, \varphi(v), \cdots, \varphi^{m-1}(v)$ span $V_i$, we observe that any $w \in V_i$ can be written as $gv$ for some $g \in R$, since $R/(c_i)$ is a cyclic $R$-module. Writing $g = c_iq + r$ for $q, r \in R$ with $\deg r < \deg c_i = m$, we have $w = gv = rv$. It is easy to see that $rv$ is a $k$-linear combination of $v, \varphi(v), \cdots, \varphi^{m-1}(v)$.

Finally, the precise description of the matrix given by (4.6) follows easily by inspection. □

With notation as above, if the field $k$ is algebraically closed, then the non-zero prime elements of $R$ are exactly the linear polynomials $X - \lambda$, for $\lambda \in k$. Hence Corollary 4.1.7 implies that the $R$-module $M$ is isomorphic to a direct sum of the form $R/(X - \lambda_1)^{a_1} \oplus \cdots \oplus R/(X - \lambda_\ell)^{a_\ell}$, where $\lambda_1, \cdots, \lambda_\ell$ are not necessarily distinct. If $W_j$ denotes the ($\varphi$-invariant!) subspace of $V$ corresponding to the direct summand $R/(X - \lambda_j)^{a_j}$, then $W_j$ is $a_j$-dimensional. If $w$ generates the cyclic $R$-module $R/(X - \lambda_j)^{a_j}$, then arguments similar to the ones we have seen already show that $\{w, (\varphi - \lambda_j)w, \cdots, (\varphi - \lambda_j)^{a_j - 1}w\}$ is a basis of $W_j$, and that the matrix of $\varphi_j = \varphi|_{W_j}$ with respect to this basis is

$$\begin{pmatrix} \lambda_j & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda_j & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda_j & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_j & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda_j \end{pmatrix}, \tag{4.7}$$

a so-called *Jordan block* of dimension $a_j$.

**Definition 4.3.5.** With notation as above (assuming $k$ algebraically closed), the *Jordan canonical matrix* of $\varphi$ is the block matrix obtained by putting the $\ell$ Jordan blocks of type (4.7) together.

*Remark* 4.3.6. Our next result will imply among other things that both the rational canonical form and, if $k$ is algebraically closed, the Jordan canonical form of $\varphi$ are unique up to the order of the blocks; this justifies the (admittedly somewhat sloppy) use of the preposition *the* in Definitions 4.3.4 and 4.3.5.

Let us now connect these considerations to the theory covered in §4.2. The following proposition is extremely useful in practice: it allows one to study the rational canonical and Jordan canonical matrices of a vector space endomorphism via the calculation of the Smith normal form of an auxiliary matrix.

**Proposition 4.3.7.** *Let $k$ be a field and let $V$ be a finite dimensional vector space over $k$. Let $\varphi : V \to V$ be a $k$-linear endomorphism and let $A = (a_{ij})_{1 \leq i,j \leq n}$ be the $(n \times n)$-matrix of $\varphi$ with respect to some fixed basis of $V$, where $n = \dim V$. Consider the matrix $XI_n - A$, with entries in $R = k[X]$.*

*If $d_1 \mid d_2 \mid \cdots \mid d_n$ are the diagonal entries of a matrix over $R$ in Smith normal form which is $R$-equivalent to $XI_n - A$, and if $M$ is the $R$-module obtained from $V$ obtained by letting $X$ act through $\varphi$ (in the sense of Example 2.1.11), then we have an isomorphism of $R$-modules*

$$M \cong R/(d_1) \oplus R/(d_2) \oplus \cdots \oplus R/(d_n). \tag{4.8}$$

*Proof.* Let $\{v_1, \cdots, v_n\}$ be the chosen basis for $V$. Consider the "standard basis" $\{e_1, \cdots, e_n\}$ for $R^n$, and the surjective $R$-module homomorphism $\pi : R^n \to M$ given by $e_i \mapsto v_i$ for $i = 1, \cdots, n$.

Let $f_j = Xe_j - \sum_{i=1}^n a_{ij} e_i$, then $\pi(f_j) = 0$ for $j = 1, \cdots, n$. We claim that $f_1, \cdots, f_n$ generate the $R$-module $\ker \pi$. To prove this, let $N$ be the $R$-submodule of $\ker \pi$ generated by $f_1, \cdots, f_n$ and let $C$ be the $k$-vector space generated by $e_1, \cdots, e_n$; this is a subspace of $R^n$ seen as a $k$-vector space. We have $Xe_j \in N + C = \{s + t : s \in N, t \in C\}$ for $j = 1, \cdots, n$, whence $XC \subseteq N + C$. It follows that

$$X^2C \subseteq X(N + C) \subseteq XN + XC \subseteq XN + N + C \subseteq N + C$$

and hence (via induction)

$$X^kC \subseteq N + C \text{ for all } k \geq 1.$$

Since $e_1, \cdots, e_n$ generate the $R$-module $R^n$, we obtain $R^n = N + C$.

If $\gamma \in \ker \pi$, then $\gamma = \delta + \sum_{i=1}^n c_i e_i$ for certain $\delta \in N$ and $c_1, \cdots, c_n \in k$. Applying $\pi$ to both sides yields $0 = \sum_{i=1}^n c_i v_i$, and therefore $c_1 = \cdots = c_n = 0$; indeed, $v_1, \cdots, v_n$ are linearly independent elements of the $k$-vector space $C$. We conclude that $\gamma = \delta \in N$, i.e., $\ker \pi = N = \langle f_1, \cdots, f_n \rangle$.

Let us now consider the short exact sequence of $R$-modules

$$0 \longrightarrow \ker \pi \longrightarrow R^n \xrightarrow{\pi} M \longrightarrow 0,$$

which gives a concrete description of $M$ in terms of generators $e_1, \cdots, e_n$ and relations $f_1, \cdots, f_n$. The arguments from §4.1 and §4.2 show that the list of invariant factors of the $R$-module $M$ can be obtained from the computation of a Smith normal form of the matrix over $R$ which expresses $f_1, \cdots, f_n$ in terms of $e_1, \cdots, e_n$. In our setting, this matrix is nothing but $XI_n - A^t$, where $A^t$ denotes the transpose of $A$.

The result now follows from the elementary observation that $XI_n - A^t$ and its transpose $XI_n - A$ have the same Smith normal form over $R$ (up to units). Indeed, if $P$ and $Q$ are invertible $(n \times n)$-matrices over $R$ such that $P(XI_n - A^t)Q$ is in Smith normal form, in particular diagonal, then so is $Q^t(XI_n - A)P^t$, with the same list of entries on the diagonal. This finishes the proof. □

**Example 4.3.8.** Let $\varphi : \mathbf{C}^3 \to \mathbf{C}^3$ be the linear endomorphism given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \text{where } A = \begin{pmatrix} 1 & 1 & -1 \\ -2 & 3 & -1 \\ -2 & 1 & 1 \end{pmatrix}.$$

To find the rational and Jordan canonical forms of $A$, we need to transform the matrix $XI_3 - A$ into a matrix in Smith normal form over $R = \mathbf{C}[X]$, via elementary operations. This leads us from

$$\begin{pmatrix} X-1 & -1 & 1 \\ 2 & X-3 & 1 \\ 2 & -1 & X-1 \end{pmatrix} \quad \text{to} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix}$$

where $c = X^3 - 5X^2 + 8X - 4 = (X-1)(X-2)^2$.

Therefore the rational and Jordan canonical forms of $A$ are

$$\begin{pmatrix} 0 & 0 & 4 \\ 1 & 0 & -8 \\ 0 & 1 & 5 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

For another example of this type of computation, we refer to [Rotman, Example 9.66].

We end this chapter with a brief discussion of the notions of *minimal* and *characteristic* polynomial.

**Definition 4.3.9.** Let $k$ be a field and let $V$ be a finite dimensional $k$-vector space. If $\varphi : V \to V$ is a linear endomorphism, then a generator for the (principal) ideal $\{f \in k[X] : f(\varphi) = 0\}$ of $k[X]$ is called a *minimal polynomial* of $\varphi$; this polynomial is unique up to multiplication by an element of $k$.

**Proposition 4.3.10.** *With notation as in Proposition 4.3.7, $d_n$ is a minimal polynomial of $\varphi$. Moreover $\varphi$ is diagonalisable if and only if $d_n$ does not have any multiple roots.*

*Proof.* For the first statement we observe that the isomorphism (4.8) implies that for any $f \in R$, $f(\varphi)$ is the zero endomorphism of $V$ if and only if $d_1, \cdots, d_n$ divide $f$. Since $d_1 \mid d_2 \mid \cdots \mid d_n$, this condition is equivalent to $d_n$ dividing $f$, in other words: $d_n$ generates $\{f \in R : f(\varphi) = 0\}$.

For the second part, we note that if $d_n$ does not have any multiple roots, then the same holds true for $d_1, \cdots, d_{n-1}$; our previous results then imply that the Jordan canonical matrix of $\varphi$ is diagonal.

Conversely, if $\varphi$ is diagonalisable (with eigenvalues $\lambda_1, \cdots, \lambda_p$), then there certainly exists a basis of eigenvectors $\{v_1, \cdots, v_n\}$ for $V$. An easy computation shows that $(\varphi - \lambda_1 \mathrm{Id}_V) \cdots (\varphi - \lambda_p \mathrm{Id}_V)$ kills each of $v_1, \cdots, v_n$, and hence is the zero endomorphism. This means that $d_n$ divides $(X - \lambda_1) \cdots (X - \lambda_p)$, and *a fortiori* that $d_n$ does not have any multiple roots. □

If $V$ is a vector space of dimension $n$ over a field $k$, then the *characteristic polynomial* of a $k$-linear endomorphism $\varphi : V \to V$ is $P_\varphi = \det(XI_n - A) \in k[X]$, where $A$ is the matrix of $\varphi$ with respect to some choice of basis for $V$ (recall that the characteristic polynomial does not depend on the basis chosen).

**Proposition 4.3.11.** *With notation as in Proposition 4.3.7, the characteristic polynomial of $\varphi$ is $d_1 \cdots d_n$ (up to multiplication by an element of $k$). In particular, we have $P_\varphi(\varphi) = 0$ (Cayley–Hamilton).*

*Proof.* If $S$ and $T$ are invertible matrices over $k[X]$ such that $S(XI_n - A)T$ is in Smith normal form over $k[X]$, with diagonal entries $d_1 \mid d_2 \mid \cdots \mid d_n$, then clearly $P_\varphi = \det(XI_n - A)$ and $d_1 \cdots d_n$ differ by a factor $\det S \det T \in k[X]$. Since both $S$ and $T$ are invertible, it follows that $\det S \det T \in k$. □

## Exercises

*Easy exercise* 4.1. Let $n \geq 2$ be an integer. Let $F = \mathbf{Z}^2$ and let $M$ be the $\mathbf{Z}$-submodule of $F$ generated by $\{(1, n), (n, 1), (n, n)\}$. Make Proposition 4.1.2 explicit for the $\mathbf{Z}$-modules $F$ and $M$: find $x_1, x_2 \in F$ and $a_1, a_2 \in \mathbf{Z}$ such that $(x_1, x_2)$ is a basis for $F$ and $(a_1x_1, a_2x_2)$ is a basis for $M$.

*Easy exercise* 4.2. Give an example of a PID $R$ and an $R$-module $M$ which is torsion free, but not free. (It follows from Theorem 4.1.5, such an $M$ cannot be finitely generated as an $R$-module…)

*Easy exercise* 4.3. Let $k$ be a field, and let $R = k[X]$. Are the following statements true or false?

(a) Any finitely generated $R$-module which is finite dimensional as a $k$-vector space is torsion.

(b) The field $k$ is finite if and only if every finitely generated torsion $R$-module is finite (as a set).

*Easy exercise* 4.4. The *exponent* of an abelian group $G$ is the positive generator of the ideal $\mathrm{Ann}_{\mathbf{Z}}(G)$ in $\mathbf{Z}$. Construct two non-isomorphic finite abelian groups having the same order and the same exponent.

*Exercise* 4.5. Let $R$ be a PID. Let $p_1, p_2 \in R$ be prime elements and let $e_1, e_2 \in \mathbf{Z}_{>0}$.

(a) If $p_1$ and $p_2$ are non-associated, show that $\mathrm{Hom}_R(R/(p_1^{e_1}), R/(p_2^{e_2})) = 0$.

(b) If $p_1$ and $p_2$ are associated, show that $\mathrm{Hom}_R(R/(p_1^{e_1}), R/(p_2^{e_2})) \cong R/(p_1^{\min(e_1,e_2)})$.

*Exercise* 4.6. Classify all $\mathbf{Z}[i]$-modules with at most 10 elements up to isomorphism.

*Exercise* 4.7. Let $M$ be a finitely generated $\mathbf{C}[\![t]\!]$-module (see Exercise 1.6 for background on $\mathbf{C}[\![t]\!]$). Prove that there exists an integer $\ell \geq 1$ such that $t^\ell M$ is free as a $\mathbf{C}[\![t]\!]$-module.

*Exercise* 4.8. Let $R$ be a PID and let $p \in R$ be a prime element. Consider a short exact sequence of $R$-modules of the form $0 \longrightarrow R/(p) \longrightarrow M \longrightarrow R/(p) \longrightarrow 0$. Show that either $M \cong R/(p) \oplus R/(p)$ or $M \cong R/(p^2)$. Show also that both options may occur, and that $R/(p^2) \not\cong R/(p) \oplus R/(p)$.

*Exercise* 4.9. Let $a, b \in \mathbf{Z}$. What is the Smith canonical form of the $(2 \times 2)$-matrix $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ over $\mathbf{Z}$?

*Exercise* 4.10. Write the following abelian groups as a product of cyclic subgroups:

(a) the group $A$ with generators $a_1$, $a_2$ and $a_3$ and relations

$$7a_1 + 4a_2 + a_3 = 8a_1 + 5a_2 + 2a_3 = 9a_1 + 6a_2 + 3a_3 = 0,$$

(b) the group $B$ with generators $b_1$, $b_2$ and $b_3$ and relations

$$b_1 + b_2 + b_3 = 3b_1 + b_2 + 5b_3 = 0,$$

(c) the group $C$ with generators $c_1$, $c_2$, $c_3$ and $c_4$ and relations

$$2c_1 + 2c_2 + 6c_3 + 4c_4 = 10c_1 + 10c_2 + 2c_4 = 0.$$

*Exercise* 4.11. Let $A$ and $B$ be two $(n \times n)$-matrices over a field $k$. Show that $A$ and $B$ are similar over $k$, i.e., that $B = P^{-1}AP$ for some invertible $(n \times n)$-matrix $P$ over $k$, if and only if $XI_n - A$ and $XI_n - B$ have the same Smith normal form over $k[X]$ (up to multiplication by elements of $k$).

*Exercise* 4.12. Find the rational and Jordan canonical forms for the following matrices over $\mathbf{C}$:

$$A = \begin{pmatrix} 15 & -6 & -4 \\ 17 & -6 & -5 \\ 18 & -8 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ -2 & -1 & -1 & -1 \end{pmatrix}.$$

*Exercise* 4.13. Let $A$ be a $(4 \times 4)$-matrix over $\mathbf{R}$. Assume that the only eigenvalues of $A$ are $i$ and $-i$. Show that $A$ is similar (over $\mathbf{R}$) to one of the matrices

$$\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

(You may want to use the result of Exercise 4.11.)

*Exercise* 4.14. Let $k$ be an algebraically closed field. Prove that every $(n \times n)$-matrix $A$ over $k$ can be written in the form $A = D + N$, where $D$ is diagonalisable, $N$ is nilpotent and $DN = ND$. This is the so-called *Jordan decomposition* of $A$. (*Hint: use the existence of a Jordan canonical form for* $A$.)

*Hard exercise* 4.15. Prove that every square matrix over $\mathbf{R}$ is similar (over $\mathbf{R}$) to a diagonal block matrix, in which every block is of one of the following types:

$$\begin{pmatrix} r & 0 & 0 & \cdots & 0 \\ 1 & r & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & r & 0 \\ 0 & 0 & \cdots & 1 & r \end{pmatrix} \; (r \in \mathbb{R}) \text{ or } \begin{pmatrix} p & q & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -q & p & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & p & q & 0 & 0 & \cdots & 0 \\ 0 & 1 & -q & p & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & p & q & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & -q & p & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & 1 & 0 & p & q \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -q & p \end{pmatrix} \; (p, q \in \mathbf{R}).$$

*Hard exercise* 4.16. Consider the (non-finitely generated) $\mathbf{Z}$-module $M = \prod_{p \text{ prime}} \mathbf{Z}/p$. Show that the torsion submodule $T(M)$ is *not* a direct summand of $M$. (Compare with Theorem 4.1.5.)

# Chapter 5

# Fractions

The goal of this (very short) chapter is to generalise the construction of the fraction field of an integral domain to a much more flexible setting, called *localisation* of arbitrary rings and modules.

## 5.1 Localisation

The basic tool for the construction of localisations is the notion of *multiplicative subset* of a ring.

**Definition 5.1.1.** Let $R$ be a ring. A *multiplicative subset* of $R$ is a subset $S \subseteq R$ which contains $1$ and which is closed under multiplication: whenever $s, t \in S$, we have $st \in S$.

Given a ring $R$ and a multiplicative subset $S$ of $R$, we define a relation $\sim$ on $R \times S$ as follows: we have $(r_1, s_1) \sim (r_2, s_2)$ if and only if there exists $s \in S$ such that $s(s_2 r_1 - s_1 r_2) = 0$ in $R$.

*Quick question* 5.1.2. Check that $\sim$ (as defined above) is an equivalence relation.

We denote the set of equivalence classes of $\sim$ by $S^{-1}R$, and we denote the class of $(r, s) \in R \times S$ by $r/s$. Let us define addition and multiplication of such "fractions with denominators in $S$" as follows:

$$\frac{r_1}{s_1} + \frac{r_2}{s_2} = \frac{r_1 s_2 + r_2 s_1}{s_1 s_2}, \quad \frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1 r_2}{s_1 s_2} \quad \text{for } r_1, r_2 \in R,\ s_1, s_2 \in S.$$

*Quick question* 5.1.3. Check that these operations are well-defined and turn $S^{-1}R$ into a ring.

**Definition 5.1.4.** If $R$ is a ring and if $S \subseteq R$ is an arbitrary multiplicative subset, then the ring $S^{-1}R$ as defined above is called the *localisation* or *ring of fractions* of $R$ with respect to $S$.

In the setting of this definition, we have a natural homomorphism of rings $\iota : R \to S^{-1}R : r \mapsto r/1$. This homomorphism is not necessarily injective: if the multiplicative subset $S$ of $R$ contains a zero divisor $s$, and if $r \in R$ is a non-zero element such that $sr = 0$, then $r/1 = 0$ in $S^{-1}R$ (why?).

*Quick question* 5.1.5. With notation as above, check that $S^{-1}R$ is the zero ring if and only if $0 \in S$.

**Example 5.1.6.** A ring $R$ is an integral domain if and only if $S = R \setminus \{0\}$ is a multiplicative subset; in that case, $S^{-1}R$ is simply the fraction field of $R$ from your undergraduate abstract algebra course.

Rings of fractions can be characterised using the following universal property.

**Proposition 5.1.7.** *Let $R$ be a ring and let $S \subseteq R$ be a multiplicative subset. Let $\varphi : R \to \widetilde{R}$ be a homomorphism of rings such that $\varphi(S) \subseteq \widetilde{R}^\times$. Then $\varphi$ factors uniquely through the map $\iota : R \to S^{-1}R$, in other words: there exists a unique homomorphism $\psi : S^{-1}R \to \widetilde{R}$ such that $\varphi = \psi \circ \iota$.*

*Proof.* If $\psi : S^{-1}R \to \widetilde{R}$ is a homomorphism for which $\varphi = \psi \circ \iota$, then $\psi(r/1) = \varphi(r)$ for all $r \in R$. Therefore $\psi(1/s) = \psi((s/1)^{-1}) = \psi(s/1)^{-1} = \varphi(s)^{-1}$ for all $s \in S$, hence $\psi(r/s) = \varphi(r)\varphi(s)^{-1}$ for all $r \in R$ and $s \in S$. We conclude that there is at most one candidate for $\psi$. It is not hard to see that this candidate works, in the sense that it is both well-defined and a homomorphism. □

In particular, we know from Proposition 5.1.7 (and also from its proof) that the localisation $S^{-1}R$ and the homomorphism $\iota : R \to S^{-1}R$ have the following properties:

- if $s \in S$, then $\iota(s)$ is a unit in $S^{-1}R$;
- if $r \in R$ such that $\iota(r) = 0$, then $sr = 0$ for some $s \in S$;
- every element of $S^{-1}R$ is of the form $\iota(r)\iota(s)^{-1}$ for some $r \in R$ and $s \in S$.

Conversely, these properties characterise $S^{-1}R$, in the following sense – we leave the proof to the reader:

**Proposition 5.1.8.** *Let $R$ and $\widetilde{R}$ be rings, let $S \subseteq R$ be a multiplicative subset and let $\varphi : R \to \widetilde{R}$ be a homomorphism of rings with the following properties:*

- *if $s \in S$, then $\varphi(s)$ is a unit in $\widetilde{R}$;*
- *if $r \in R$ such that $\varphi(r) = 0$, then $sr = 0$ for some $s \in S$;*
- *every element of $\widetilde{R}$ is of the form $\varphi(r)\varphi(s)^{-1}$ for some $r \in R$ and $s \in S$.*

*Then there exists a unique isomorphism $\psi : S^{-1}R \to \widetilde{R}$ such that $\varphi = \psi \circ \iota$.*

**Example 5.1.9.** Let $R$ be a ring. If $\mathfrak{p}$ is a prime ideal, then $R \setminus \mathfrak{p}$ is a multiplicative subset of $R$. The *localisation of $R$ at $\mathfrak{p}$* is the ring $R_\mathfrak{p} = (R \setminus \mathfrak{p})^{-1}R$. The elements of $R_\mathfrak{p}$ of the form $r/s$ with $r \in \mathfrak{p}$ and $s \in R \setminus \mathfrak{p}$ form an ideal, which (by Proposition 1.1.18) is the unique maximal ideal in $R_\mathfrak{p}$ (denoted $\mathfrak{p}R_\mathfrak{p}$). In particular, $R_\mathfrak{p}$ is a local ring. For example, if $p$ is a prime number, then the localisation $\mathbf{Z}_{(p)}$ consists of all rational numbers whose denominator is prime to $p$.

**Example 5.1.10.** Let $R$ be ring and let $f \in R$. The set $S = \{f^n : n \in \mathbf{Z}_{\geq 0}\}$ is a multiplicative subset of $R$. We write $R_f$ for the localisation $S^{-1}R$. The elements of $R_f$ are all of the form $r/f^n$ for some $r \in R$ and $n \in \mathbf{Z}_{\geq 0}$. (Note that the homomorphism $\iota : R \to R_f$ is not injective if $f$ is a zero divisor.)

Given a ring $R$ and a multiplicative subset $S \subseteq R$, we can also localise $R$-modules with respect to $S$:

**Definition 5.1.11.** Let $R$ be a ring, let $S \subseteq R$ be a multiplicative subset and let $M$ be an $R$-module. Define a relation $\sim$ on $M \times S$ as follows: $(x_1, s_1) \sim (x_2, s_2)$ if and only if $s(s_2x_1 - s_1x_2) = 0$ for some $s \in S$. Then $\sim$ is an equivalence relation. Let $S^{-1}M$ be the set of equivalence classes, and denote the equivalence class of $(x, s)$ by $x/s$; then $S^{-1}M$ is an $S^{-1}R$-module, the *localisation of $M$ with respect to $S$*, with scalar multiplication given by $r/s_1 \cdot x/s_2 = (rx)/(s_1s_2)$.

As in Examples 5.1.9 and 5.1.10, if $S = R \setminus \mathfrak{p}$ for some prime ideal $\mathfrak{p}$, we write $M_\mathfrak{p}$ instead of $S^{-1}M$; if $S = \{f^n : n \in \mathbf{Z}_{\geq 0}\}$ for some $f \in R$, we write $M_f$ instead of $S^{-1}M$.

Given a ring $R$, a multiplicative subset $S \subseteq R$ and a homomorphism of $R$-modules $f : M \to N$, we get an induced homomorphism of $S^{-1}R$-modules $S^{-1}f : S^{-1}M \to S^{-1}N : x/s \mapsto f(x)/s$. Given another homomorphism of $R$-modules $g : N \to P$, we clearly have $S^{-1}g \circ S^{-1}f = S^{-1}(g \circ f)$.

**Proposition 5.1.12.** *With notation as above, if $M \xrightarrow{f} N \xrightarrow{g} P$ is exact, then so is*

$$S^{-1}M \xrightarrow{S^{-1}f} S^{-1}N \xrightarrow{S^{-1}g} S^{-1}P.$$

*Proof.* It is obvious that $\operatorname{im} S^{-1}(f) \subseteq \ker S^{-1}(g)$. Conversely, if $y/s \in \ker S^{-1}(g)$, then $g(y)/s = 0$ in $S^{-1}P$, whence $s'g(y) = 0$ for some $s' \in S$. It follows that $s'y \in \ker g = \operatorname{im} f$. Writing $s'y = f(x)$ for some $x \in M$, we conclude that $y/s = f(x)/(ss') \in \operatorname{im} S^{-1}(f)$, as desired. □

Proposition 5.1.12 implies (why?) that if $N$ is a submodule of the $R$-module $M$, then $S^{-1}N$ is a submodule of the $S^{-1}R$-module $S^{-1}M$, so that we can state the following (proofs left to the reader):

**Corollary 5.1.13.** *Let $R$ be a ring, and let $S \subseteq R$ be a multiplicative set. Given an $R$-module $M$ and submodules $N$ and $P$, we have $S^{-1}(N + P) = S^{-1}N + S^{-1}P$ and $S^{-1}(N \cap P) = S^{-1}N \cap S^{-1}P$.*

*Moreover, the $S^{-1}R$-modules $S^{-1}(M/N)$ and $(S^{-1}M)/(S^{-1}N)$ are isomorphic.*

## 5.2 Local versus global

It is often very useful to study the properties of algebraic objects (such as modules) defined over some commutative ring $R$ by analysing them "one prime at a time", by which we mean: after localising at a prime ideal $\mathfrak{p}$ of $R$. Information obtained after localising is said to be "local", and to challenge is to translate this local data to "global" information about the objects of study.

As a first (and possibly underwhelming) example, we have the basic fact that a module over a ring is the zero module if and only if it is the zero module *locally* at all primes of the ring:

**Proposition 5.2.1.** *Let $R$ be a ring and let $M$ be an $R$-module. Then $M = 0$ if and only if $M_{\mathfrak{p}} = 0$ for all prime ideals $\mathfrak{p}$ (or equivalently, $M_{\mathfrak{m}} = 0$ for all maximal ideals $\mathfrak{m}$).*

*Proof.* Assume that $M_{\mathfrak{p}} = 0$ for all prime ideals $\mathfrak{p}$. If $M \neq 0$, there exists a non-zero element $x \in M$. Then $\operatorname{Ann}_R(x)$ is contained in some maximal ideal $\mathfrak{m}$ of $R$. Since $x/1 = 0$ in $M_{\mathfrak{m}}$, it follows that $x$ is killed by a scalar contained in $R \setminus \mathfrak{m}$. This is impossible, since $\operatorname{Ann}_R(x) \subseteq \mathfrak{m}$. □

Similarly, injectivity of a homomorphism of $R$-modules can be checked locally at all primes of $R$:

**Proposition 5.2.2.** *Let $R$ be a ring, and let $\varphi : M \to N$ be a homomorphism of $R$-modules. Then $\varphi$ is injective if and only if $\varphi_{\mathfrak{p}} : M_{\mathfrak{p}} \to N_{\mathfrak{p}}$ is injective for all prime ideals $\mathfrak{p}$ of $R$.*

*Proof.* If $0 \to M \xrightarrow{\varphi} N$ is exact, then $0 \to M_{\mathfrak{p}} \xrightarrow{\varphi_{\mathfrak{p}}} N_{\mathfrak{p}}$ is exact for all primes $\mathfrak{p}$ by Proposition 5.1.12.

Conversely, assume that $\varphi_{\mathfrak{p}}$ is injective for all primes $\mathfrak{p}$. The natural sequence $0 \to \ker\varphi \to M \to N$ is exact, hence so is $0 \to (\ker\varphi)_{\mathfrak{p}} \to M_{\mathfrak{p}} \to N_{\mathfrak{p}}$ for all $\mathfrak{p}$ (again by Proposition 5.1.12). Our assumptions imply that $(\ker\varphi)_{\mathfrak{p}} = 0$ for all $\mathfrak{p}$. It now follows from Proposition 5.2.1 that $\ker\varphi = 0$. □

*Quick question* 5.2.3. Does Proposition 5.2.2 remain true if one replaces "injective" by "surjective"?

We will encounter other examples of so-called *local properties* in subsequent chapters. To get a good grip on such properties, it is often important to understand the relationship between the ideals of a given ring and the ideals of a localisation. Let us therefore take a brief look at this correspondence.

If $R$ is a ring, and if $S$ is a multiplicative subset of $R$, then any ideal $I$ of $R$ generates an ideal in $S^{-1}R$ (denoted by $S^{-1}I$) which consists of all elements of $S^{-1}R$ the form $x/s$, where $x \in I$ and $s \in S$. In fact every ideal of $S^{-1}R$ is of this form: given an ideal $J$ of $S^{-1}R$, it is not hard to see that $J = S^{-1}I$, where $I = \iota^{-1}(J)$ (and where, as before, $\iota : R \to S^{-1}R$ is the map given by $x \mapsto x/1$).

However, many ideals tend to "expand" after localisation: if $I$ intersects $S$, then the ideal $S^{-1}I$ of $S^{-1}R$ contains 1 and hence is equal to $S^{-1}R$ itself. In particular, we have the following:

**Proposition 5.2.4.** *With notation as above, the prime ideals of $S^{-1}R$ are in bijective correspondence with the prime ideals of $R$ which do not intersect the multiplicative subset $S$. In particular, if $\mathfrak{p}$ is a prime ideal of $R$, then the prime ideals of $R_{\mathfrak{p}}$ are in bijection with the prime ideals of $R$ contained in $\mathfrak{p}$.*

*Proof.* If $\mathfrak{q}$ is a prime ideal in $S^{-1}R$, then $\mathfrak{p} = \iota^{-1}(\mathfrak{q})$ is a prime ideal in $R$ such that $\mathfrak{q} = S^{-1}\mathfrak{p}$; it is easy to see that $\mathfrak{p} \cap S = \emptyset$. Conversely, let $\mathfrak{p}$ be a prime ideal in $R$. If $\overline{S}$ denotes the image of $S$ in $R/\mathfrak{p}$, then we have an isomorphism of rings $S^{-1}R/S^{-1}\mathfrak{p} \cong \overline{S}^{-1}(R/\mathfrak{p})$. Since the right hand side is contained in the fraction field of the integral domain $R/\mathfrak{p}$, it follows that the left hand side is either the zero ring or an integral domain, in other words, that $S^{-1}\mathfrak{p}$ is either a prime ideal or equal to $S^{-1}R$. Since the latter possibility occurs precisely when $\mathfrak{p} \cap S \neq \emptyset$, this yields the desired result. □

# Exercises

*Easy exercise* 5.1. Let $R$ be a ring. If $S, T \subseteq R$ are multiplicative subsets, then $ST = \{st : s \in S,\ t \in T\}$ is clearly again a multiplicative subset of $R$. If $U$ denotes the image of $T$ in $S^{-1}R$, show that we have an isomorphism of rings $(ST)^{-1}(R) \cong U^{-1}(S^{-1}R)$.

*Easy exercise* 5.2. Prove the following generalisation of Proposition 5.2.1: if $R$ is a ring, if $I$ is an ideal of $R$ and if $M$ is an $R$-module such that $M_{\mathfrak{p}} = 0$ for all primes $\mathfrak{p}$ of $R$ which contain $I$, then $M = IM$.

*Easy exercise* 5.3. Let $R$ be a ring, let $S \subseteq R$ be a multiplicative subset. If $I, J$ are ideals of $R$, show that the following equalities of ideals hold in the localised ring $S^{-1}R$: $S^{-1}(I + J) = S^{-1}I + S^{-1}J$, $S^{-1}(IJ) = (S^{-1}I)(S^{-1}J)$, $S^{-1}(I \cap J) = S^{-1}I \cap S^{-1}J$ and $S^{-1}\sqrt{I} = \sqrt{S^{-1}I}$.

*Exercise* 5.4. Let $R$ be a ring and let $\mathfrak{p}$ be a prime ideal in $R$. Show that the residue field of the local ring $R_{\mathfrak{p}}$ is naturally isomorphic to the fraction field of the integral domain $R/\mathfrak{p}$.

*Exercise* 5.5. Let $R$ be an integral domain, and let $S \subseteq R$ be a multiplicative subset.

(a) If $M$ is an $R$-module, show that the torsion $T(S^{-1}M)$ of $S^{-1}M$ is isomorphic to $S^{-1}T(M)$.

(b) Deduce from this that "being torsion free" is a local property of $R$-modules, i.e., the $R$-module $M$ is torsion free if and only if the $R_{\mathfrak{p}}$-module $M_{\mathfrak{p}}$ is torsion free, for all prime ideals $\mathfrak{p}$ of $R$.

*Exercise* 5.6. Let $R$ be a ring, let $S \subseteq R$ be a multiplicative subset and let $M$ be a *finitely generated* $R$-module. Show that $S^{-1}(\mathrm{Ann}_R(M)) = \mathrm{Ann}_R(S^{-1}M)$, and show that this equality may fail if the condition that $M$ be finitely generated is omitted.

*Exercise* 5.7. Show that "having trivial nilradical" (also called "being *reduced*") is a local property, i.e., prove that a ring $R$ has no non-zero nilpotent elements if and only if $R_{\mathfrak{p}}$ has no non-zero nilpotent elements, for all prime ideals $\mathfrak{p}$ of $R$. Show also that "being an integral domain" is *not* a local property.

# Chapter 6

# Tensor products

Linear maps (homomorphisms of modules over rings) are great because of their ubiquity and simplicity, and they are certainly one of the central topics of this course. Alas, many interesting maps are not linear... For example, if $R$ is a ring, then the multiplication map $R^2 \to R : (x, y) \mapsto xy$ is not linear; it is however *bilinear*. More generally, multilinear maps are omnipresent, and obviously it is important to develop a good framework for the study of such maps. One way to achieve this is by "approximating" them in the best possible way by linear maps; this is where the notion of *tensor product* comes in.

## 6.1 Tensor product of modules

**Definition 6.1.1.** Let $R$ be a ring. Let $M$ and $N$ be $R$-modules. A map $f : M \times N \to P$ with values in an $R$-module $P$ is *$R$-bilinear* if the following equalities hold for all $x, x' \in M$, $y, y' \in N$ and $r \in R$:

$$f(x+x', y) = f(x, y) + f(x', y), \quad f(x, y+y') = f(x, y) + f(x, y'), \quad f(rx, y) = f(x, ry) = rf(x, y).$$

With notation as above, if $f : M \times N \to P$ is $R$-bilinear and if $\varphi : P \to Q$ is a homomorphism of $R$-modules, then the composition $\varphi \circ f$ is again $R$-bilinear. The following theorem says that there is a *universal* $R$-bilinear map, from which all others can be obtained uniquely by composition:

**Theorem 6.1.2.** *Let $R$ be a ring, and let $M$ and $N$ be $R$-modules. There exist an $R$-module $P_0$ and an $R$-bilinear map $f_0 : M \times N \to P_0$ such that for every $R$-bilinear map $f : M \times N \to P$ with values in some $R$-module $P$, there exists a unique $R$-module homomorphism $\varphi : P_0 \to P$ such that $f = \varphi \circ f_0$.*

*Moreover, the pair $(P_0, f_0)$ is unique up to unique isomorphism: if $(P_1, f_1)$ is another pair with the same property, then there exists a unique $R$-module isomorphism $\psi : P_0 \to P_1$ such that $\psi \circ f_0 = f_1$.*

In fact the uniqueness part of Theorem is easy to prove; the existence part requires more work.

*Proof.* Let us first prove uniqueness. Let $(P_0, f_0)$ and $(P_1, f_1)$ be pairs with the required properties. Since $f_1 : M \times N \to P_1$ is $R$-bilinear, there exists a unique $R$-module homomorphism $\psi : P_0 \to P_1$ such that $f_1 = \psi \circ f_0$. Reversing the roles of $f_0$ and $f_1$, we see that there is a unique $R$-module homomorphism $\chi : P_1 \to P_0$ such that $f_0 = \chi \circ f_1$. Now $(\chi \circ \psi) \circ f_0 : M \times N \to P_0$ is $R$-bilinear, but so is $\mathrm{Id}_{P_0} \circ f_0 : M \times N \to P_0$; from the uniqueness requirement in the statement of the theorem, we therefore conclude that $\chi \circ \psi = \mathrm{Id}_{P_0}$. Similarly, we get $\psi \circ \chi = \mathrm{Id}_{P_1}$, proving that both $\psi$ and $\chi$ are isomorphisms.

Let us now tackle existence: we will construct such a pair $(P_0, f_0)$ explicitly. Let $F = R^{(M \times N)}$ be the free $R$-module with basis $M \times N$; given $(x, y) \in M \times N$, let $e_{(x,y)}$ be the corresponding basis vector.

The map $M \times N \to F : (x, y) \mapsto e_{(x,y)}$ has no reason to be $R$-bilinear, hence we will force it to become $R$-bilinear, by imposing relations via a quotient construction. Let $G$ be the submodule of $F$ generated by all elements of the following form (where $x, x' \in M$, $y, y' \in N$ and $r \in R$ are arbitrary):

$$e_{(x+x',y)} - e_{(x,y)} - e_{(x',y)},\ \ e_{(x,y+y')} - e_{(x,y)} - e_{(x,y')},\ e_{(rx,y)} - re_{(x,y)},\ e_{(x,ry)} - re_{(x,y)}.$$

Let $P_0 = F/G$, and let $f_0$ be the composition $M \times N \to F \twoheadrightarrow P_0$; then $f_0$ is $R$-bilinear by construction.

Let us check that the pair $(P_0, f_0)$ satisfies our requirements. If $f : M \times N \to P$ is an $R$-bilinear map with values in some $R$-module $P$, then there is a unique homomorphism $\overline{f} : R^{(M \times N)} \to P$ given by $\overline{f}(e_{(x,y)}) = f(x, y)$ for all $x \in M,\ y \in N$. The $R$-bilinearity of $f$ implies that $\overline{f}(G) = 0$, in other words: $\overline{f}$ induces a homomorphism $\varphi : P_0 \to P$ such that $f = \varphi \circ f_0$. Finally, $\varphi$ is unique: $P_0$ is generated by the images of the elements $(x, y) \in M \times N$, and $\varphi$ must send the image of $(x, y)$ to $f(x, y)$. □

**Definition 6.1.3.** With notation as in Theorem 6.1.2 above, we call a pair $(P_0, f_0)$ as in the statement of the theorem a *tensor product* of the $R$-modules $M$ and $N$. Since the pair $(P_0, f_0)$ is unique in a strong sense (up to unique isomorphism), we will also say *the* tensor product rather than *a* tensor product.

We will use the notation $M \otimes_R N$ for $P_0$, and $\otimes : M \times N \to M \otimes_R N : (x, y) \mapsto x \otimes y$ for $f_0$.

The construction used to prove the existence part of Theorem 6.1.2 may seem a bit opaque. In practice it will often be easier to use the universal property satisfied by the tensor product, together with the fact that $M \otimes_R N$ is generated by the so-called *elementary tensors*: elements of the form $x \otimes y$, where $x \in M$ and $y \in N$. Note that **not** all elements of $M \otimes_R N$ are elementary tensors in general, but obviously every element of $M \otimes_R N$ can be written as a sum of elementary tensors (often in many different ways!).

Summarizing, we have the following "rules" at our disposal to compute in the $R$-module $M \otimes_R N$:

$$(x + x') \otimes y = x \otimes y + x' \otimes y,\ \ x \otimes (y + y') = x \otimes y + x \otimes y',\ \ r(x \otimes y) = (rx) \otimes y = x \otimes (ry)$$

for all $x, x' \in M$, $y, y' \in N$ and $r \in R$. In particular, we have $x \otimes 0 = 0 \otimes y = 0$ for all $x \in M$, $y \in N$.

*Quick question* 6.1.4. Let $R$ be a ring, and let $M$ and $N$ be $R$-modules. If $M = \langle x_i \rangle_{i \in I}$ and $N = \langle y_j \rangle_{j \in J}$, check that $M \otimes_R N = \langle x_i \otimes y_j \rangle_{i \in I,\, j \in J}$. This implies in particular that if both $M$ and $N$ are finitely generated, then so is their tensor product $M \otimes_R N$.

*Remark* 6.1.5. The notation $x \otimes y$ is inherently ambiguous, since it is not a priori clear to which tensor product it belongs. For example, if $M$ and $N$ are $R$-modules with submodules $M'$ and $N'$ respectively, then it may happen for some $x \in M'$ and $y \in N'$ that $x \otimes y$ is zero as an element of $M \otimes_R N$, but not as an element of $M' \otimes_R N'$. Indeed: take $R = \mathbf{Z}$, $M = \mathbf{Z}$ and $N = \mathbf{Z}/2$. Let $M' = 2\mathbf{Z}$ and $N' = N$. Then $2 \otimes \overline{1}$ is zero as an element of $M \otimes_R N$ since $2 \otimes \overline{1} = (2 \cdot 1) \otimes \overline{1} = 1 \otimes (2 \cdot \overline{1}) = 1 \otimes \overline{2} = 1 \otimes \overline{0} = 0$.

However, $2 \otimes \overline{1}$ is non-zero in $M' \otimes_R N'$: we will check this using an ad hoc argument, but we will soon develop better methods to prove this. Consider $f : M' \times N' \to N' : (2a, \overline{b}) \mapsto \overline{ab}$. It is not hard to check that this map is well-defined and $R$-bilinear; therefore $f$ factors over $\otimes : M' \times N' \to M' \otimes_R N'$. However, $f((2, \overline{1})) = \overline{1} \in N'$ is non-zero, hence $2 \otimes \overline{1}$ cannot be zero in $M' \otimes_R N'$!

In general, it can be quite hard to decide whether a given element of a tensor product is actually zero. Of course we know that $\sum_{i=1}^{n} x_i \otimes y_i = 0$ in $M \otimes_R N$ if and only if for every $R$-module $P$ and every $R$-bilinear map $f : M \times N \to P$, we have $\sum_{i=1}^{n} f(x_i, y_i) = 0$, but this criterion is not easy to use in practice. Let us look at another example showing that (perhaps unexpected?) cancellations may occur.

**Example 6.1.6.** We have $\mathbf{Z}/2 \otimes_{\mathbf{Z}} \mathbf{Z}/3 = 0$. Indeed, we have $\overline{a} \otimes \overline{b} = \overline{3a} \otimes \overline{b} = \overline{a} \otimes 3\overline{b} = \overline{a} \otimes \overline{0} = 0$ for all $a, b \in \mathbf{Z}$; hence every elementary tensor is zero, and therefore the tensor product is zero as well. In other words: there exist no non-zero bilinear maps from $\mathbf{Z}/2 \times \mathbf{Z}/3$ to *any* abelian group.

More generally, we have $\mathbf{Z}/m \otimes_{\mathbf{Z}} \mathbf{Z}/n \cong \mathbf{Z}/d$, where $d = \gcd(m, n)$. Indeed, let us first observe that $\overline{a} \otimes \overline{b} = ab(\overline{1} \otimes \overline{1})$ for all $a, b \in \mathbf{Z}$, so that $\mathbf{Z}/m \otimes_{\mathbf{Z}} \mathbf{Z}/n$ must be the cyclic abelian group generated by $\overline{1} \otimes \overline{1}$. Next, since $m(\overline{1} \otimes \overline{1}) = 0$ and similarly $n(\overline{1} \otimes \overline{1}) = 0$, we have $d(\overline{1} \otimes \overline{1}) = 0$; therefore the order of our cyclic group divides $d$. But the map $f : \mathbf{Z}/m \times \mathbf{Z}/n \to \mathbf{Z}/d : (\overline{a}, \overline{b}) \mapsto \overline{ab}$ is well-defined (why?) and $\mathbf{Z}$-bilinear. The induced map $\varphi : \mathbf{Z}/m \otimes_{\mathbf{Z}} \mathbf{Z}/n \to \mathbf{Z}/d$ maps $\overline{1} \otimes \overline{1}$ to the element $\overline{1} \in \mathbf{Z}/d$ of order $d$; in particular, $\mathbf{Z}/m \otimes_{\mathbf{Z}} \mathbf{Z}/n$ has order at least $d$. We conclude that $\varphi$ must be an isomorphism.

Let us now develop some of the basic rules to do computations involving tensor products. In each case, the definition of the tensor product via its universal property plays a key part in the proof.

**Proposition 6.1.7.** *Let $R$ be a ring. If $M$ and $N$ are $R$-modules, there exists a unique isomorphism of $R$-modules $M \otimes_R N \to N \otimes_R M$ defined at the level of elementary tensors by $m \otimes n \mapsto n \otimes m$.*

*Proof.* The map $f_0 : M \times N \to N \otimes_R M : (m, n) \mapsto n \otimes m$ is well-defined and $R$-bilinear. By Theorem 6.1.2, it suffices to show that an arbitrary $R$-bilinear map $f : M \times N \to P$ factors uniquely over $f_0$, i.e., that $f = \varphi \circ f_0$ for a uniquely defined $R$-module homomorphism $\varphi : N \otimes_R M \to P$.

Note that $f_0$ is the composition of the maps $\sigma : M \times N \to N \times M$ and $\otimes : N \times M \to N \otimes_R M$. It is clear that there exists a unique $R$-bilinear map $f^\sigma : N \times M \to P$ such that $f = f^\sigma \circ \sigma$. Now we see (again by Theorem 6.1.2) that there exists a unique $R$-module homomorphism $\varphi : N \otimes_R M \to P$ such that $f^\sigma = \varphi \circ \otimes$. It follows that $f = f^\sigma \circ \sigma = (\varphi \circ \otimes) \circ \sigma = \varphi \circ (\otimes \circ \sigma) = \varphi \circ f_0$ for a uniquely defined $R$-module homomorphism $\varphi$, which is exactly what we had to prove. □

The following result can be proven in a very similar way.

**Proposition 6.1.8.** *Let $R$ be a ring, and let $M$ be an $R$-module. There exists a unique isomorphism of $R$-modules $R \otimes_R M \to M$ defined at the level of elementary tensors by $a \otimes m \to am$.*

In fact, Proposition 6.1.8 is almost a tautology: the key bilinear map to consider here is the scalar multiplication $R \times M \to M$ which makes $M$ into an $R$-module. It is not hard to see that the inverse of the map $R \otimes_R M \to M$ in the statement is the map $M \to R \otimes_R M : m \mapsto 1 \otimes m$ (check this!).

Moreover, tensor products commute with direct sums. In particular, the tensor product of two free $R$-modules is again free, and the tensor product of $R^m$ and $R^n$ is isomorphic to $R^{mn}$:

**Proposition 6.1.9.** *Let $R$ be a ring, let $M$ be an $R$-module and let $(N_\alpha)_{\alpha \in A}$ be a family of $R$-modules. Then*

$$M \otimes_R \left( \bigoplus_{\alpha \in A} N_\alpha \right) \cong \bigoplus_{\alpha \in A} (M \otimes_R N_\alpha).$$

*Proof.* We will construct homomorphisms in both directions which are inverses to each other.

First consider the map $M \times \left(\bigoplus_{\alpha \in A} N_\alpha\right) \to \bigoplus_{\alpha \in A}(M \otimes_R N_\alpha) : (m, (n_\alpha)_{\alpha \in A}) \mapsto (m \otimes n_\alpha)_{\alpha \in A}$. Since this map is $R$-bilinear, it factors over $M \otimes_R \left(\bigoplus_{\alpha \in A} N_\alpha\right)$ by Theorem 6.1.2, whence the existence of an $R$-module homomorphism $\varphi : M \otimes_R \left(\bigoplus_{\alpha \in A} N_\alpha\right) \to \bigoplus_{\alpha \in A}(M \otimes_R N_\alpha)$ defined at the level of elementary tensors by the formula $\varphi(m \otimes (n_\alpha)_{\alpha \in A}) = (m \otimes n_\alpha)_{\alpha \in A}$.

For each $\alpha \in A$, let $\iota_\alpha : N_\alpha \to \bigoplus_{\alpha \in A} N_\alpha$ be the inclusion. The map $M \times N_\alpha \to M \otimes_R \left(\bigoplus_{\alpha \in A} N_\alpha\right)$ given by $(m, n_\alpha) \mapsto m \otimes \iota_\alpha(n_\alpha)$ is $R$-bilinear, hence we obtain (again by Theorem 6.1.2) an induced homomorphism of $R$-modules $f_\alpha : M \otimes_R N_\alpha \to M \otimes_R \left(\bigoplus_{\alpha \in A} N_\alpha\right)$ defined at the level of elementary tensors by $m \otimes n_\alpha \mapsto m \otimes \iota_\alpha(n_\alpha)$. Now consider $\psi : \bigoplus_{\alpha \in A}(M \otimes_R N_\alpha) \to M \otimes_R \left(\bigoplus_{\alpha \in A} N_\alpha\right)$ which sends an $A$-tuple $(x_\alpha)_{\alpha \in A}$ to $\sum_{\alpha \in A} f_\alpha(x_\alpha)$. We claim that $\psi$ is a two-sided inverse for $\varphi$.

Checking that $\psi \circ \varphi = \mathrm{Id}$ can be done on elementary tensors (why?): we have

$$(\psi \circ \varphi)((m \otimes (n_\alpha)_{\alpha\in A}) = \psi((m \otimes n_\alpha)_{\alpha\in A}) = \sum_{\alpha\in A} f_\alpha(m \otimes n_\alpha) = \sum_{\alpha\in A} m \otimes \iota_\alpha(n_\alpha) = m \otimes (n_\alpha)_{\alpha\in A}.$$

We leave it to the reader to check that $\varphi \circ \psi = \mathrm{Id}$. □

The analogue of Proposition 6.1.9 fails for direct products, see Exercise 6.12.

**Proposition 6.1.10.** *Let $R$ be a ring, and let $M$ be an $R$-module. For any ideal $I$ of $R$, we have*

$$M \otimes_R R/I \cong M/IM.$$

*Proof.* We sketch the argument and leave the details to the reader: there exists a well-defined $R$-module homomorphism $\varphi : M \otimes_R R/I \to M/IM$, defined on elementary tensors by $\varphi(m \otimes \overline{r}) = \overline{rm}$. Now the map $\psi : M/IM \to M \otimes_R R/I : \overline{m} \mapsto m \otimes \overline{1}$ is a well-defined inverse for $\varphi$. □

*Quick question* 6.1.11. Prove Proposition 6.1.10 a second time, now using the universal property in the statement of Theorem 6.1.2, i.e., show directly that any $R$-bilinear map $M \times R/I \to P$ to an arbitrary $R$-module $P$ factors uniquely over the bilinear map $M \times R/I \to M/IM : (m, \overline{r}) \to \overline{rm}$.

*Quick question* 6.1.12. Let $I$ and $J$ be ideals in a ring $R$. Show that $R/I \otimes_R R/J \cong R/(I + J)$.

The following notion generalises Definition 6.1.1:

**Definition 6.1.13.** Let $R$ be a ring, and let $M_1, \cdots, M_n$ be $R$-modules. A map $f : M_1 \times \cdots \times M_n \to P$ with values in an $R$-module $P$ is *$R$-multilinear* if it is $R$-linear in each variable. This means that if $x_1 \in M_1, \cdots, x_{n-1} \in M_{n-1}$ are arbitrary elements, then the map $M_n \to P : y \mapsto f(x_1, \cdots, x_{n-1}, y)$ is an $R$-module homomorphism, and similarly for the other coordinates.

With notation as above, it is possible to define a "multi-tensor product" $M_1 \otimes \cdots \otimes M_n$, generated by all elementary tensors $x_1 \otimes \cdots \otimes x_n$, where $x_i \in M_i$ for $1 \le i \le n$. We leave the details to the reader and simply state the following generalisation of Theorem 6.1.2:

**Proposition 6.1.14.** *Let $R$ be a ring, and let $M_1, \cdots, M_n$ be $R$-modules.*

*There exist an $R$-module $P_0$ and an $R$-multilinear map $f_0 : M_1 \times \cdots \times M_n \to P_0$ such that for every $R$-multilinear map $f : M_1 \times \cdots \times M_n \to P$ with values in some $R$-module $P$, there exists a unique $R$-module homomorphism $\varphi : P_0 \to P$ such that $f = \varphi \circ f_0$.*

*Moreover, the pair $(P_0, f_0)$ is unique up to unique isomorphism: if $(P_1, f_1)$ is another pair with the same property, then there exists a unique isomorphism $\psi : P_0 \to P_1$ such that $\psi \circ f_0 = f_1$.*

This new piece of terminology allows us to state the following result (proof left to the reader):

**Proposition 6.1.15.** *Let $R$ be a ring and let $M$, $N$ and $P$ be $R$-modules. There exist unique isomorphisms*

$$(M \otimes_R N) \otimes_R P \longrightarrow M \otimes_R (N \otimes_R P) \longrightarrow M \otimes_R N \otimes_R P$$

*defined at the level of elementary tensors by*

$$(x \otimes y) \otimes z \mapsto x \otimes (y \otimes z) \mapsto x \otimes y \otimes z.$$

Now that we have understood (at least to some extent. . . ) how to construct tensor products of modules, it is quite simple to define tensor products of *homomorphisms* of $R$-modules.

Given a ring $R$ and two homomorphisms $f : M \to M'$ and $g' : N \to N'$ of $R$-modules, the map

$$M \times N \to M' \otimes_R N' : (x, y) \mapsto f(x) \otimes g(y)$$

is easily seen to be $R$-bilinear, hence induces an $R$-module homomorphism

$$f \otimes g : M \otimes_R N \to M' \otimes_R N'$$

defined at the level of elementary tensors by $(f \otimes g)(x \otimes y) = f(x) \otimes g(y)$ for all $x \in M$ and $y \in N$.

Given additional homomorphisms $f' : M' \to M''$ and $g : N' \to N''$ of $R$-modules, it is clear that the homomorphisms $(f' \circ f) \otimes (g' \circ g)$ and $(f' \otimes g') \circ (f \otimes g)$ agree on all elementary tensors in $M \otimes_R N$, hence must be equal:

$$(f' \circ f) \otimes (g' \circ g) = (f' \otimes g') \circ (f \otimes g). \tag{6.1}$$

This is the *functorial behaviour* of the tensor product, which we will encounter in the next chapter.

We end this rather long section with the notions of *restriction* and *extension* of scalars. If $f : R \to S$ is a homomorphism of rings, and if $N$ is an $S$-module, then $N$ has a natural $R$-module structure: given $r \in R$ and $y \in N$, define $ry$ to be $f(r)y$. This $R$-module is said to be obtained from $N$ by *restriction of scalars*. (In particular, $f$ defines an $R$-module structure on $S$ itself.)

On the other hand, if $M$ is an $R$-module, then we can form the $R$-module $M_S = S \otimes_R M$, where $S$ is equipped with the $R$-module structure we have just seen. It is not hard to see that $M_S$ can be turned into an $S$-module: let $s_1(s_2 \otimes x) = s_1 s_2 \otimes x$ for all $s_1, s_2 \in S$ and $x \in M$. The $S$-module $M_S$ is said to be obtained from $M$ by *extension of scalars* from $R$ to $S$. Note that if $M = \langle x_i \rangle_{i \in I}$, then $M_S = \langle 1 \otimes x_i \rangle_{i \in I}$.

**Example 6.1.16.** Localisation of a module, as introduced in §5.1, is a special case of extension of scalars. Indeed, given a ring $R$, a multiplicative subset $S$ and an $R$-module $M$, we have a natural isomorphism

$$\varphi : S^{-1}R \otimes_R M \to S^{-1}M,$$

constructed as follows. First, note that the map $S^{-1}R \times M \to S^{-1}M : (r/s, m) \mapsto rm/s$ is $R$-bilinear, hence factors through $\varphi$; on elementary tensors, $\varphi$ is given by $\varphi(r/s \otimes m) = rm/s$. Hence $\varphi$ is surjective.

To show that $\varphi$ is also injective, let $\sum_{i=1}^n r_i/s_i \otimes m_i \in \ker \varphi$. Set $s = \prod_{i=1}^n s_i$ and $t_i = \prod_{j \neq i} s_j$, then

$$\sum_{i=1}^n \frac{r_i}{s_i} \otimes m_i = \sum_{i=1}^n \frac{r_i t_i}{s} \otimes m_i = \sum_{i=1}^n \frac{1}{s} \otimes r_i t_i m_i = \frac{1}{s} \otimes \sum_{i=1}^n r_i t_i m_i.$$

Let $m = \sum_{i=1}^n r_i t_i m_i$, then we have $\varphi(1/s \otimes m) = m/s = 0$. This means that $tm = 0$ for some $t \in S$, whence $1/s \otimes m = 1/st \otimes tm = 1/st \otimes 0 = 0$, as desired. We conclude that $\varphi$ is an isomorphism.

The following result says that localisation and tensor product "commute":

**Proposition 6.1.17.** *Let $R$ be a ring, let $S$ be a multiplicative subset, and let $M$ and $N$ be $R$-modules. There is an isomorphism of $S^{-1}R$-modules*

$$f : S^{-1}M \otimes_{S^{-1}R} S^{-1}N \longrightarrow S^{-1}(M \otimes_R N)$$

*given on elementary tensors by*

$$f\left(\frac{m}{s} \otimes \frac{n}{t}\right) = \frac{m \otimes n}{st}.$$

*In particular, if $\mathfrak{p}$ is a prime ideal in $R$, we have $M_\mathfrak{p} \otimes_{R_\mathfrak{p}} N_\mathfrak{p} \cong (M \otimes_R N)_\mathfrak{p}$ as $R_\mathfrak{p}$-modules.*

## 6.2 Tensor product of algebras

Rings have addition and multiplication; modules over rings have addition and *scalar* multiplication. It is often useful to consider objects which have both an "internal" multiplication and a scalar multiplication, with elements of some other ring. This leads us to the notion of *algebra* over a given ring:

**Definition 6.2.1.** Let $R$ be a ring. An *$R$-algebra* is a pair $(S, f)$, where $S$ is a ring and $f : R \to S$ is a homomorphism of rings, and where $S$ is considered as an $R$-module via restriction of scalars.

It is important to note that the ring and module structures on $S$ in the above definition are compatible, in a sense which the reader should try to formulate for him- or herself. We will often be deliberately sloppy and say "let $S$ be an $R$-algebra" rather than "let $(S, f)$ be an $R$-algebra".

**Example 6.2.2.** Let $K$ be a field. Then a $K$-algebra is nothing but a ring which contains $K$ as a subring.

*Quick question* 6.2.3. Check that an arbitrary ring is automatically a **Z**-algebra in a unique way.

**Definition 6.2.4.** Let $R$ be a ring, and let $(S_1, f_1)$ and $(S_2, f_2)$ be $R$-algebras. A *homomorphism of $R$-algebras* from $(S_1, f_1)$ to $(S_2, f_2)$ is a homomorphism of rings $g : S_1 \to S_2$ such that $g \circ f_1 = f_2$.

The condition $g \circ f_1 = f_2$ in the definition ensures that $g$ be a homomorphism of $R$-modules.

**Definition 6.2.5.** A ring homomorphism $f : R \to S$ is *finite*, or equivalently, $S$ is said to be a *finite $R$-algebra*, if $S$ is finitely generated as an $R$-module. The homomorphism $f$ is *of finite type*, or equivalently, $S$ is said to be a *finitely generated* $R$-algebra, if there exists a surjective homomorphism of $R$-algebras from the polynomial ring $R[X_1, \cdots, X_n]$, for some $n \geq 1$, onto $S$.

The point is now that the tensor product of two $R$-algebras is not only an $R$-module, but again an $R$-algebra in a natural way. Indeed, let $R$ be a ring and let $(S, f)$ and $(T, g)$ be two $R$-algebras. Since $S$ and $T$ are certainly $R$-modules, we may form the tensor product $U = S \otimes_R T$, which is again an $R$-module. We will now define a multiplication on this tensor product. Consider the map

$$S \times T \times S \times T \longrightarrow U : (s, t, s', t') \mapsto ss' \otimes tt'.$$

This map is clearly $R$-multilinear (see Definition 6.1.13), hence defines an $R$-module homomorphism

$$S \otimes_R T \otimes_R S \otimes_R T \longrightarrow U.$$

This means that we have an $R$-module homomorphism $U \otimes_R U \to U$, or equivalently, an $R$-bilinear map $\mu : U \times U \to U$, with the property that $\mu(s \otimes t, s' \otimes t') = ss' \otimes tt'$ for all $s, s' \in S$ and $t, t' \in T$. We have therefore defined a multiplication on $U = S \otimes_R T$, given on elementary tensors by $(s \otimes t)(s' \otimes t') = ss' \otimes tt'$ and in general by the formula

$$\left(\sum_i (s_i \otimes t_i)\right)\left(\sum_j (s'_j \otimes t'_j)\right) = \sum_{i,j} (s_i s'_j \otimes t_i t'_j).$$

The reader should check that this turns $U = S \otimes_R T$ into a commutative ring, with identity element $1 \otimes 1$, and that the map $R \to U : r \mapsto f(r) \otimes 1 = 1 \otimes g(r)$ turns $U$ into an $R$-algebra.

**Example 6.2.6.** Let $R$ be a ring. Then $R[X_1, \cdots, X_m] \otimes_R R[Y_1, \cdots, Y_n] \cong R[Z_1, \cdots, Z_{m+n}]$ as $R$-algebras. Indeed, we have an isomorphism $R[X_1, \cdots, X_m] \otimes_R R[Y_1, \cdots, Y_n] \to R[Z_1, \cdots, Z_{m+n}]$ of $R$-modules which sends $X_1 \otimes 1, \cdots, X_m \otimes 1$ to $Z_1, \cdots, Z_m$ and $1 \otimes Y_1, \cdots, 1 \otimes Y_n$ to $Z_{m+1}, \cdots, Z_{m+n}$. It is not hard to see from the description of the multiplication on $R[X_1, \cdots, X_m] \otimes_R R[Y_1, \cdots, Y_n]$ given above that this isomorphism of $R$-modules is also an isomorphism of rings.

**Example 6.2.7.** Let $R$ be a ring, let $S$ be an $R$-algebra and let $I$ be an ideal in $R[X_1, \cdots, X_m]$. Then we have an isomorphism of $R$-algebras $S \otimes_R R[X_1, \cdots, X_m]/I \cong S[X_1, \cdots, X_m]/I$, where in the right hand side $I$ is viewed as an ideal in $S[X_1, \cdots, X_m]$ via the map $R \to S$.

This describes in very concrete terms the extension of scalars, from $R$ to another (arbitrary) $R$-algebra $S$, of the finitely generated $R$-algebra $R[X_1, \cdots, X_m]/I$.

**Example 6.2.8.** The inclusion $\mathbf{R} \hookrightarrow \mathbf{C}$ makes $\mathbf{C}$ into an $\mathbf{R}$-algebra. Let us compute the tensor product $\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C}$, first as an $\mathbf{R}$-module, then as an $\mathbf{R}$-algebra. As an $\mathbf{R}$-module, $\mathbf{C}$ is simply a two-dimensional real vector space with basis $\{1, i\}$. Hence $\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C}$ is four-dimensional with basis $\{1 \otimes 1, 1 \otimes i, i \otimes 1, i \otimes i\}$ (but this does not matter: all four-dimensional real vector spaces are isomorphic anyway).

Let us now compute $\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C}$ as an $\mathbf{R}$-algebra. We have (using Example 6.2.7)

$$\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C} \cong \mathbf{R}[X]/(X^2+1) \otimes_{\mathbf{R}} \mathbf{C} \cong \mathbf{C}[X]/(X^2+1) \cong \mathbf{C}[X]/(X+i) \times \mathbf{C}[X]/(X-i) \cong \mathbf{C} \times \mathbf{C}.$$

Of course the $\mathbf{R}$-algebra structure yields richer information on $\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C}$ than the $\mathbf{R}$-vector space structure!

## 6.3 The Hom-$\otimes$ adjunction

Let $R$ be a ring and let $M$, $N$ and $P$ be $R$-modules. There is a natural map

$$\varphi : \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)) \longrightarrow \mathrm{Hom}_R(M \otimes_R N, P) \tag{6.2}$$

defined as follows. Given $f \in \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P))$, we get for every $m \in M$ an $R$-module homomorphism $f(m) : N \to P$. Hence, given a pair $(m, n) \in M \times N$, we get the element $f(m)(n) \in P$.

*Quick question* 6.3.1. Check that the assignment $M \times N \to P$ obtained in this way is $R$-bilinear.

Now Theorem 6.1.2 tells us that the $R$-bilinear map $M \times N \to P$ obtained from $f$ corresponds to a unique $R$-module homomorphism $M \otimes_R N \to P$ which we denote by $\varphi(f)$.

*Quick question* 6.3.2. Check that this construction yields an $R$-module homomorphism

$$\varphi : \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)) \longrightarrow \mathrm{Hom}_R(M \otimes_R N, P) : f \mapsto \varphi(f).$$

**Proposition 6.3.3.** *With notation as above, the map $\varphi$ is an* isomorphism *of $R$-modules.*

*Proof.* Let us construct a two-sided inverse for $\varphi$. Any $g \in \mathrm{Hom}_R(M \otimes_R N, P)$ corresponds to a unique $R$-bilinear map $b : M \times N \to P$. Given $m \in M$, the map $b(m, -) : N \to P : n \mapsto b(m, n)$ is $R$-linear, and the assignment $\psi(g) : M \to \mathrm{Hom}_R(N, P) : m \mapsto b(m, -)$ is an $R$-module homomorphism. Now

$$\psi : \mathrm{Hom}_R(M \otimes_R N, P) \to \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)) : g \mapsto \psi(g)$$

is an $R$-module homomorphism as well, and it is not hard to check that this is the desired inverse. □

The map (6.2) constructed above is the so-called *Hom-$\otimes$ adjunction*, to which we will come back in the final two chapters of this course, using the professional language of categories and functors.

*Remark* 6.3.4. The method used to construct $\varphi$ and $\psi$ is called *currying* in computer science.

## 6.4 Flatness

We have seen in §5.1 that localisation is *exact*: if $R$ is a ring, then localising an exact sequence of $R$-modules with respect to a multiplicative subset of $R$ yields another exact sequence. However, such a property fails for tensor products in general, as shown by the following example.

**Example 6.4.1.** Consider the exact sequence of abelian groups $0 \to \mathbf{Z} \xrightarrow{\cdot 2} \mathbf{Z}$. Extension of scalars from $\mathbf{Z}$ to $\mathbf{Z}/2$ yields the sequence $0 \to \mathbf{Z} \otimes_{\mathbf{Z}} \mathbf{Z}/2 \xrightarrow{f} \mathbf{Z} \otimes_{\mathbf{Z}} \mathbf{Z}/2$, where $f = (\cdot 2) \otimes \mathrm{Id}$. This sequence is no longer exact, because $f$ is the zero morphism: indeed, for any elementary tensor $x \otimes y \in \mathbf{Z} \otimes_{\mathbf{Z}} \mathbf{Z}/2$, we have $f(x \otimes y) = 2x \otimes y = x \otimes 2y = x \otimes 0 = 0$. Since $\mathbf{Z} \otimes_{\mathbf{Z}} \mathbf{Z}/2 \cong \mathbf{Z}/2 \neq 0$, $f$ is not injective.

However, not all is lost: tensor product is still *right exact*, in the following sense.

**Proposition 6.4.2.** *Let $R$ be a ring, and assume that $M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3 \longrightarrow 0$ is an exact sequence of $R$-modules. If $N$ is an arbitrary $R$-module, then the induced sequence*

$$M_1 \otimes_R N \xrightarrow{f_1 \otimes \mathrm{Id}} M_2 \otimes_R N \xrightarrow{f_2 \otimes \mathrm{Id}} M_3 \otimes_R N \longrightarrow 0 \tag{6.3}$$

*is still exact.*

There are various ways to prove this, but the best method (by far) uses the Hom-$\otimes$ adjunction.

*Proof.* By Proposition 2.4.7, (6.3) is exact if and only if for any $R$-module $P$, the induced sequence

$$0 \longrightarrow \mathrm{Hom}_R(M_3 \otimes_R N, P) \longrightarrow \mathrm{Hom}_R(M_2 \otimes_R N, P) \longrightarrow \mathrm{Hom}_R(M_1 \otimes_R N, P)$$

is exact. The Hom-$\otimes$ adjunction from the previous section allows us to identify this sequence to

$$0 \longrightarrow \mathrm{Hom}_R(M_3, \mathrm{Hom}_R(N, P)) \longrightarrow \mathrm{Hom}_R(M_2, \mathrm{Hom}_R(N, P)) \longrightarrow \mathrm{Hom}_R(M_1, \mathrm{Hom}_R(N, P)).$$

Since we know that the latter sequence is exact (again by Proposition 2.4.7), we are done. □

Moreover some modules do preserve exactness in general, which leads us to the following definition:

**Definition 6.4.3.** Let $R$ be a ring. An $R$-module $N$ is said to be *flat* if $N$ preserves exactness, in the following sense: if $M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3$ is an exact sequence of $R$-modules, then so is the sequence

$$M_1 \otimes_R N \xrightarrow{f_1 \otimes \mathrm{Id}} M_2 \otimes_R N \xrightarrow{f_2 \otimes \mathrm{Id}} M_3 \otimes_R N.$$

*Quick question* 6.4.4. Let $R$ be a ring. Show that a free $R$-module (of arbitrary rank) is flat. Show also that if $S$ is a multiplicative subset of $R$, then $S^{-1}R$ is a flat $R$-module.

The following proposition yields an alternative characterisation of flatness.

**Proposition 6.4.5.** *Let $R$ be a ring and let $N$ be an $R$-module. The following statements are equivalent:*

*(a) $N$ is a flat $R$-module;*

*(b) tensoring with $N$ preserves* short *exact sequences: if $0 \to M_1 \to M_2 \to M_3 \to 0$ is an exact sequence of $R$-modules, then so is the sequence $0 \to M_1 \otimes_R N \to M_2 \otimes_R N \to M_3 \otimes_R N \to 0$;*

*(c) if $f : M \to M'$ is an injective $R$-module homomorphism, then so is $f \otimes \mathrm{Id} : M \otimes_R N \to M' \otimes_R N$.*

*Proof.* The equivalence of (a) and (b) follows from the fact that any exact sequence can be split up into short exact sequences (Lemma 2.4.5). The equivalence of (b) and (c) follows from Proposition 6.4.2. □

It turns out that flatness is a local property, in the sense of §5.2:

**Proposition 6.4.6.** *Let $R$ be a ring and let $M$ be an $R$-module. Then $M$ is flat if and only if for every prime ideal $\mathfrak{p}$ of $R$, the localisation $M_{\mathfrak{p}}$ is a flat $R_{\mathfrak{p}}$-module.*

*Proof.* If $M$ is a flat $R$-module, then $M_{\mathfrak{p}}$ is a flat $R_{\mathfrak{p}}$-module by Example 6.1.16 and Proposition 5.1.12. Conversely, assume that $M_{\mathfrak{p}}$ is a flat $R_{\mathfrak{p}}$-module for all primes $\mathfrak{p}$. Let $f : N \to P$ be an injective $R$-module homomorphism. Then $f_{\mathfrak{p}} : N_{\mathfrak{p}} \to P_{\mathfrak{p}}$ is injective for all primes $\mathfrak{p}$, and therefore so is

$$\mathrm{Id} \otimes f_{\mathfrak{p}} : M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} N_{\mathfrak{p}} \longrightarrow M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} P_{\mathfrak{p}}$$

(indeed, $M_{\mathfrak{p}}$ is flat over $R_{\mathfrak{p}}$). It now follows from Proposition 6.1.17 that the $R_{\mathfrak{p}}$-module homomorphism

$$(\mathrm{Id} \otimes f)_{\mathfrak{p}} : (M \otimes_R N)_{\mathfrak{p}} \longrightarrow (M \otimes_R P)_{\mathfrak{p}}$$

is injective for all primes $\mathfrak{p}$, and from Proposition 5.2.2 that $\mathrm{Id} \otimes f : M \otimes_R N \to M \otimes_R P$ is injective.

Using Proposition 6.4.5, we now conclude that $M$ is a flat $R$-module, as required. □

# Exercises

*Easy exercise* 6.1. Let $A$ be a torsion abelian group. Show that $A \otimes_{\mathbf{Z}} \mathbf{Q} = 0$.

*Easy exercise* 6.2. Let $V$ be a two-dimensional $\mathbf{C}$-vector space, with basis $\{v, w\}$. Show that the element $v \otimes w + w \otimes v$ is not an elementary tensor in $V \otimes_{\mathbf{C}} V$. *Hint: use the fact that $\{v \otimes v, v \otimes w, w \otimes v, w \otimes w\}$ is a basis for the 4-dimensional vector space $V \otimes_{\mathbf{C}} V$ (why?), and argue by contradiction.*

*Easy exercise* 6.3. Let $R$ be a ring, and let $M$ be an $R$-module. Denote by $M[X]$ the set of polynomial expressions in $X$ with "coefficients in $M$": expressions of the form $m_0 + m_1 X + \cdots + m_n X^n$ where $m_0, \cdots, m_n \in M$. Show that $M[X]$ is an $R[X]$-module in a natural way, and that $M[X] \cong M \otimes_R R[X]$.

*Easy exercise* 6.4. Let $R$ be a ring, let $S$ be a multiplicative subset of $R$. Let $M$ be an arbitrary $R$-module. Show that every element of $S^{-1}R \otimes_R M$ is an elementary tensor.

*Exercise* 6.5. Give an example of a non-zero abelian group $A$ such that $A \otimes_{\mathbf{Z}} A = 0$.

*Exercise* 6.6. Let $R$ be a ring, and let $(M_i)_{i \in I}$ be a collection of $R$-modules. Prove that $\bigoplus_{i \in I} M_i$ is a flat $R$-module if and only if $M_i$ is a flat $R$-module for all $i \in I$. Deduce that $R[X]$ is a flat $R$-algebra.

*Exercise* 6.7. Let $R$ be a ring and let $M$, $N$ be flat $R$-modules. Show that $M \otimes_R N$ is a flat $R$-module.

*Exercise* 6.8. Let $R$ be an integral domain. Show that any flat $R$-module is torsion free.

*Exercise* 6.9. Let $R$ be an integral domain, and let $M$ be an $R$-module. Let $\rho(M) = \dim_K(M \otimes_R K)$.

(a) If $R$ is a PID, show that $\rho(M)$ is equal to the rank of $M$, as defined in Theorem 4.1.5.

(b) Show that $M$ contains a free $R$-submodule of rank $\rho(M)$.

(c) If $N$ is an $R$-submodule of $M$, show that $\rho(M) = \rho(N) + \rho(M/N)$.

(d) Deduce from (c) that if $R^m \to R^n$ is an injective $R$-module homomorphism, then $m \leq n$.

*Exercise* 6.10. Let $R$ be a ring, let $S$ be an $R$-algebra. Given an $R$-module $M$ and an $S$-module $N$, construct a natural bijection $\mathrm{Hom}_S(S \otimes_R M, N) \to \mathrm{Hom}_R(M, N)$.

*Exercise* 6.11. Let $R$ be a ring and let $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$ be a *split* short exact sequence. Show that for an arbitrary $R$-module $N$, the induced sequence

$$0 \to M_1 \otimes_R N \xrightarrow{f \otimes \mathrm{Id}} M_2 \otimes_R N \xrightarrow{g \otimes \mathrm{Id}} M_3 \otimes_R N \to 0$$

is again exact.

*Exercise* 6.12. Let $G_n = \mathbf{Z}/2^n$ for all $n \geq 1$. Show that $\left(\prod_{n\geq 1} G_n\right) \otimes_{\mathbf{Z}} \mathbf{Q} \not\cong \prod_{n\geq 1}(G_n \otimes_{\mathbf{Z}} \mathbf{Q})$. Therefore tensor products do not commute with direct products in general; see Proposition 6.1.9.

*Exercise* 6.13. In this exercise we show that there exist a commutative ring $R$, $R$-modules $M$ and $N$, together with elements $m \in M$ and $n \in N$ which are *not* torsion elements, such that $m \otimes n \in M \otimes_R N$ is a torsion element. We sketch the construction; the assignment is to provide full details for all steps.

> *Let $R = \mathbf{C}[X,Y]/(X^2, XY, Y^2)$. Let $x$ and $y$ be the cosets of $X$ and $Y$ in $R$. Let $M$ be the $R$-module with generators $e_1$ and $e_2$ and relations $xe_1 = ye_2$ and $xe_2 = 0$. Let $N$ be the $R$-module with generators $f_1$ and $f_2$ and the relation $yf_1 = xf_2$. Then $e_1$ is not a torsion element in $M$, and the same holds for $f_1$ in $N$. However, a calculation shows that $x(e_1 \otimes f_1) = 0$ in $M \otimes_R N$. Hence $e_1 \otimes f_1$ is a torsion element.*

*Exercise* 6.14. Let $k$ be an arbitrary field. Let $R$ be the polynomial ring $k[X,Y]$, and let $\mathfrak{m} = (X,Y)$. The goal of this exercise is to compute the tensor product $\mathfrak{m} \otimes_R \mathfrak{m}$.

(a) Check that

$$0 \to R \xrightarrow{\psi} R \oplus R \xrightarrow{\varphi} \mathfrak{m} \to 0$$

is exact, where $\psi$ and $\varphi$ are defined by $\psi(f) = (fY, -fX)$ and $\varphi(g,h) = gX + hY$.

(b) Deduce from (a) that there exists a short exact sequence of the form

$$0 \to \mathfrak{m} \xrightarrow{\widetilde{\psi}} \mathfrak{m} \oplus \mathfrak{m} \xrightarrow{\widetilde{\varphi}} \mathfrak{m} \otimes_R \mathfrak{m} \to 0.$$

Give an explicit description for the morphisms $\widetilde{\psi}$ and $\widetilde{\varphi}$.

(c) Consider the $R$-module homomorphism $\mu : \mathfrak{m} \otimes_R \mathfrak{m} \to \mathfrak{m}$ given by $\mu(m \otimes m') = mm'$.

(1) Check that the image of $\mu$ is equal to $\mathfrak{m}^2 = (X^2, XY, Y^2)$.

(2) Check that $\mu \circ \widetilde{\varphi} = \varphi \circ \iota$, where $\iota : \mathfrak{m} \oplus \mathfrak{m} \to R \oplus R$ is the canonical injection.

(3) Deduce from this that the ker $\mu$ is a cyclic $R$-module, with generator $X \otimes Y - Y \otimes X$.

(4) Check that $X \otimes Y - Y \otimes X \neq 0$ in $\mathfrak{m} \otimes_R \mathfrak{m}$.

(5) Prove that $\mathrm{Ann}_R(X \otimes Y - Y \otimes X) = \mathfrak{m}$.

(d) Construct a homomorphism of $R$-modules $\lambda : \mathfrak{m}^2 \to \mathfrak{m} \otimes_R \mathfrak{m}$ such that $\mu \circ \lambda$ is the identity. (*Hint: it suffices to define $\lambda(X^2)$, $\lambda(XY)$ and $\lambda(Y^2)$, but be careful...*)

(e) Deduce that $\mathfrak{m} \otimes_R \mathfrak{m} = \ker \mu \oplus \mathrm{im}\, \lambda$ and hence $\mathfrak{m} \otimes_R \mathfrak{m} \cong \mathfrak{m}^2 \oplus R/\mathfrak{m}$.

*Hard exercise* 6.15. Let $R$ be a local ring. If $M$, $N$ are finitely generated $R$-modules with $M \otimes_R N = 0$, show that $M = 0$ or $N = 0$.

# Chapter 7

# Category theory

In many fields of mathematics, one studies an interesting class of objects, together with a class of maps with favourable properties between such objects. For example: in group theory, one studies groups, and functions between groups which respect the group structure, namely homomorphisms of groups; in topology, one studies topological spaces and maps between these spaces which respect the topological structure, namely continuous maps. The language of categories formalises the informal idea of a class of "interesting objects" together with a class of "maps with favourable properties" between these. This chapter serves as an introduction to this language, which is ubiquitous in pure mathematics.

## 7.1 Categories

The central notion of this chapter is the following.

**Definition 7.1.1.** A *category* $\mathcal{C}$ consists of the following data:

(1) a class $\mathrm{ob}(\mathcal{C})$ of *objects*,

(2) for each pair of objects $X, Y \in \mathrm{ob}(\mathcal{C})$, a class $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ of *morphisms* from $X$ to $Y$,

(3) for each object $X \in \mathrm{ob}(\mathcal{C})$, an *identity morphism* $1_X \in \mathrm{Hom}_{\mathcal{C}}(X, X)$,

(4) for each triple of objects $X, Y, Z \in \mathrm{ob}(\mathcal{C})$, a map

$$\mathrm{Hom}_{\mathcal{C}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}}(Y, Z) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(X, Z) : (f, g) \mapsto gf$$

called *composition*, satisfying the following requirements:

- given objects $X, Y, Z, T \in \mathrm{ob}(\mathcal{C})$ and morphisms $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, $g \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$ and $h \in \mathrm{Hom}_{\mathcal{C}}(Z, T)$, the equality $h(gf) = (hg)f$ holds ("composition is associative");
- given objects $X, Y \in \mathrm{ob}(\mathcal{C})$ and a morphism $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, we have $f1_X = f = 1_Y f$.

*Remark* 7.1.2. The use of the word "class" in the above definition is not innocent, but we will ignore the set-theoretic subtleties hidden in this word; see [Rotman, Introduction to §7.2] for some background.

**Example 7.1.3.** The category of sets, denoted by Set, is defined as follows: take $\mathrm{ob}(\mathsf{Set})$ to be the class of all sets, and let $\mathrm{Hom}_{\mathsf{Set}}(X, Y)$ be the set of all functions from $X$ to $Y$; the identity morphism $\mathrm{id}_X$ is simply the identity map on $X$, and composition of morphisms is the traditional composition of functions.

**Example 7.1.4.** The category Top of topological spaces is defined by taking $\mathrm{ob}(\mathsf{Top})$ to be the class of all topological spaces, and $\mathrm{Hom}_{\mathsf{Top}}(X, Y)$ the set of continuous maps from $X$ to $Y$. The identity morphism $\mathrm{id}_X$ is the identity map on $X$, and composition of morphisms is composition of continuous functions (recall from your topology course that the composition of continuous maps is again continuous!).

**Example 7.1.5.** Let $R$ be a ring. The category ${}_R\mathsf{Mod}$ of $R$-modules is defined by taking $\mathrm{ob}({}_R\mathsf{Mod})$ to be the class of all (left) $R$-modules. Morphisms in ${}_R\mathsf{Mod}$ are given by $\mathrm{Hom}_{{}_R\mathsf{Mod}}(M, N) = \mathrm{Hom}_R(M, N)$. The identity morphisms are the obvious ones, and composition of morphisms is simply composition of $R$-module homomorphisms (recall that the composition of two $R$-linear maps is again $R$-linear).

*Quick question* 7.1.6. Define the category Ring of (commutative, unital) rings, the category Grp of groups and the category Ab of abelian groups yourself, and give some other natural examples of categories.

In all of these examples of categories, the objects are sets with extra structure, and the morphisms are functions which “respect” this structure. This is the case for many commonly used categories, but certainly not for all of them, as the following examples will show.

**Example 7.1.7.** Let $S$ be a set. Then we can define a category $\mathcal{C}$ with $\mathrm{ob}(\mathcal{C}) = S$ and, for all $x, y \in S$, $\mathrm{Hom}_{\mathcal{C}}(x, y) = 1_x$ if $x = y$ and $\mathrm{Hom}_{\mathcal{C}}(x, y) = \emptyset$ if $x \neq y$. This is the *discrete category* indexed by $S$, in which the only morphisms are the identity morphisms.

**Example 7.1.8.** There is a category $\mathcal{C}$ with precisely two objects $X$ and $Y$, and with three morphisms: the identity morphisms $1_X$ and $1_Y$, and a map $f : X \to Y$. This category can be represented as follows:

$$\begin{tikzcd} X \arrow[loop left, "1_X"] \arrow[r, "f"] & Y \arrow[loop right, "1_Y"] \end{tikzcd}$$

**Example 7.1.9.** Any group can be regarded as a category with a single object. Indeed, if $G$ is a group, we have a category $\mathrm{B}G$ with $\mathrm{ob}(\mathrm{B}G) = \{\star\}$ and $\mathrm{Hom}_{\mathrm{B}G}(\star, \star) = G$, where composition of morphisms is multiplication in $G$, and where $1_\star$ is the identity element of $G$.

**Example 7.1.10.** Consider an arbitrary preordered set, i.e., a set $S$ equipped with a relation $\leq$ which is reflexive and transitive. Then there is a category $\mathcal{C}$ with $\mathrm{ob}(\mathcal{C}) = S$, in which morphisms are given by $\mathrm{Hom}_{\mathcal{C}}(x, y) = \{f_{(x,y)}\}$ if $x \leq y$, and $\mathrm{Hom}_{\mathcal{C}}(x, y) = \emptyset$ if $x \nleq y$. Composition of morphisms is given by $f_{(y,z)} \circ f_{(x,y)} = f_{(x,z)}$ whenever applicable (i.e., whenever $x \leq y \leq z$). This category is *thin*, which means that there is at most one morphism between every two objects of the category.

There are (at least) two easy ways to construct new categories from old ones.

**Definition 7.1.11.** Let $\mathcal{C}$ be a category. The *dual* or *opposite* category of $\mathcal{C}$, denoted by $\mathcal{C}^{\mathrm{opp}}$, is the category obtained from $\mathcal{C}$ by “reversing arrows”. This means that $\mathrm{ob}(\mathcal{C}^{\mathrm{opp}}) = \mathrm{ob}(\mathcal{C})$ and that morphisms are given by $\mathrm{Hom}_{\mathcal{C}^{\mathrm{opp}}}(X, Y) = \mathrm{Hom}_{\mathcal{C}}(Y, X)$, with composition done “the other way around”:

$$\mathrm{Hom}_{\mathcal{C}^{\mathrm{opp}}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}^{\mathrm{opp}}}(Y, Z) \longrightarrow \mathrm{Hom}_{\mathcal{C}^{\mathrm{opp}}}(X, Z) : (f, g) \mapsto fg$$

where $f : Y \to X$, $g : Z \to Y$ and $fg : Z \to X$ are morphisms in $\mathcal{C}$.

*Quick question* 7.1.12. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. Explain how to construct a *product category* $\mathcal{C} \times \mathcal{D}$ with objects pairs of the form $(X, Y)$, where $X \in \mathrm{ob}(\mathcal{C})$ and $Y \in \mathrm{ob}(\mathcal{D})$, and for which

$$\mathrm{Hom}_{\mathcal{C}\times\mathcal{D}}((X, Y), (X', Y')) = \mathrm{Hom}_{\mathcal{C}}(X, X') \times \mathrm{Hom}_{\mathcal{D}}(Y, Y').$$

We will now define special types of objects and morphisms in categories, treating morphisms first. In general it makes no sense whatsoever to say that a morphism in a category is bijective, injective or surjective, since the objects of a category are not even sets in general. However, there are special kinds of morphisms that behave very much like bijective, injective and surjective maps.

**Definition 7.1.13.** Let $f : X \to Y$ be a morphism in a category $\mathcal{C}$. Then $f$ is said to be an *isomorphism* if there exists a morphism $g : Y \to X$ in $\mathcal{C}$ such that $gf = 1_X$ and $fg = 1_Y$.

**Example 7.1.14.** An isomorphism in Set is a bijective function. An isomorphism in Grp is a group isomorphism; similarly for Ring. An isomorphism in Top is a homeomorphism; note that this is stronger than being a continuous bijection! Finally, if $G$ is a group, then *every* morphism in B$G$ is an isomorphism.

**Definition 7.1.15.** Let $f : X \to Y$ be a morphism in a category $\mathcal{C}$. Then $f$ is said to be a *monomorphism* (or sometimes also *monic*) if for every object $W$ and every pair of morphisms $g_1, g_2 : W \to X$ with $fg_1 = fg_2$, we have $g_1 = g_2$. Moreover, $f$ is said to be an *epimorphism* (or *epic*) if for every object $Z$ and every pair of morphisms $h_1, h_2 : Y \to Z$ with $h_1 f = h_2 f$, we have $h_1 = h_2$.

**Example 7.1.16.** In Set, the monomorphism are the injections, and the epimorphism are the surjections.

It is clear that an injection is a monomorphism. Conversely, let $f : X \to Y$ be a monomorphism. If $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$, let $W = \{\star\}$ and consider $g_i : W \to X : \star \mapsto x_i$ for $i \in \{1, 2\}$. Then clearly $fg_1 = fg_2$ and therefore $g_1 = g_2$, i.e., $x_1 = x_2$, showing that $f$ is injective.

It is clear that a surjection is an epimorphism. Conversely, let $f : X \to Y$ be an epimorphism. If $f$ is not surjective, pick $y_0 \in Y \setminus f(X)$ and set $Z = \{0, 1\}$. Define $g_1 : Y \to Z$ by $g_1(y) = 0$ for all $y \in Y$, and $g_2 : Y \to Z$ by $g_2(y_0) = 1$ and $g_2(y) = 0$ if $y \neq y_0$. Then $g_1 f = g_2 f$, but clearly $g_1 \neq g_2$; contradiction! We conclude that $f$ is surjective.

*Remark* 7.1.17. A similar argument shows that the monomorphisms in Grp are precisely the injective group homomorphisms. It is also true that epimorphisms in Grp are precisely the surjective group homomorphisms, but proving this is quite a bit harder – see Exercise 7.28.

**Example 7.1.18.** Let $R$ be a ring. Then the monomorphisms in ${}_R$Mod are the injective $R$-module homomorphisms, and the epimorphisms in ${}_R$Mod are the surjective $R$-module homomorphisms.

Indeed, if $f : M \to N$ is a monomorphism, consider the embedding $\iota : \ker f \hookrightarrow M$ and the zero map $0 : \ker f \to M$. Then $f \circ \iota = f \circ 0$ and therefore $\iota = 0$, in other words: $f$ is injective. A similar argument (now using the cokernel) proves the statement about epimorphisms.

Let us now look at special objects, rather than special morphisms:

**Definition 7.1.19.** Let $\mathcal{C}$ be a category. An object $X$ of $\mathcal{C}$ is *initial* if for every object $Y$, there is a unique morphism $X \to Y$ in $\mathcal{C}$. Moreover, $X$ is *final* if for every object $Y$, there is a unique morphism $Y \to X$ in $\mathcal{C}$. Finally, $X$ is a *zero object* in $\mathcal{C}$ if $X$ is both initial and final.

**Example 7.1.20.** In Set, the empty set $\emptyset$ is initial and a singleton $\{\star\}$ is final; the same is true for Top. In Grp, the trivial group $\{1\}$ is a zero object. In ${}_R$Mod, the zero module is a zero object. In Ring, the ring of integers $\mathbf{Z}$ is initial, whereas the zero ring 0 is final.

**Example 7.1.21.** Let $R$ be a ring, and let $M$ and $N$ be $R$-modules. The category Bil$(M, N, -)$ is defined as follows: objects are pairs $(P, f)$ consisting of an $R$-module $P$ and an $R$-bilinear map $f : M \times N \to P$; morphisms from $(P, f)$ to $(Q, g)$ are $R$-module homomorphisms $\varphi : P \to Q$ such that $\varphi \circ f = g$. Then Bil$(M, N, -)$ has an initial object, namely $(M \otimes_R N, \otimes)$. (Does Bil$(M, N, -)$ have a final object?)

Nothing guarantees the existence of initial or final objects in a given category, but if they do exist, they are unique in the following strong sense:

**Proposition 7.1.22.** *Final and initial objects in a category are unique up to unique isomorphism.*

*Proof.* We will prove the statement about final objects (the proof for initial objects is very similar). Let $X_1$ and $X_2$ be two final objects in a category $\mathcal{C}$. Since $X_2$ is final, there exists a unique map $f : X_1 \to X_2$, and since $X_1$ is final, there exists a unique map $g : X_2 \to X_1$. Consider the composition $gf : X_1 \to X_1$. Since $X_1$ is final, there exists a unique map $X_1 \to X_1$; however, we already know one such map, namely $1_{X_1}$, so that $gf = 1_{X_1}$. Similarly, we obtain $fg = 1_{X_2}$. It follows that $f$ and $g$ are isomorphisms. □

*Quick question* 7.1.23. Let $\mathcal{C}$ be a category which has a zero object. Given two objects $X$ and $Y$ in $\mathcal{C}$, how would you define a "zero morphism" from $X$ to $Y$?

## 7.2 Functors

What makes category theory such a powerful language is the fact that we do not only have morphisms between objects in a given category, but also "morphisms of categories"; this is where *functors* come in.

**Definition 7.2.1.** Let $\mathcal{C}$, $\mathcal{D}$ be categories. A *covariant functor* $F : \mathcal{C} \to \mathcal{D}$ consists of the following data:

(a) for every object $X$ in $\mathcal{C}$, an object $F(X)$ in $\mathcal{D}$;

(b) for every morphism $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, a morphism $F(f) \in \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$,

such that

- for every object $X$ in $\mathcal{C}$, we have $F(1_X) = 1_{F(X)}$,
- given $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ and $g \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$, we have $F(gf) = F(g)F(f)$ in $\mathrm{Hom}_{\mathcal{D}}(F(X), F(Z))$.

To make the notation a bit lighter, we will often write $FX$ and $Ff$ rather than $F(X)$ and $F(f)$; whenever we say *functor*, we mean a *covariant* functor as defined above.

*Quick question* 7.2.2. Given categories $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ and functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{E}$, check that $GF : \mathcal{C} \to \mathcal{E}$ defined by $(GF)(X) = G(F(X))$ and $(GF)(f) = G(F(f))$ is again a functor.

Let us present a bunch of examples.

**Example 7.2.3.** Let $\mathcal{C}$ be a category. The *identity functor* $1_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ is the covariant functor which leaves all objects and morphisms in $\mathcal{C}$ unchanged. (Obviously this functor is extremely boring. . . )

**Example 7.2.4.** Let $R$ be a ring. There is a functor ${}_R\mathsf{Mod} \to \mathsf{Ab}$ which sends an $R$-module $M$ to the abelian group $M$, and an $R$-module homomorphism $f : M \to N$ to the same homomorphism, now seen as a group homomorphism. In other words: this functor "forgets" the $R$-module structure and remembers only the group structure. Similarly, we have forgetful functors $\mathsf{Top} \to \mathsf{Set}$ (forgetting the topology), $\mathsf{Ring} \to \mathsf{Ab}$ (forgetting the multiplication), $\mathsf{Grp} \to \mathsf{Set}$ (forgetting the group structure), et cetera.

**Example 7.2.5.** Given a group $G$, the *abelianisation* $G^{\mathsf{ab}}$ of $G$ is $G^{\mathsf{ab}} = G/[G, G]$, where $[G, G]$ denotes the subgroup of $G$ generated by all *commutators*: elements of the form $xyx^{-1}y^{-1}$ where $x, y \in G$. If $f : G \to H$ is a morphism in the category $\mathsf{Grp}$, then $f([G, G]) \subseteq [H, H]$; therefore $f$ induces a morphism $f^{\mathsf{ab}} : G^{\mathsf{ab}} \to H^{\mathsf{ab}}$ in $\mathsf{Ab}$. It is an easy exercise to check that these constructions together yield a well-defined functor $\mathsf{ab} : \mathsf{Grp} \to \mathsf{Ab}$, the so-called *abelianisation functor*.

For example, let $\mathcal{S}_3$ be the symmetric group on a set of 3 elements, then we have $\mathcal{S}_3^{\mathsf{ab}} \cong \mathbf{Z}/2$. Fix an embedding $\iota : \mathbf{Z}/3 \hookrightarrow \mathcal{S}_3$, then $\iota^{\mathsf{ab}}$ is the zero morphism $0 : \mathbf{Z}/3 \to \mathbf{Z}/2$. On the other hand, the abelianisation of the identity $\mathrm{Id} : \mathbf{Z}/3 \to \mathbf{Z}/3$ is again $\mathrm{Id} : \mathbf{Z}/3 \to \mathbf{Z}/3$.

**Example 7.2.6.** Let $G$ and $H$ be groups; let $\mathrm{B}G$ and $\mathrm{B}H$ be the associated categories with a single object, see Example 7.1.9. Then a functor $\mathrm{B}G \to \mathrm{B}H$ is nothing but a group homomorphism $G \to H$.

**Example 7.2.7.** Let $S$ and $T$ be pre-ordered sets; let $\mathcal{C}_S$ and $\mathcal{C}_T$ be the associated categories, in the sense of Example 7.1.10. Then a functor $\mathcal{C}_S \to \mathcal{C}_T$ is nothing but an order-preserving function $S \to T$.

**Example 7.2.8.** Let $R$ be a ring. We have a functor $\mathsf{Free} : \mathsf{Set} \to {}_R\mathsf{Mod}$ which sends a set $S$ to the free $R$-module $R^{(S)}$ with basis $S$. Given a morphism $f : S \to T$ in $\mathsf{Set}$, we get an induced morphism $\mathsf{Free}(f) : R^{(S)} \to R^{(T)}$ in ${}_R\mathsf{Mod}$, simply by sending the standard basis vector of $R^{(S)}$ corresponding to $s \in S$ to the standard basis vector of $R^{(T)}$ corresponding to $f(s) \in T$ (this defines $\mathsf{Free}(f)$ completely).

**Example 7.2.9.** Let $R$ be a ring. If $M$ is an $R$-module, there is a functor $M \otimes_R - : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}$ which sends an $R$-module $N$ to $M \otimes_R N$, and a homomorphism of $R$-modules $f : N_1 \to N_2$ to the induced homomorphism $\mathrm{Id}_M \otimes f : M \otimes_R N_1 \to M \otimes_R N_2$; the functoriality follows from equality (6.1).

**Example 7.2.10.** Next to the forgetful functor $\mathsf{Ring} \to \mathsf{Grp}$ which forgets the multiplication, there is another natural functor $\mathsf{Ring} \to \mathsf{Grp}$: consider for each ring $R$ the set of units $R^\times$, which is a group under multiplication, and for each homomorphism of rings $f : R \to S$ the restriction $f^\times : R^\times \to S^\times$.

The following example is slightly more sophisticated.

**Example 7.2.11.** Let $\mathsf{Top}_\star$ be the category of pointed topological spaces: pairs $(X, x_0)$ consisting of a topological space $X$ together with a base point $x_0 \in X$. A morphism $(X, x_0) \to (Y, y_0)$ in $\mathsf{Top}_\star$ is simply a continuous map $f : X \to Y$ for which $f(x_0) = y_0$. There exists a functor $\pi_1 : \mathsf{Top}_\star \to \mathsf{Grp}$ which sends a pointed topological space $(X, x_0)$ to its *fundamental group* $\pi_1(X, x_0)$, and which sends a morphism $(X, x_0) \to (Y, y_0)$ in $\mathsf{Top}_\star$ to a group homomorphism $\pi_1(X, x_0) \to \pi_1(Y, y_0)$.

The next two examples are fundamental and involve very general categories.

**Example 7.2.12.** Let $\mathcal{C}$ be a category such that for all $X, Y \in \mathrm{ob}(\mathcal{C})$, the class $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a set. Given $X \in \mathrm{ob}(\mathcal{C})$, define a functor $F : \mathcal{C} \to \mathsf{Set}$ as follows: given $Y \in \mathrm{ob}(\mathcal{C})$, let $F(Y) = \mathrm{Hom}_{\mathcal{C}}(X, Y)$; given $f \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$, define $F(f) : \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{C}}(X, Z)$ by $F(f)(g) = fg$. It is not hard to check that $F$ is a functor, which we will denote by $\mathrm{Hom}_{\mathcal{C}}(X, -)$ in the sequel.

**Example 7.2.13.** Let $I$ be a preordered set (with order relation $\leq$). Let $\mathcal{I}$ be the thin category constructed from $I$ (see Example 7.1.10). Given a category $\mathcal{C}$, a functor $F : \mathcal{I} \to \mathcal{C}$ consists of the following data:

- for every $i \in I$, an object $X_i$ in $\mathcal{C}$;
- for every pair $(i, j) \in I^2$ with $i \leq j$, a morphism $f_{ij} \in \mathrm{Hom}_{\mathcal{C}}(X_i, X_j)$ such that $f_{ii} = 1_{X_i}$ and

$$f_{ik} = f_{jk} \circ f_{ij} \;\text{ for all }\; (i, j, k) \in I^3 \;\text{ such that }\; i \leq j \leq k.$$

We call such a functor a *diagram* or *inductive system* in $\mathcal{C}$ indexed by $I$; the $f_{ij}$ are called *transition maps*.

In fact it is possible to define a notion of diagram in a category $\mathcal{C}$ indexed by an arbitrary category $\mathcal{I}$, rather than a category associated to a preordered set; see this link for more details.

Next to covariant functors, we also have their contravariant counterparts:

**Definition 7.2.14.** Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *contravariant functor* from $\mathcal{C}$ to $\mathcal{D}$ is a covariant functor $F : \mathcal{C}^{\mathrm{opp}} \to \mathcal{D}$. In concrete terms, this means that $F$ consists of the following data:

(a) for every object $X$ in $\mathcal{C}$, an object $F(X)$ in $\mathcal{D}$;

(b) for every morphism $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, a morphism $F(f) \in \mathrm{Hom}_{\mathcal{D}}(F(Y), F(X))$,

such that

- for every object $X$ in $\mathcal{C}$, we have $F(1_X) = 1_{F(X)}$,
- given $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ and $g \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$, we have $F(gf) = F(f)F(g)$ in $\mathrm{Hom}_{\mathcal{D}}(F(Z), F(X))$.

In other words: the only difference with the notion of a covariant functor is the fact that a contravariant functor *reverses* the order of composition, whereas a covariant functor preserves the order.

**Example 7.2.15.** Let $\mathcal{C}$ be a category such that for all $X, Y \in \mathrm{ob}(\mathcal{C})$, the class $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a set. Given $X \in \mathrm{ob}(\mathcal{C})$, define a functor $F : \mathcal{C} \to \mathsf{Set}$ as follows: given $Y \in \mathrm{ob}(\mathcal{C})$, let $F(Y) = \mathrm{Hom}_{\mathcal{C}}(Y, X)$; given $f \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$, define $F(f) : \mathrm{Hom}_{\mathcal{C}}(Z, X) \to \mathrm{Hom}_{\mathcal{C}}(Y, X)$ by $F(f)(g) = gf$. It is not hard to check that $F$ is a *contravariant* functor, which we will denote by $\mathrm{Hom}_{\mathcal{C}}(-, X)$.

**Example 7.2.16.** Let $k$ be a field, and let $\mathsf{Vec}_k$ be the category of vector spaces over $k$. There is a contravariant functor from $\mathsf{Vec}_k$ to itself, which sends a $k$-vector space $V$ to its dual $V^\vee$, and a $k$-linear map $f : V \to W$ to the induced map $f^\vee : W^\vee \to V^\vee : \varphi \mapsto \varphi \circ f$.

Finally, it is sometimes useful to consider functors with multiple arguments.

**Definition 7.2.17.** Let $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{D}$ be categories, and let $\mathcal{C}_1 \times \mathcal{C}_2$ be the product of $\mathcal{C}_1$ and $\mathcal{C}_2$, as defined in Quick Question 7.1.12. A functor $F : \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{D}$ assigns to a pair of objects $(X_1, X_2)$, with $X_1 \in \mathrm{ob}(\mathcal{C}_1)$ and $X_2 \in \mathrm{ob}(\mathcal{C}_2)$, an object $F(X_1, X_2)$ in $\mathcal{D}$, and to a pair of morphisms $(f_1, f_2)$, with $f_1 \in \mathrm{Hom}_{\mathcal{C}}(X_1, Y_1)$ and $f_2 \in \mathrm{Hom}_{\mathcal{C}}(X_2, Y_2)$, a morphism $F(f_1, f_2) \in \mathrm{Hom}_{\mathcal{D}}(F(X_1, X_2), F(Y_1, Y_2))$.

This can be combined with the notion of a covariant functor:

**Example 7.2.18.** Let $\mathcal{C}$ be a category such that for all objects $X$ and $Y$ in $\mathcal{C}$, the class $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a set. We have a functor $\mathrm{Hom}_{\mathcal{C}}(-, -) : \mathcal{C}^{\mathrm{opp}} \times \mathcal{C} \to \mathsf{Set}$: we send a pair $(X, Y) \in \mathrm{ob}(\mathcal{C}^{\mathrm{opp}} \times \mathcal{C})$ to $\mathrm{Hom}_{\mathcal{C}}(X, Y)$, and given morphisms $f : X' \to X$ and $g : Y \to Y'$ in $\mathcal{C}$, we send $(f, g)$ to

$$\mathrm{Hom}_{\mathcal{C}}(f, g) : \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{C}}(X', Y') : \alpha \mapsto g\alpha f.$$

We say that $\mathrm{Hom}_{\mathcal{C}}(-, -)$ is contravariant in the first argument, and covariant in the second argument.

**Example 7.2.19.** Let $R$ be a ring. We have a functor $- \otimes_R - : {}_R\mathsf{Mod} \times {}_R\mathsf{Mod} \to \mathsf{Set}$ which sends a pair of $R$-modules $(M, N)$ to $M \otimes_R N$, and a pair of $R$-module homomorphisms $(f : M \to M', g : N \to N')$ to $f \otimes g : M \otimes_R N \to M' \otimes_R N'$. This is an example of a so-called *bifunctor*.

## 7.3 Natural transformations

Now that we have defined functors as “morphisms between categories”, it will be useful to develop a notion of “morphisms between functors” as well; this brings us to the concept of a *natural transformation*.

**Definition 7.3.1.** Let $\mathcal{C}$ and $\mathcal{D}$ be categories, and let $F, G : \mathcal{C} \to \mathcal{D}$ be functors. A *natural transformation* $\eta$ from $F$ to $G$ (denoted by $\eta : F \Rightarrow G$) consists of the following data:

- for every object $X$ in $\mathcal{C}$, a morphism $\eta_X : FX \to GX$ in $\mathcal{D}$,

subject to the requirement that

- for every morphism $f : X \to Y$ in $\mathcal{C}$, the following square in $\mathcal{D}$ commutes:

$$\begin{tikzcd} FX \arrow[r, "Ff"] \arrow[d, "\eta_X"] & FY \arrow[d, "\eta_Y"] \\ GX \arrow[r, "Gf"] & GY \end{tikzcd} .$$

We say that $\eta$ is a *natural isomorphism* if $\eta_X$ is an isomorphism in $\mathcal{D}$, for every object $X$ in $\mathcal{C}$.

**Example 7.3.2.** Let $k$ be a field. If $V$ is a $k$-vector space, we have a *natural* map $\eta_V : V \to V^{\vee\vee}$ from $V$ to its double dual, obtained by sending $v \in V$ to $\mathrm{ev}_v : V^\vee \to k : f \mapsto f(v)$. The word “natural” is often used informally, meaning roughly “independent of the choice of a basis”, but it has a deeper meaning.

Indeed, the collection $(\eta_V)_V$, where $V$ ranges over all $k$-vector spaces, is a natural transformation from the identity functor $\mathrm{Id} : \mathsf{Vec}_k \to \mathsf{Vec}_k$ to the double dual functor $(-)^{\vee\vee} : \mathsf{Vec}_k \to \mathsf{Vec}_k$.

This means that if $f : V \to W$ is a morphism in the category $\mathsf{Vec}_k$ of $k$-vector spaces, then

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

commutes. This is not a natural isomorphism; however, if we restrict Id and $(-)^{\vee\vee}$ to the subcategory $\mathsf{fVec}_k$ of *finite dimensional* $k$-vector spaces, then the restriction of $\eta$ becomes a natural isomorphism.

**Example 7.3.3.** Let $\mathcal{C}$ be the category with objects all finite sets, and morphisms bijections between finite sets. Let us define functors $\mathrm{Perm} : \mathcal{C} \to \mathsf{Set}$ and $\mathrm{Ord} : \mathcal{C} \to \mathsf{Set}$, as follows. The functor Perm sends an object $X$ in $\mathcal{C}$ to the set of permutations on $X$, and a morphism $f : X \to Y$ in $\mathcal{C}$ to the morphism $\mathrm{Perm}(f) : \mathrm{Perm}(X) \to \mathrm{Perm}(Y) : g \mapsto fgf^{-1}$ in $\mathsf{Set}$. The functor Ord sends an object $X$ in $\mathcal{C}$ to the set of total orders on $X$, which can be thought of as sequences of inequalities $x_1 < \cdots < x_n$, where $X = \{x_1, \cdots, x_n\}$. A morphism $f : X \to Y$ in $\mathcal{C}$ induces $\mathrm{Ord}(f) : \mathrm{Ord}(X) \to \mathrm{Ord}(Y)$ in $\mathsf{Set}$, simply by mapping a total order $x_1 < \cdots < x_n$ to the total order $f(x_1) < \cdots < f(x_n)$.

The functors Perm and Ord are pointwise isomorphic: this means that for every object $X$ in $\mathcal{C}$, we have $\mathrm{Perm}(X) \cong \mathrm{Ord}(X)$ in $\mathsf{Set}$. However, these functors are not *naturally* isomorphic. Indeed, assume that $\eta : \mathrm{Perm} \Rightarrow \mathrm{Ord}$ is a natural transformation, then

$$\eta_Y \circ \mathrm{Perm}(f) = \mathrm{Ord}(f) \circ \eta_X$$

as maps from $\mathrm{Perm}(X)$ to $\mathrm{Ord}(Y)$, for every morphism $f : X \to Y$ in $\mathcal{C}$. Evaluating this equality on $\mathrm{Id}_X$ (the trivial permutation on $X$) yields $\eta_Y(\mathrm{Id}_Y) = \mathrm{Ord}(f)(\eta_X(\mathrm{Id}_X))$, for all $f : X \to Y$ in $\mathcal{C}$. However, if we take $X = Y = \{a, b\}$ and if $f : X \to Y$ is given by $f(a) = b$ and $f(b) = a$, then we get a contradiction (indeed, the order on the right hand side is "opposite" to the one on the left hand side).

**Example 7.3.4.** Let $G$ and $H$ be groups. We have seen that a functor $F : \mathrm{B}G \to \mathrm{B}H$ is nothing but a group homomorphism $G \to H$ (Example 7.2.6). Given two functors $F_1, F_2 : \mathrm{B}G \to \mathrm{B}H$ corresponding to homomorphisms $f_1, f_2 : G \to H$, natural transformations from $F_1$ to $F_2$ can be described in concrete terms: a natural transformation $\eta : F_1 \Rightarrow F_2$ is nothing but a *conjugation* $f_2 = cf_1c^{-1}$, where $c \in H$ is a fixed element. Of course such a conjugation need not exist for all pairs of homomorphisms $(f_1, f_2)$.

**Example 7.3.5.** Let $S$ and $T$ be pre-ordered sets, with associated categories $\mathcal{C}_S$ and $\mathcal{C}_T$ (Example 7.1.10). Then a functor $\mathcal{C}_S \to \mathcal{C}_T$ is nothing but an order preserving function $S \to T$ (see (Example 7.2.7). If $F, G : \mathcal{C}_S \to \mathcal{C}_T$ are functors corresponding to order preserving functions $f, g : S \to T$, then a natural transformation $\eta : F \Rightarrow G$ is simply an inequality $f \leq g$ on objects, i.e., $f(s) \leq g(s)$ in $T$ for all $s \in S$.

We will now define the notion of *equivalence* of categories, which tells us whether two categories are "essentially the same". The first attempt to define such a notion could be to say that categories $\mathcal{C}$ and $\mathcal{D}$ are *isomorphic* if and only if there exist functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ such that $GF = 1_{\mathcal{C}}$ and $FG = 1_{\mathcal{D}}$. However, this definition is way too rigid to be practical, as shown by the following example.

**Example 7.3.6.** Let $\mathcal{C}$ be the category with a unique object $\star$ and a unique morphism $1_\star$. Let $\mathcal{D}$ be the category with two objects $X$ and $Y$, and four morphisms: $1_X$, $1_Y$, $f : X \to Y$ and $g : Y \to X$. Then $f$ and $g$ are isomorphisms and therefore one would like to say that $\mathcal{C}$ and $\mathcal{D}$ are equivalent, since $\mathcal{D}$ consists of "two copies of $\star$". However, $\mathcal{C}$ and $\mathcal{D}$ are not isomorphic in the sense of the definition given above.

The following definition has the necessary flexibility built in.

**Definition 7.3.7.** Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A functor $F : \mathcal{C} \to \mathcal{D}$ is an *equivalence of categories* if there exists a functor $G : \mathcal{D} \to \mathcal{C}$ together with a pair of natural isomorphisms

$$\varepsilon : FG \overset{\sim}{\Longrightarrow} 1_{\mathcal{D}},\ \eta : GF \overset{\sim}{\Longrightarrow} 1_{\mathcal{C}}.$$

Given $F : \mathcal{C} \to \mathcal{D}$, a functor $G : \mathcal{D} \to \mathcal{C}$ with these properties is called a *quasi-inverse* of $F$. The categories $\mathcal{C}$ and $\mathcal{D}$ are said to be *equivalent* if there exists an equivalence $F : \mathcal{C} \to \mathcal{D}$.

*Quick question* 7.3.8. Show that equivalence of categories defines an equivalence relation.

Constructing a quasi-inverse for an equivalence is generally not easy in concrete situations. Therefore we will now develop an alternative and rather explicit criterion for a functor to be an equivalence of categories. To formulate this criterion, we need some new terminology:

**Definition 7.3.9.** Let $F : \mathcal{C} \to \mathcal{D}$ be a functor. We say that $F$ is

(a) *full* if for all objects $X$ and $Y$ in $\mathcal{C}$, the map $\mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(FX, FY)$ is surjective;

(b) *faithful* if for all objects $X$ and $Y$ in $\mathcal{C}$, the map $\mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(FX, FY)$ is injective;

(c) *fully faithful* if $F$ is both full and faithful.

*Quick question* 7.3.10. Let $F : \mathcal{C} \to \mathcal{D}$ be a fully faithful functor. Show that a morphism $f$ in $\mathcal{C}$ is an isomorphism if and only if $Ff$ is an isomorphism in $\mathcal{D}$. Show also that two objects $X$ and $Y$ in $\mathcal{C}$ are isomorphic if and only if $FX$ and $FY$ are isomorphic in $\mathcal{D}$.

**Definition 7.3.11.** A functor $F : \mathcal{C} \to \mathcal{D}$ is *essentially surjective* if and only if for every object $Z$ in $\mathcal{D}$, there exists an object $X$ in $\mathcal{C}$ such that $FX$ and $Z$ are isomorphic in $\mathcal{D}$.

Our criterion is now very simple to state:

**Theorem 7.3.12.** *A functor is an equivalence if and only if it is fully faithful and essentially surjective.*

*Proof.* We leave it as a simple exercise for the reader to show that if the functor $F : \mathcal{C} \to \mathcal{D}$ is an equivalence of categories, then $F$ is fully faithful and essentially surjective.

Conversely, assume that $F$ is both fully faithful and essentially surjective; let us construct a quasi-inverse $G : \mathcal{D} \to \mathcal{C}$. Essential surjectivity means that for every object $X$ in $\mathcal{D}$, there exists an object $GX$ in $\mathcal{C}$ together with an isomorphism $\alpha_X : F(GX) \to X$. Let us choose such a pair $(GX, \alpha_X)$ for every object $X$ in $\mathcal{D}$; the goal is to turn the map $X \mapsto GX$ (defined on objects) into a full-fledged functor.

Given $f \in \mathrm{Hom}_{\mathcal{D}}(X, Y)$, we have an induced morphism $\alpha_Y^{-1} f \alpha_X \in \mathrm{Hom}_{\mathcal{D}}(F(GX), F(GY))$. Since the natural map $\mathrm{Hom}_{\mathcal{C}}(GX, GY) \to \mathrm{Hom}_{\mathcal{D}}(F(GX), F(GY))$ is bijective, there exists a unique pre-image for $\alpha_Y^{-1} f \alpha_X$ under this map, which we denote by $Gf$, so that $F(Gf) = \alpha_Y^{-1} f \alpha_X$. It is not hard to check that this construction indeed defines a functor $G : \mathcal{D} \to \mathcal{C}$ (exercise for the reader).

Finally, we observe that $G$ is a quasi-inverse for $F$; indeed, the maps $\alpha_X$ together define a natural isomorphism $\alpha : FG \overset{\sim}{\Longrightarrow} 1_{\mathcal{D}}$. To construct a natural isomorphism $\beta : GF \overset{\sim}{\Longrightarrow} 1_{\mathcal{C}}$, let $\beta_X : GFX \to X$ be the unique pre-image of $\alpha_{FX}$ under the bijection $\mathrm{Hom}_{\mathcal{C}}(GFX, X) \to \mathrm{Hom}_{\mathcal{D}}(FGFX, FX)$. Then $\beta_X$ is again an isomorphism, and it is not hard to check that $\beta$ obtained in this way is natural. □

For a more interesting example of an equivalence of categories than Example 7.3.6, see Exercise 7.12.

## 7.4 Representability

**Disclaimer.** All categories in this section will be assumed to be *locally small*, which means that the class of morphisms between any two objects of the category is in fact a set.

Let $\mathcal{C}$ be a category. For any object $X$ in $\mathcal{C}$, we have the functors

$$h^X := \mathrm{Hom}_{\mathcal{C}}(X, -) : \mathcal{C} \to \mathsf{Set}, \quad h_X := \mathrm{Hom}_{\mathcal{C}}(-, X) : \mathcal{C}^{\mathrm{opp}} \to \mathsf{Set}$$

defined in Examples 7.2.12 and 7.2.15.

**Theorem 7.4.1** (Yoneda's lemma)**.** *Let $X$ be an object in $\mathcal{C}$. If $F : \mathcal{C} \to \mathsf{Set}$ is a covariant functor, there is a bijection between the set of natural transformations $\mathrm{Nat}(h^X, F)$ from $h^X$ to $F$, and the set $F(X)$. Similarly, if $F : \mathcal{C}^{\mathrm{opp}} \to \mathsf{Set}$ is covariant, there is a bijection between $\mathrm{Nat}(h_X, F)$ and $F(X)$.*

*Proof.* Let $\eta : h^X \Rightarrow F$ be a natural transformation. Then $\eta$ defines an element of $F(X)$, as follows: $\eta$ induces a morphism $\eta_X : h^X(X) = \mathrm{Hom}_{\mathcal{C}}(X, X) \to F(X)$ in $\mathsf{Set}$, and hence we get the "distinguished" element $\eta_X(1_X) \in F(X)$. Conversely, we will prove that for any $\alpha \in F(X)$, there exists exactly one natural transformation $\eta : h^X \Rightarrow F$ such that $\eta_X(1_X) = \alpha$; this will clearly suffice to prove the theorem.

Defining $\eta$ means defining for every object $Y$ in $\mathcal{C}$ a morphism $\eta_Y : h^X(Y) \to F(Y)$ in $\mathsf{Set}$ such that if $f : Y \to Z$ is a morphism in $\mathcal{C}$, then the following diagram commutes:

$$\begin{tikzcd}
h^X(Y) \arrow[r, "h^X(f)"] \arrow[d, "\eta_Y"] & h^X(Z) \arrow[d, "\eta_Z"] \\
FY \arrow[r, "Ff"] & FZ
\end{tikzcd}$$

We claim that fixing $\eta_X(1_X)$ determines $\eta$ (if it exists) completely. Indeed, taking $Y = X$ in the above diagram and choosing $f \in h^X(Z)$ arbitrarily, we get $(\eta_Z \circ h^X(f))(1_X) = (Ff \circ \eta_X)(1_X)$, or equivalently, $\eta_Z(f) = Ff(\eta_X(1_X))$. Hence the "value" of $\eta_X(1_X)$ determines all of $\eta$.

What remains to be done is showing that choosing $\eta_Z(f) = Ff(\eta_X(1_X))$ consistently – that is, for all possible $Z$ and all possible $f \in h^X(Z)$ – indeed yields a natural transformation. To check this, we need to show that if $g : Y \to Z$ is a morphism in $\mathcal{C}$, then the diagram

$$\begin{tikzcd}
h^X(Y) \arrow[r, "h^X(g)"] \arrow[d, "\eta_Y"] & h^X(Z) \arrow[d, "\eta_Z"] \\
FY \arrow[r, "Fg"] & FZ
\end{tikzcd}$$

commutes. Indeed, if $f \in h^X(Y)$, then we have

$$(\eta_Z \circ h^X(g))(f) = \eta_Z(gf) = F(gf)(\eta_X(1_X)) = (Fg \circ Ff)(\eta_X(1_X)) = (Fg \circ \eta_Y)(f).$$

The proof in the contravariant case proceeds along similar lines. □

**Corollary 7.4.2.** *Let $\mathcal{C}$ be a category. If $X$ and $Y$ are objects in $\mathcal{C}$, then we have bijections*

$$\mathrm{Nat}(h^X, h^Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(Y, X), \;\; \mathrm{Nat}(h_X, h_Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, Y).$$

*Given $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, we denote by*

$$h^f \in \mathrm{Nat}(h^Y, h^X), \;\; h_f \in \mathrm{Nat}(h_X, h_Y)$$

*the corresponding natural transformations.*

Using Yoneda's lemma, we can construct the so-called *Yoneda embedding*.

**Theorem 7.4.3.** *Let $\mathcal{C}$ be a category. Denote by $\mathsf{Fun}(\mathcal{C}, \mathsf{Set})$ the category with objects all functors from $\mathcal{C}$ to $\mathsf{Set}$, and with morphisms natural transformations between functors. Define $\mathsf{Fun}(\mathcal{C}^{\mathrm{opp}}, \mathsf{Set})$ similarly. There exist fully faithful functors (the so-called* Yoneda embeddings*)*

$$h^{-} : \mathcal{C}^{\mathrm{opp}} \to \mathsf{Fun}(\mathcal{C}, \mathsf{Set}) : X \mapsto h^X,\ (f : X \to Y) \mapsto (h^f : h^Y \Rightarrow h^X)$$

*and*

$$h_{-} : \mathcal{C} \to \mathsf{Fun}(\mathcal{C}^{\mathrm{opp}}, \mathsf{Set}) : X \mapsto h_X,\ (f : X \to Y) \mapsto (h_f : h_X \Rightarrow h_Y).$$

*Proof.* The statement that $h^{-}$ and $h_{-}$ are functors follows from the definitions and is left as an exercise for the reader. The fact that these functors are fully faithful is simply a rephrasing of Corollary 7.4.2. □

**Corollary 7.4.4.** *If $X$ and $Y$ are objects in a category $\mathcal{C}$ such that $h_X$ (resp. $h^X$) and $h_Y$ (resp. $h^Y$) are naturally isomorphic, then $X$ and $Y$ are isomorphic as well.*

*Proof.* This follows from Theorem 7.4.3, together with Quick question 7.3.10. □

We are now ready to introduce the crucial notion of *representability* of a functor.

**Definition 7.4.5.** Let $\mathcal{C}$ be a category. A functor $F : \mathcal{C} \to \mathsf{Set}$ is said to be *representable* if there exists an object $X$ in $\mathcal{C}$ such that $h^X$ and $F$ are naturally isomorphic. If $\sigma : h^X \Rightarrow F$ is a natural isomorphism, the pair $(X, \sigma)$ is said to be a *representation* of $F$. Similarly, a functor $G : \mathcal{C}^{\mathrm{opp}} \to \mathsf{Set}$ is representable, with representation $(Y, \tau)$, if there exists an object $Y$ in $\mathcal{C}$ and a natural isomorphism $\tau : h_Y \Rightarrow G$.

Note that if the pair $(X, \sigma)$ represents $F : \mathcal{C} \to \mathsf{Set}$, then the object $X$ is uniquely determined up to isomorphism (this follows immediately from Corollary 7.4.4). Let us now discuss some examples.

**Example 7.4.6.** The forgetful functor $\mathrm{Forget} : \mathsf{Grp} \to \mathsf{Set}$ is represented by the pair $(\mathbf{Z}, \sigma)$, where $\sigma : h^{\mathbf{Z}} \Rightarrow \mathrm{Forget}$ is the natural isomorphism given on objects by $\sigma_G : h^{\mathbf{Z}}(G) \to \mathrm{Forget}(G) : f \mapsto f(1)$.

Similarly, the forgetful functor $\mathrm{Forget} : \mathsf{Ring} \to \mathsf{Set}$ is represented by the pair $(\mathbf{Z}[X], \sigma)$, where $\sigma : h^{\mathbf{Z}[X]} \Rightarrow \mathrm{Forget}$ is the natural isomorphism given by $\sigma_R : h^{\mathbf{Z}[X]}(R) \to \mathrm{Forget}(R) : f \mapsto f(X)$.

In both of these cases, the naturality of $\sigma$ needs to be checked, but this is not hard (exercise!).

The forgetful functor $\mathrm{Forget} : \mathsf{Field} \to \mathsf{Set}$, defined on the category of fields, is *not* representable, since $\mathsf{Field}$ is "disconnected": there is no single field which can represent Forget, since there do not exist any morphisms at all between fields of different characteristics.

**Example 7.4.7.** Let $R$ be a ring. If $M$ and $N$ are $R$-modules, the functor $\mathrm{Bil}_R(M, N, -) : {}_R\mathsf{Mod} \to \mathsf{Set}$ is defined on objects by sending an $R$-module $P$ to the set of $R$-bilinear maps $f : M \times N \to P$, and on morphisms by sending an $R$-module homomorphism $\varphi : P \to Q$ to the map of sets

$$(f : M \times N \to P) \mapsto (\varphi \circ f : M \times N \to Q).$$

Then $\mathrm{Bil}_R(M, N)$ is represented by $(M \otimes_R N, \sigma)$, where $\sigma : h^{M \otimes_R N} \Rightarrow \mathrm{Bil}_R(M, N, -)$ is the natural isomorphism obtained from the universal property of the tensor product (checking naturality is not hard).

**Example 7.4.8.** Consider the functor $\mathrm{Unit} : \mathsf{Ring} \to \mathsf{Set}$ which sends a ring $R$ to its set of units $R^\times$ – see Example 7.2.10, where we interpreted this as a functor from $\mathsf{Ring}$ to $\mathsf{Grp}$. The functor Unit is represented by $(\mathbf{Z}[X, \frac{1}{X}], \sigma)$, where $\sigma : h^{\mathbf{Z}[X, \frac{1}{X}]} \Rightarrow \mathrm{Unit}$ is the natural isomorphism given on a ring $R$ by $h^{\mathbf{Z}[X, \frac{1}{X}]}(R) \to \mathrm{Unit}(R) = R^\times : f \mapsto f(X)$. (As before, naturality needs to be checked.)

**Example 7.4.9.** Consider the functor $\mathrm{Square} : \mathsf{Ring} \to \mathsf{Set}$, which sends a ring $R$ to its set of squares $\{x^2 : x \in R\}$. This functor is *not* representable. Indeed, assume that we have a representation $(R, \sigma)$. Then we have, for every ring $S$, an isomorphism $\sigma_S : h^R(S) = \mathrm{Hom}_{\mathsf{Ring}}(R, S) \xrightarrow{\sim} \mathrm{Square}(S)$ which is natural in $S$. Let $\alpha = \sigma_R(1_R) \in \mathrm{Square}(R)$ be the image of $1_R \in h^R(R)$.

We will show that $\alpha$ is a "universal square", i.e., that $\alpha$ has the following universal property: for every ring $S$ and every square $\beta \in S$, there exists a unique homomorphism $R \to S$ which sends $\alpha$ to $\beta$. Indeed, given any homomorphism $f : R \to S$, the diagram

$$\begin{tikzcd}
h^R(R) \arrow[r, "h^R(f)"] \arrow[d, "\sigma_R"] & h^R(S) \arrow[d, "\sigma_S"] \\
\mathrm{Square}(R) \arrow[r, "\mathrm{Square}(f)"] & \mathrm{Square}(S)
\end{tikzcd}$$

commutes. The proof of Yoneda's lemma shows that $\beta = \mathrm{Square}(f)(\sigma_R(1_R))$ determines $f$ completely, and conversely, that any choice of $\beta \in \mathrm{Square}(S)$ defines a unique $f : R \to S$ such that $f(\alpha) = \beta$.

But this is impossible: we have seen that there should be only one ring homomorphism $f : R \to \mathbf{Z}[X]$ such that $f(\alpha) = X^2$, but if $\varphi : \mathbf{Z}[X] \to \mathbf{Z}[X]$ is the unique ring automorphism which sends $X$ to $-X$, then $\varphi \circ f : R \to \mathbf{Z}[X]$ is a different homomorphism with the same property. Contradiction!

## 7.5 Adjunction

**Disclaimer.** As in the previous section, all categories in this section will be assumed to be *locally small*, which means that the class of morphisms between any two objects of the category is in fact a set.

**Definition 7.5.1.** Let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ be functors. An adjunction between $F$ and $G$ is a natural isomorphism $\alpha : \mathrm{Hom}_{\mathcal{D}}(F(-), -) \overset{\sim}{\Longrightarrow} \mathrm{Hom}_{\mathcal{C}}(-, G(-))$ of functors $\mathcal{C}^{\mathrm{opp}} \times \mathcal{D} \to \mathsf{Set}$. If such an adjunction exists, we say that $F$ is *left adjoint* to $G$, and that $G$ is *right adjoint* to $F$.

More concretely, an adjunction consists of the data (for all $X \in \mathrm{ob}(\mathcal{C})$ and $Y \in \mathrm{ob}(\mathcal{D})$) of a bijection

$$\alpha_{X,Y} : \mathrm{Hom}_{\mathcal{D}}(FX, Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, GY)$$

such that for every $f : X \to X'$ in $\mathcal{C}$, the square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(FX, Y) \arrow[r, "\alpha_{X,Y}"] & \mathrm{Hom}_{\mathcal{C}}(X, GY) \\
\mathrm{Hom}_{\mathcal{D}}(FX', Y) \arrow[r, "\alpha_{X',Y}"] \arrow[u, "-\circ Ff"] & \mathrm{Hom}_{\mathcal{C}}(X', GY) \arrow[u, "-\circ f"']
\end{tikzcd}$$

commutes, and similarly, for every $g : Y \to Y'$ in $\mathcal{D}$, the square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(FX, Y) \arrow[r, "\alpha_{X,Y}"] \arrow[d, "g\circ -"] & \mathrm{Hom}_{\mathcal{C}}(X, GY) \arrow[d, "Gg\circ -"] \\
\mathrm{Hom}_{\mathcal{D}}(FX, Y') \arrow[r, "\alpha_{X,Y'}"] & \mathrm{Hom}_{\mathcal{C}}(X, GY')
\end{tikzcd}$$

commutes.

*Remark* 7.5.2. This terminology comes from linear algebra: if $V$ and $W$ are vector spaces over some field equipped with inner products, then linear maps $f : V \to W$ and $g : W \to V$ are called *adjoint* if the equality $\langle f(v), w\rangle_W = \langle v, g(w)\rangle_V$ holds for all $v \in V$ and $w \in W$.

If $(F : \mathcal{C} \to \mathcal{D}, G : \mathcal{D} \to \mathcal{C})$ is a pair of adjoint functors, with adjunction $\alpha$, then we obtain a bijection

$$\alpha_{X,FX} : \mathrm{Hom}_{\mathcal{D}}(FX, FX) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, GFX)$$

for every object $X$ in $\mathcal{C}$. The image of $1_{FX}$ under this map is a morphism $\eta_X : X \to GFX$ in $\mathcal{C}$.

*Quick question* 7.5.3. Show that the morphisms $\eta_X$ together give a natural transformation $\eta : 1_{\mathcal{C}} \Rightarrow GF$.

Similarly, we obtain a natural transformation $\varepsilon : FG \Rightarrow 1_{\mathcal{D}}$.

**Definition 7.5.4.** With notation as above, the natural transformations $\eta : 1_{\mathcal{C}} \Rightarrow GF$ and $\varepsilon : FG \Rightarrow 1_{\mathcal{D}}$ are called the *unit* and *counit* of the adjunction between $F$ and $G$.

The unit $\eta$ and counit $\varepsilon$ indicate that an adjunction between two functors should be thought of as a generalisation of an equivalence of categories: rather than the existence of a pair of natural isomorphisms $1_{\mathcal{C}} \overset{\sim}{\Longrightarrow} GF$ and $FG \overset{\sim}{\Longrightarrow} 1_{\mathcal{D}}$ (which would exist if $G$ were a quasi-inverse for $F$), we only require the existence of natural transformations, satisfying a mild compatibility condition (see Exercise 7.14).

Adjoint pairs of functors are ubiquitous in algebra: let us give some examples.

**Example 7.5.5.** Let $A$ be a set. For all sets $X$ and $Y$, we have a canonical bijection

$$\alpha_{X,Y} : \mathrm{Hom}_{\mathsf{Set}}(X \times A, Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathsf{Set}}(X, \mathrm{Hom}_{\mathsf{Set}}(A, Y))$$

obtained by "currying" (see §6.3): we send $f : X \times A \to Y$ to the map

$$X \to \mathrm{Hom}_{\mathsf{Set}}(A, Y) : x \mapsto (a \mapsto f(x, a)),$$

and conversely, we send $g : X \to \mathrm{Hom}_{\mathsf{Set}}(A, Y)$ to

$$X \times A \to Y : (x, a) \mapsto g(x)(a).$$

Then $\alpha$ defines an adjunction, meaning that the functor

$$- \times A : \mathsf{Set} \to \mathsf{Set} : X \mapsto X \times A, (f : X \to Y) \mapsto (f \times 1_A : X \times A \to Y \times A)$$

is left adjoint to the functor $\mathrm{Hom}_{\mathsf{Set}}(A, -) : \mathsf{Set} \to \mathsf{Set}$.

*Quick question* 7.5.6. Compute the unit and counit of the adjunction from Example 7.5.5.

**Example 7.5.7.** The adjunction from the previous example can be upgraded to the category of $R$-modules, where $R$ is an arbitrary ring: if $N$ is an $R$-module, we have an adjunction between the functors

$$\otimes_R N : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}$$

and

$$\mathrm{Hom}_R(N, -) : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}.$$

Of course this statement is not new for us, see §6.3: this simply says that we have isomorphisms

$$\mathrm{Hom}_R(M \otimes_R N, P) \xrightarrow{\sim} \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P))$$

for all $R$-modules $M$ and $P$, functorial in both arguments (contravariant in $M$ and covariant in $P$).

*Quick question* 7.5.8. Compute again the unit and counit of the adjunction from Example 7.5.7.

**Example 7.5.9.** Let $R$ be a ring. Let $\mathsf{Forget} : {}_R\mathsf{Mod} \to \mathsf{Set}$ be the functor which forgets the $R$-module structure, and let $\mathsf{Free} : \mathsf{Set} \to {}_R\mathsf{Mod}$ be the functor from Example 7.2.8. Then $\mathsf{Free}$ is a left adjoint to $\mathsf{Forget}$. This means that we have, for every $R$-module $M$ and every set $S$, an isomorphism

$$\mathrm{Hom}_R(\mathsf{Free}(S), M) \xrightarrow{\sim} \mathrm{Hom}_{\mathsf{Set}}(S, \mathsf{Forget}(M))$$

which is functorial in both $S$ and $M$ (contravariant in $S$, covariant in $M$).

Indeed, given a set $S$ and an $R$-module $M$, a morphism $S \to M$ of sets defines a unique $R$-module homomorphism $R^{(S)} \to M$, and vice versa. Moreover, if $S \to T$ is a morphism of sets and if $M \to N$ is an $R$-module homomorphism, then the obvious diagrams

$$\begin{tikzcd}
\mathrm{Hom}_R(R^{(S)}, M) \arrow[r] & \mathrm{Hom}_{\mathsf{Set}}(S, M) \\
\mathrm{Hom}_R(R^{(T)}, M) \arrow[r] \arrow[u] & \mathrm{Hom}_{\mathsf{Set}}(T, M) \arrow[u]
\end{tikzcd}
\quad \text{and} \quad
\begin{tikzcd}
\mathrm{Hom}_R(R^{(S)}, M) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathsf{Set}}(S, M) \arrow[d] \\
\mathrm{Hom}_R(R^{(S)}, N) \arrow[r] & \mathrm{Hom}_{\mathsf{Set}}(S, N)
\end{tikzcd}$$

commute.

We will cover more examples in the exercises. We end this section with a uniqueness statement:

**Proposition 7.5.10.** *Let $F : \mathcal{C} \to \mathcal{D}$ be a functor. If both $G_1 : \mathcal{D} \to \mathcal{C}$ and $G_2 : \mathcal{D} \to \mathcal{C}$ are right adjoint to $F$, then $G_1$ and $G_2$ are naturally isomorphic. Similarly, if both $F_1 : \mathcal{C} \to \mathcal{D}$ and $F_2 : \mathcal{C} \to \mathcal{D}$ are left adjoint to $G : \mathcal{D} \to \mathcal{C}$, then $F_1$ and $F_2$ are naturally isomorphic.*

*Proof.* We prove the statement on right adjoints; the proof can easily be adapted to the case of left adjoints.

Choosing adjunctions between $F$ and $G_1$ and between $F$ and $G_2$ yields bijections

$$\mathrm{Hom}_{\mathcal{C}}(X, G_1 Y) \xleftarrow{\sim} \mathrm{Hom}_{\mathcal{D}}(FX, Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, G_2 Y)$$

which are functorial in both $X$ and $Y$ (contravariant in $X$, covariant in $Y$).

Hence we have, for all objects $X$ in $\mathcal{C}$ and $Y$ in $\mathcal{D}$, functorial isomorphisms

$$\mathrm{Hom}_{\mathcal{C}}(X, G_1 Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, G_2 Y).$$

Functoriality in $X$ means that we have a natural isomorphism

$$\mathrm{Hom}_{\mathcal{C}}(-, G_1 Y) \overset{\sim}{\Longrightarrow} \mathrm{Hom}_{\mathcal{C}}(-, G_2 Y)$$

in $\mathsf{Fun}(\mathcal{C}^{\mathrm{opp}}, \mathsf{Set})$ for every object $Y$ in $\mathcal{D}$, which by Yoneda's lemma (Corollary 7.4.4) comes from a unique isomorphism in $\mathcal{C}$ which we denote by $\gamma_Y : G_1 Y \longrightarrow G_2 Y$. Functoriality in $Y$ (why?) of these isomorphisms implies that we obtain a natural isomorphism $\gamma : G_1 \Longrightarrow G_2$, as desired. □

## 7.6 Limits

**Disclaimer.** All categories in this section will be assumed to be *small*, which means that the class of objects and the classes of morphisms between any two objects of the category are sets.

We end our crash course on category theory with a brief introduction to the theory of (co)limits, which generalises many known concepts encountered in previous courses on pure mathematics. We will start with a simple special case: the notion of (co)product.

**Definition 7.6.1.** Let $X$ and $Y$ be objects in a category $\mathcal{C}$.

A *product* of $X$ and $Y$ is a triple $(P, \pi_X, \pi_Y)$, where $P$ is an object and $\pi_X : P \to X$, $\pi_Y : P \to Y$ are morphisms in $\mathcal{C}$, with the following property: given an object $Q$ and morphisms $q_X : Q \to X$ and $q_Y : Q \to Y$ in $\mathcal{C}$, there exists a unique morphism $u : Q \to P$ such that $q_X = \pi_X u$ and $q_Y = \pi_Y u$.

A *coproduct* of $X$ and $Y$ is a triple $(C, \iota_X, \iota_Y)$, where $C$ is an object and $\iota_X : X \to C$, $\iota_Y : Y \to C$ are morphisms in $\mathcal{C}$, with the following property: given an object $D$ and morphisms $j_X : X \to D$ and $j_Y : Y \to D$ in $\mathcal{C}$, there exists a unique morphism $v : C \to D$ such that $j_X = v\iota_X$ and $j_Y = v\iota_Y$.

The “universal” properties of the product and coproduct are visualised in the diagrams

$$\begin{tikzcd} Q \arrow[drr, bend left, "q_Y"] \arrow[dr, dashed, "u"] \arrow[ddr, bend right, "q_X"'] & & \\ & P \arrow[r, "\pi_Y"] \arrow[d, "\pi_X"'] & Y \\ & X & \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} & X \arrow[d, "\iota_X"] \arrow[ddr, bend left, "j_X"] & \\ Y \arrow[r, "\iota_Y"'] \arrow[drr, bend right, "j_Y"'] & C \arrow[dr, dashed, "v"] & \\ & & D \end{tikzcd}$$

respectively.

Nothing guarantees that the (co)product of a pair of objects need exist in an arbitrary category, but *if* a (co)product exists, then it is unique in a strong sense. Let us prove this for the product – the case of coproducts can be handled using identical arguments (exercise for the reader).

Given a category $\mathcal{C}$ and objects $X$ and $Y$ in $\mathcal{C}$, define a new category $\mathcal{C}_{/(X,Y)}$ with objects triples of the form $(Q, q_X, q_Y)$, where $Q$ is an object in $\mathcal{C}$, $q_X \in \operatorname{Hom}_{\mathcal{C}}(Q, X)$ and $q_Y \in \operatorname{Hom}_{\mathcal{C}}(Q, Y)$. A morphism from $(Q, q_X, q_Y)$ to $(Q', q'_X, q'_Y)$ is defined to be a morphism $\varphi \in \operatorname{Hom}_{\mathcal{C}}(Q, Q')$ such that $q_X = q'_X\varphi$ and $q_Y = q'_Y\varphi$. It is easy to check that $\mathcal{C}_{/(X,Y)}$ is indeed a category.

Definition 7.6.1 says that $(P, \pi_X, \pi_Y)$ is a product of $X$ and $Y$ if and only if $(P, \pi_X, \pi_Y)$ is a *final* object of $\mathcal{C}_{/(X,Y)}$. By Proposition 7.1.22, we know that $(P, \pi_X, \pi_Y)$ is unique up to unique isomorphism in $\mathcal{C}_{/(X,Y)}$. This means precisely that if $(P, \pi_X, \pi_Y)$ and $(P', \pi'_X, \pi'_Y)$ are two products of $X$ and $Y$ in $\mathcal{C}$, then there exists a unique *isomorphism* $\varphi : P \to P'$ in $\mathcal{C}$ such that $\pi_X = \pi'_X\varphi$ and $\pi_Y = \pi'_Y\varphi$.

If the product of $X$ and $Y$ in $\mathcal{C}$ exists, then we usually denote it by $X \times Y$ or sometimes $X \prod Y$, even though this is sloppy notation: a product is in fact a *triple* consisting of one object and two morphisms. Similarly, the coproduct of $X$ and $Y$ in $\mathcal{C}$ (if it exists) will be denoted by $X \coprod Y$ or sometimes $X \sqcup Y$.

**Example 7.6.2.** In the category Set, the product of sets $X$ and $Y$ is simply the Cartesian product $X \times Y$, together with the coordinate projections $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$. The coproduct of $X$ and $Y$ is the disjoint union $X \sqcup Y$, together with the injections $\iota_X : X \to X \sqcup Y$ and $\iota_Y : Y \to X \sqcup Y$.

*Remark* 7.6.3. Let $\mathcal{C}$ be a category, and let $X$ and $Y$ be objects in $\mathcal{C}$. Then $X$ and $Y$ have a product in $\mathcal{C}$ if and only if the functor $\mathcal{C}^{\mathrm{opp}} \to \mathsf{Set} : T \mapsto \operatorname{Hom}_{\mathcal{C}}(T, X) \times \operatorname{Hom}_{\mathcal{C}}(T, Y)$ is representable; if so, the product $X \times Y$ represents this functor. Similarly, $X$ and $Y$ have a coproduct in $\mathcal{C}$ if and only if the functor $\mathcal{C} \to \mathsf{Set} : T \mapsto \operatorname{Hom}_{\mathcal{C}}(X, T) \times \operatorname{Hom}_{\mathcal{C}}(Y, T)$ is representable; if so, the coproduct $X \coprod Y$ represents this functor. Note that this gives a new proof of the uniqueness of (co)products (see Corollary 7.4.4).

**Example 7.6.4.** In the category Grp, the product of $G_1$ and $G_2$ is the direct product $G_1 \times G_2$, with elements $\{(g_1, g_2) : g_1 \in G_1, g_2 \in G_2\}$ and componentwise multiplication. In Ab, the product of $G_1$ and $G_2$ is the direct sum $G_1 \oplus G_2$ (which is really the same thing as the direct product. . . ). On the other hand, in the category Cyc of cyclic groups, products of two objects do not exist in general; for example, $\mathbf{Z}/2$ and $\mathbf{Z}/4$ do not have a product in Cyc (why?). More generally, one can show that $\mathbf{Z}/m$ and $\mathbf{Z}/n$ have a product in Cyc (namely $\mathbf{Z}/(mn)$) if and only if $m$ and $n$ are coprime integers.

In the category Grp, the coproduct of $G_1$ and $G_2$ is the so-called *free product* $G_1 \star G_2$. In Ab, the coproduct of $G_1$ and $G_2$ is again $G_1 \oplus G_2$, so that products and coproducts of two objects in Ab coincide.

*Quick question* 7.6.5. Let $R$ be a ring, and let $M$ and $N$ be $R$-modules. Show that the $R$-module $M \oplus N$ is both the product and the coproduct of $M$ and $N$ in the category ${}_R\mathsf{Mod}$.

**Example 7.6.6.** Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. The product of these spaces in Top is the Cartesian product $X \times Y$, equipped with the so-called product topology. The coproduct of these spaces in Top is the disjoint union $X \sqcup Y$, equipped with the topology $\{U \sqcup V : U \in \mathcal{T}_X, V \in \mathcal{T}_Y\}$.

We can easily generalise the notions of product and coproduct to arbitrary sets of indices:

**Definition 7.6.7.** Let $(X_i)_{i\in I}$ be a collection of objects in a category $\mathcal{C}$.

A product of $(X_i)_{i\in I}$ is a tuple $(P,(\pi_i)_{i\in I})$, where $P$ is an object and $\pi_i : P \to X_i$ are morphisms in $\mathcal{C}$, with the property that if $Q$ is any object in $\mathcal{C}$ equipped with morphisms $q_i : Q \to X_i$ for all $i \in I$, then there exists a unique morphism $u : Q \to P$ in $\mathcal{C}$ such that $q_i = \pi_i u$ for all $i \in I$.

A coproduct of $(X_i)_{i\in I}$ is a tuple $(C,(\iota_i)_{i\in I})$, where $C$ is an object and $\iota_i : X_i \to C$ are morphisms in $\mathcal{C}$, with the property that if $D$ is any object in $\mathcal{C}$ equipped with morphisms $j_i : X_i \to D$ for all $i \in I$, then there exists a unique morphism $v : C \to D$ in $\mathcal{C}$ such that $j_i = v\iota_i$ for all $i \in I$.

As before, we have that products and coproducts of arbitrarily many objects are unique up to unique isomorphism. The product (resp. coproduct) of $(X_i)_{i\in I}$ will be denoted by $\prod_{i\in I} X_i$ (resp. $\coprod_{i\in I} X_i$).

*Quick question* 7.6.8. Rephrase the existence of arbitrary (co)products in terms of representable functors, following Remark 7.6.3.

*Quick question* 7.6.9. Show that arbitrary products and coproducts exist in the categories Set, Top and ${}_R$Mod, where $R$ is an arbitrary ring. (For the latter category, see also Definition 2.2.1.)

**Example 7.6.10.** Let TorAb be the category of torsion abelian groups. The product of finitely many torsion abelian groups is simply the direct product. However, if $(A_i)_{i\in A}$ is an infinite collection of torsion abelian groups, then the direct product $\prod_{i\in I} A_i$ is still the product in Ab, but not necessarily in TorAb, since $\prod_{i\in I} A_i$ need not be torsion; for example, the direct product $\prod_{n\geq 1} \mathbf{Z}/2^n$ is not torsion, since $(\overline{1},\overline{1},\cdots)$ is not a torsion element.

However, the *torsion subgroup* $T\left(\prod_{i\in I} A_i\right)$ of $\prod_{i\in I} A_i$ is a product of $(A_i)_{i\in I}$ in TorAb; this follows essentially from the fact that if $f : B \to \prod_{i\in I} A_i$ is a morphism in Ab and if $B$ is torsion, then $f(B)$ lies in the torsion subgroup of $\prod_{i\in I} A_i$, and hence $f$ factors uniquely through $T\left(\prod_{i\in I} A_i\right)$.

*Remark* 7.6.11. If $(f_i : X_i \to Y_i)_{i\in I}$ is a collection of morphisms in a category $\mathcal{C}$ in which the products $\prod_{i\in I} X_i$ and $\prod_{i\in I} Y_i$ both exist, then we get an induced morphism $\prod_{i\in I} f_i : \prod_{i\in I} X_i \to \prod_{i\in I} Y_i$, as follows: $\prod_{i\in I} f_i$ is the unique morphism obtained from the morphisms $f_i\pi_{X_i} : \prod_{i\in I} X_i \to Y_i$ using the universal property of the product. Similarly, if the coproducts $\coprod_{i\in I} X_i$ and $\coprod_{i\in I} Y_i$ exist, we obtain a unique morphism $\coprod_{i\in I} f_i : \coprod_{i\in I} X_i \to \coprod_{i\in I} Y_i$ from the morphisms $\iota_{Y_i} f_i : X_i \to \coprod_{i\in I} Y_i$.

In general, morphisms from coproducts to products are easy to describe.

Let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be collections of objects in a category $\mathcal{C}$. Assume that the coproduct $\coprod_{i\in I} X_i$ and the product $\prod_{j\in J} Y_j$ both exist. Then every morphism $f : \coprod_{i\in I} X_i \to \prod_{j\in J} Y_j$ in $\mathcal{C}$ induces a "matrix of morphisms" $(f_{ij} : X_i \to Y_j)_{i\in I, j\in J}$, where $f_{ij} = \pi_{Y_j} f \iota_{X_i}$.

Conversely, given morphisms $(f_{ij} : X_i \to Y_j)_{i\in I, j\in J}$, we obtain a unique morphism $f$ as above, with the property that $f_{ij} = \pi_{Y_j} f \iota_{X_i}$ for all $i \in I$ and $j \in J$. There are two ways to proceed:

– for each $i \in I$, the morphisms $(f_{ij})_{j\in J}$ induce a morphism $g_i : X_i \to \prod_{j\in J} Y_j$. The morphisms $(g_i)_{i\in I}$ together define a morphism $f_1 : \coprod_{i\in I} X_i \to \prod_{j\in J} Y_j$.

– for each $j \in J$, the morphisms $(f_{ij})_{i\in I}$ induce a morphism $h_j : \coprod_{i\in I} X_i \to Y_j$. The morphisms $(h_j)_{j\in J}$ together define a morphism $f_2 : \coprod_{i\in I} X_i \to \prod_{j\in J} Y_j$

*Quick question* 7.6.12. With notation as above, show that $f_1$ and $f_2$ are equal, and that the morphism $f$ produced by these equivalent constructions satisfies the equalities $f_{ij} = \pi_{Y_j} f \iota_{X_i}$ for all $i \in I$ and $j \in J$.

Let us now look at a construction which is at the same time a special case and a generalisation of the notion of (co)product. We will need the following notions for the construction.

**Definition 7.6.13.** Let $\mathcal{C}$ be a category. If $X$ is an object in $\mathcal{C}$, the *slice category* of $\mathcal{C}$ above $X$, denoted $\mathcal{C}_{/X}$, is the category whose objects are morphisms $Y \to X$ in $\mathcal{C}$, such that morphisms from $f_1 : Y_1 \to X$ to $f_2 : Y_2 \to X$ in $\mathcal{C}_{/X}$ are morphisms $g : Y_1 \to Y_2$ in $\mathcal{C}$ for which $f_1 = f_2 g$. The *coslice category* of $\mathcal{C}$ under $X$, denoted ${}^{X/}\mathcal{C}$, is the category whose objects are morphisms $X \to Y$ in $\mathcal{C}$, such that morphisms from $f_1 : X \to Y_1$ to $f_2 : X \to Y_2$ in ${}^{X/}\mathcal{C}$ are morphisms $g : Y_1 \to Y_2$ in $\mathcal{C}$ for which $f_2 = g f_1$.

Morphisms in $\mathcal{C}_{/X}$ and ${}^{X/}\mathcal{C}$ are therefore commutative triangles of the forms

$$\begin{tikzcd}
Y_1 \arrow[rr, "g"] \arrow[dr, "f_1"'] & & Y_2 \arrow[dl, "f_2"] \\
& X &
\end{tikzcd}
\quad \text{and} \quad
\begin{tikzcd}
& X \arrow[dl, "f_1"'] \arrow[dr, "f_2"] & \\
Y_1 \arrow[rr, "g"] & & Y_2
\end{tikzcd}$$

respectively. We are now ready to define pullbacks and pushouts.

**Definition 7.6.14.** Let $\mathcal{C}$ be a category, and let $Z$ be an object in $\mathcal{C}$. Given two morphisms $f : X \to Z$ and $g : Y \to Z$ in $\mathcal{C}$, we define the *pullback* of $f$ and $g$ to be the product of $f$ and $g$ in the slice category $\mathcal{C}_{/Z}$. In the dual situation, given two morphisms $f : Z \to X$ and $g : Z \to Y$ in $\mathcal{C}$, we define the *pushout* of $f$ and $g$ to be the coproduct of $f$ and $g$ in the coslice category ${}^{Z/}\mathcal{C}$.

Translating the definition from $\mathcal{C}_{/Z}$ and ${}^{Z/}\mathcal{C}$ to the category $\mathcal{C}$ we started with, we see that the defining universal properties of the pullback and pushout are described by the following diagrams:

$$\begin{tikzcd}
Q \arrow[drr, bend left, "q_Y"] \arrow[ddr, bend right, "q_X"'] \arrow[dr, dashed, "u"] & & \\
& P \arrow[r, "\pi_Y"] \arrow[d, "\pi_X"'] & Y \arrow[d, "g"] \\
& X \arrow[r, "f"] & Z
\end{tikzcd}
\quad \text{and} \quad
\begin{tikzcd}
Z \arrow[r, "f"] \arrow[d, "g"] & X \arrow[d, "\iota_X"] \arrow[ddr, bend left, "j_X"] & \\
Y \arrow[r, "\iota_Y"'] \arrow[drr, bend right, "j_Y"'] & C \arrow[dr, dashed, "v"] & \\
& & D
\end{tikzcd}$$

Let us spell out what the first diagram means in concrete terms: the pullback of $f$ and $g$ (if it exists!) is a triple $(P, \pi_X, \pi_Y)$, where $P$ is an object in $\mathcal{C}$ and $\pi_X : P \to X$ and $\pi_Y : P \to Y$ are morphisms such that $f\pi_X = g\pi_Y$, such that if $(Q, q_X, q_Y)$ is another such triple with the property that $fq_X = gq_Y$, then there exists a unique morphism $u : Q \to P$ such that $q_X = \pi_X u$ and $q_Y = \pi_Y u$.

*Quick question* 7.6.15. Spell out the meaning of the second diagram.

By construction, the pullback and pushout (if they exist) are unique up to unique isomorphism in an appropriate sense (we let the reader fill in the details). These objects will sometimes be denoted by $X \times_Z Y$ and $X \sqcup_Z Y$ respectively, if the morphisms $f$ and $g$ are clear from the context.

**Example 7.6.16.** Pullbacks exist in Set. Indeed, if $f : X \to Z$ and $g : Y \to Z$ are two morphisms in Set, then the set $\{(x, y) \in X \times Y : f(x) = g(y)\}$ satisfies the universal property.

The pullback is sometimes also called the *fibre product*, for example because a fibre of a map of sets can be seen as a pullback: if $f : X \to Z$ is a map and if $\iota_Z : \{z\} \hookrightarrow Z$ is the inclusion of an element $z \in Z$, then the pullback of $f$ and $\iota_Z$ is isomorphic to $f^{-1}(z)$. If $X$ and $Y$ are subsets of $Z$ and if $f$ and $g$ are the corresponding inclusion maps, then $X \times_Z Y$ is nothing but $X \cap Y$.

Pushouts exist as well in Set: given two maps of sets $f : Z \to X$ and $g : Z \to Y$, then the quotient $(X \sqcup Y)/\sim$, where $X \sqcup Y$ denotes the disjoint union of $X$ and $Y$, and where $\sim$ is the equivalence relation on $X \sqcup Y$ generated by $f(z) \sim g(z)$ for all $z \in Z$, satisfies the universal property.

*Quick question* 7.6.17. Using the existence of pullbacks and pushouts in Set, characterise pullbacks and pushouts in arbitrary categories using the framework of representable functors (see Remark 7.6.3).

**Example 7.6.18.** Pullbacks and pushouts exist in Top.

Indeed, if $f : X \to Z$ and $g : Y \to Z$ are continuous maps between topological spaces, then the set $\{(x, y) \in X \times Y : f(x) = g(y)\}$ from the previous example, equipped with the subspace topology inherited from the product topology on $X \times Y$, satisfies the universal property of the pullback.

Moreover, if $f : Z \to X$ and $g : Z \to Y$ are continuous, then the set $(X \sqcup Y)/\sim$ from the previous example, now equipped with the natural quotient topology, satisfies the universal property of the pushout. If $Z$ is a "common subspace" of $X$ and $Y$ and if $f$ and $g$ are the corresponding embeddings, then this yields a way to "glue" $X$ and $Y$ along $Z$.

**Example 7.6.19.** Pullbacks and pushouts exist in ${}_R\mathsf{Mod}$, for an arbitrary ring $R$.

Indeed, if $f_1 : M_1 \to N$ and $f_2 : M_2 \to N$ are two homomorphisms of $R$-modules, then the set $\{(x_1, x_2) \in M_1 \oplus M_2 : f_1(x_1) = f_2(x_2)\}$ is naturally an $R$-submodule of the direct sum $M_1 \oplus M_2$, and satisfies the universal property of the pullback.

Moreover, if $f_1 : N \to M_1$ and $f_2 : N \to M_2$ are $R$-module homomorphisms, then the quotient module $(M_1 \oplus M_2)/P$, where $P$ is the $R$-submodule $\{(f_1(y), -f_2(y)) : y \in N\}$ of $M_1 \oplus M_2$, can be seen to satisfy the universal property of the pushout.

Let us now introduce yet another pair of dual concepts: equalisers and co-equalisers.

**Definition 7.6.20.** Let $f, g : X \to Y$ be morphisms in a category $\mathcal{C}$. An *equaliser* of $f$ and $g$ is a morphism $\iota : E \to X$ in $\mathcal{C}$ such that $f\iota = g\iota$, with the property that if $i : E' \to X$ is another morphism for which $fi = gi$, then there exists a unique morphism $u : E' \to E$ such that $i = \iota u$. A *coequaliser* of $f$ and $g$ is a morphism $\pi : Y \to C$ such that $\pi f = \pi g$, with the property that if $p : Y \to C'$ is another morphism for which $pf = pg$, then there exists a unique morphism $v : C \to C'$ such that $p = v\pi$.

The defining properties of the equaliser and coequaliser are described by the diagrams

$$\begin{tikzcd}
E \arrow[r, "\iota"] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \\
E' \arrow[u, dashed, "u"] \arrow[ur, "i"'] & &
\end{tikzcd}$$

and

$$\begin{tikzcd}
X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[r, "\pi"] \arrow[dr, "p"'] & C \arrow[d, dashed, "v"] \\
& & C'
\end{tikzcd}$$

Just like (co)products, pullbacks and pushouts, (co)equalisers are unique up to unique isomorphism.

**Example 7.6.21.** Equalisers and coequalisers exist in Set, Top and ${}_R\mathsf{Mod}$. We will leave the first two cases as simple exercises for the reader, and look at the latter case in some detail.

If $f, g : M \to N$ are two morphisms in ${}_R\mathsf{Mod}$, then $f - g : M \to N$ is another morphism. The inclusion $\iota : \ker(f - g) \hookrightarrow M$ is an equaliser of $f$ and $g$. In particular, if $g = 0$, we see that the kernel of an $R$-module homomorphism is a special case of an equaliser. The same construction allows us to define "categorical kernels" in an arbitrary category with a zero object: the kernel of $f : X \to Y$ is the equaliser of $f : X \to Y$ and the zero morphism $0 : X \to Y$ (which factors over the zero object).

Similarly, $\pi : N \twoheadrightarrow \operatorname{coker}(f - g)$ is a coequaliser of $f$ and $g$, and the cokernel is a special case of a coequaliser; this allows us to define "categorical cokernels" in a category with a zero object.

We have now encountered various notions – (co)products, pullbacks and pushouts, (co)equalisers – with very similar properties and flavours. The aching question is: what do these constructions have in common? The answer is that they are special cases of an extremely general concept: (co)limits.

To define this concept, we need the notion of a *diagram* in a category. We have already encountered diagrams in a category $\mathcal{C}$ indexed by a preordered set $(I, \leq)$, see Example 7.2.13. Such a diagram can be described as a functor $X : \mathcal{I} \to \mathcal{C}$, where $\mathcal{I}$ is the thin category associated to $(I, \leq)$.

In fact there is no reason to restrict to indexing by a thin category associated to a preordered set: *any* (small) category can serve as an indexing object for a diagram. For some intuition behind this insight, we refer to this blog post. The formal definition is the following.

**Definition 7.6.22.** Let $\mathcal{I}$ be a category. A *diagram indexed by* $\mathcal{I}$ in a category $\mathcal{C}$ is a functor $X : \mathcal{I} \to \mathcal{C}$.

This means simply that for every object $i$ in $\mathcal{I}$, we have an associated object $X_i := X(i)$ in $\mathcal{C}$, and for every morphism $\varphi : i \to j$ in $\mathcal{I}$, we have an associated morphism $X(\varphi) : X_i \to X_j$.

**Definition 7.6.23.** The *limit* of a functor (or diagram) $X : \mathcal{I} \to \mathcal{C}$ consists of an object $\lim_{\mathcal{I}} X$ in $\mathcal{C}$, together with a morphism $\pi_i : \lim_{\mathcal{I}} X \to X_i$ for every object $i$ in $\mathcal{I}$, such that

(1) for every morphism $\varphi : i \to j$ in $\mathcal{I}$, we have $\pi_j = X(\varphi) \circ \pi_i$;

(2) if $T$ is an object in $\mathcal{C}$ equipped with morphisms $t_i : T \to X_i$ such that $t_j = X(\varphi) \circ t_i$ for every $\varphi : i \to j$ in $\mathcal{I}$, then there is a unique morphism $u : T \to \lim_{\mathcal{I}} X$ such that $t_i = \pi_i u$ for all $i$ in $\mathcal{I}$.

*Quick question* 7.6.24. Show that if the indexing category $\mathcal{I}$ is one of the categories

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

then we recover the notions of product, pullback and equaliser respectively. Show furthermore that the empty indexing category recovers the notion of a final object (Definition 7.1.19).

The dual notion is the colimit of a functor:

**Definition 7.6.25.** The *colimit* of a functor (or diagram) $X : \mathcal{I} \to \mathcal{C}$ consists of an object $\mathrm{colim}_{\mathcal{I}} X$ in $\mathcal{C}$, together with a morphism $\iota_i : X_i \to \mathrm{colim}_{\mathcal{I}} X$ for every object $i$ in $\mathcal{I}$, such that

(1) for every morphism $\varphi : i \to j$ in $\mathcal{I}$, we have $\iota_j \circ X(\varphi) = \iota_i$;

(2) if $T$ is an object in $\mathcal{C}$ equipped with morphisms $t_i : X_i \to T$ such that $t_j \circ X(\varphi) = t_i$ for every $\varphi : i \to j$ in $\mathcal{I}$, then there is a unique morphism $u : \mathrm{colim}_{\mathcal{I}} X \to T$ such that $t_i = u\iota_i$ for all $i$ in $\mathcal{I}$.

*Quick question* 7.6.26. Show that coproducts, pushouts, coequalisers and initital objects are colimits.

Of course (co)limits need not exist in general, but if they exist, they are unique in the appropriate sense. We can rephrase this in terms of representability of certain functors, using the following result.

**Proposition 7.6.27.** *Arbitrary limits and colimits exist in* Set.

*Proof.* We prove the result for limits, and leave the case of colimits to the reader.

Let $X : \mathcal{I} \to \mathsf{Set}$ be a functor. Then consider the following subset of $\prod_{i \in \mathrm{ob}(\mathcal{I})} X_i$:

$$L = \left\{ (x_i)_{i \in \mathrm{ob}(\mathcal{I})} : X(\varphi)(x_i) = x_j \ \text{ for all } \varphi : i \to j \text{ in } \mathcal{I} \right\}.$$

Then $L$, equipped with the various projections $\pi_i : L \to X_i$, satisfies the universal property of the limit.

Indeed, for every $x \in L$ and $\varphi : i \to j$ in $\mathcal{I}$, we have

$$\pi_j(x) = x_j = X(\varphi)(x_i) = (X(\varphi) \circ \pi_i)(x).$$

Moreover, if $(T, (t_i : T \to X_i)_{i \in \mathrm{ob}(\mathcal{I})})$ satisfies $t_j = X(\varphi) \circ t_i$ for all $\varphi : i \to j$ in $\mathcal{I}$, then the map $u : T \to L : x \mapsto (t_i(x))_{i \in \mathrm{ob}(\mathcal{I})}$ is well-defined and is the unique map such that $t_i = \pi_i u$ for all $i$. □

The following result provides what one could call the "Yoneda perspective" on the existence and characterisation of (co)limits, already hinted at in Remark 7.6.3 and Quick question 7.6.17; the proof is really not much more than a reinterpretation of Definitions 7.6.23 and 7.6.25:

**Proposition 7.6.28.** *Let $X : \mathcal{I} \to \mathcal{C}$ be a functor. Let $Y$ be an object of $\mathcal{C}$. Then $\lim_{\mathcal{I}} X$ exists and is isomorphic to $Y$ if and only if we have a natural isomorphism*

$$\mathrm{Hom}_{\mathcal{C}}(-,Y) \overset{\sim}{\Longrightarrow} \lim_{\mathcal{I}} \mathrm{Hom}_{\mathcal{C}}(-,X_i)$$

*of functors $\mathcal{C}^{\mathrm{opp}} \to \mathsf{Set}$.*

*Dually, $\mathrm{colim}_{\mathcal{I}} X$ exists and is isomorphic to $Y$ if and only if we have a natural isomorphism*

$$\mathrm{Hom}_{\mathcal{C}}(Y,-) \overset{\sim}{\Longrightarrow} \lim_{\mathcal{I}^{\mathrm{opp}}} \mathrm{Hom}_{\mathcal{C}}(X_i,-)$$

*of functors $\mathcal{C} \to \mathsf{Set}$.*

Often an important question is whether a functor is "(co)continuous" in the sense that it preserves (co)limits, or special cases of those – such as (co)products. We have the following very general result:

**Theorem 7.6.29.** *Let $F : \mathcal{C} \to \mathcal{D}$ be a functor which is left adjoint to $G : \mathcal{D} \to \mathcal{C}$.*

*If $X : \mathcal{I} \to \mathcal{C}$ is a functor such that $\mathrm{colim}_{\mathcal{I}} X$ exists in $\mathcal{C}$, then $\mathrm{colim}_{\mathcal{I}} FX$ exists in $\mathcal{D}$, and*

$$\mathrm{colim}_{\mathcal{I}}(FX) \cong F(\mathrm{colim}_{\mathcal{I}} X).$$

*If $Y : \mathcal{I} \to \mathcal{D}$ is a functor such that $\lim_{\mathcal{I}} Y$ exists in $\mathcal{D}$, then $\lim_{\mathcal{I}} GY$ exists in $\mathcal{C}$, and*

$$\lim_{\mathcal{I}}(GY) \cong G(\lim_{\mathcal{I}} Y).$$

*Proof.* We prove the result for colimits, and leave the case of limits to the reader. Using Definition 7.5.1 and Proposition 7.6.28, we obtain for any object $T$ in $\mathcal{D}$ a chain of isomorphisms

$$\begin{aligned} \mathrm{Hom}_{\mathcal{D}}(F(\mathrm{colim}_{\mathcal{I}} X_i), T) &\cong \mathrm{Hom}_{\mathcal{C}}(\mathrm{colim}_{\mathcal{I}} X_i, GT) \\ &\cong \lim_{\mathcal{I}^{\mathrm{opp}}} \mathrm{Hom}_{\mathcal{C}}(X_i, GT) \\ &\cong \lim_{\mathcal{I}^{\mathrm{opp}}} \mathrm{Hom}_{\mathcal{D}}(FX_i, T) \end{aligned}$$

functorial in $T$, whence a natural isomorphism

$$\mathrm{Hom}_{\mathcal{D}}(F(\mathrm{colim}_{\mathcal{I}} X_i), -) \overset{\sim}{\Longrightarrow} \lim_{\mathcal{I}^{\mathrm{opp}}} \mathrm{Hom}_{\mathcal{D}}(FX_i, -).$$

The result now follows from another application of Proposition 7.6.28. □

The list of applications of this result is endless! For example:

**Example 7.6.30.** Let $R$ be a ring, and let $N$ be an $R$-module. We know from Example 7.5.7 that the functor $- \otimes_R N : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}$ is left adjoint to the functor $\mathrm{Hom}_R(N,-) : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}$. Hence $- \otimes_R N$ commutes with coproducts (direct sums), pushouts and cokernels; the latter statement is simply a rephrasing of the fact that $- \otimes_R N$ is right exact. This reproves Propositions 6.1.9 and 6.4.2!

**Example 7.6.31.** The embedding functor $F : \mathsf{Ab} \to \mathsf{Grp}$, which forgets that an abelian group is actually abelian and only remembers that it is a group, is right adjoint to the abelianisation functor $\mathsf{ab} : \mathsf{Grp} \to \mathsf{Ab}$. Therefore $\mathsf{ab}$ commutes with all colimits. Hence we know that $(G \star H)^{\mathsf{ab}} \cong G^{\mathsf{ab}} \oplus H^{\mathsf{ab}}$ for arbitrary groups $G$ and $H$, and also that $\mathsf{ab}$ is right exact: if $G_1 \to G_2 \to G_3 \to 0$ is an exact sequence of groups, then so is the induced sequence $G_1^{\mathsf{ab}} \to G_2^{\mathsf{ab}} \to G_3^{\mathsf{ab}} \to 0$.

# Exercises

*Easy exercise* 7.1. Let $\mathcal{C}$ be a category. Given an object $X$ in $\mathcal{C}$, let

$$\mathrm{Aut}_{\mathcal{C}}(X) = \{f \in \mathrm{Hom}_{\mathcal{C}}(X, X) : f \text{ is an isomorphism}\}.$$

Show that $\mathrm{Aut}_{\mathcal{C}}(X)$ forms a group under composition, the *automorphism group* of $X$ in $\mathcal{C}$. Show also that if $X$ and $Y$ are isomorphic objects, then $\mathrm{Aut}_{\mathcal{C}}(X)$ and $\mathrm{Aut}_{\mathcal{C}}(Y)$ are isomorphic groups.

*Easy exercise* 7.2. Let $\mathcal{C}$ be an arbitrary category. Show that the composition of two monomorphisms (resp. epimorphisms) in $\mathcal{C}$ is again a monomorphism (resp. epimorphism). Conversely, if $f$ and $g$ are morphisms in $\mathcal{C}$ such that $fg$ is monic, show that $g$ is monic; if $fg$ is epic, show that $f$ is epic.

*Easy exercise* 7.3. A monomorphism $f : X \to Y$ in a category $\mathcal{C}$ is *split* if there exists $g \in \mathrm{Hom}_{\mathcal{C}}(Y, X)$ such that $gf = 1_X$ (such a $g$ is called a *retraction* of $f$). Similarly, an epimorphism $f : X \to Y$ in $\mathcal{C}$ is split if there exists $g \in \mathrm{Hom}_{\mathcal{C}}(Y, X)$ such that $fg = 1_Y$ (such a $g$ is called a *section* of $f$).

If $F : \mathcal{C} \to \mathcal{D}$ is a functor, show that

(a) if $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ is an isomorphism, so is $Ff$;

(b) if $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a split monomorphism (resp. split epimorphism), so is $Ff$;

(c) if $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ is monic (resp. epic), then $Ff$ need not be monic (resp. epic).

*Easy exercise* 7.4. Does the category Field of fields (which you can easily define yourself) have initial and/or final objects? What about the category $\mathsf{Field}_p$ of fields of characteristic $p > 0$, where $p$ is prime?

*Easy exercise* 7.5. Show that the forgetful functor Grp $\to$ Set is faithful, but not full; the embedding functor Ab $\to$ Grp fully faithful; the abelianisation functor Grp $\to$ Ab is neither faithful, nor full.

*Easy exercise* 7.6. Let $F : \mathcal{C} \to \mathcal{D}$ be an equivalence of categories, and let $G : \mathcal{D} \to \mathcal{C}$ be a quasi-inverse. Show that $F$ is both left and right adjoint to $G$.

*Easy exercise* 7.7. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor which is left adjoint to $G : \mathcal{D} \to \mathcal{C}$. If $X$ is an initial object in $\mathcal{C}$, show that $FX$ is initial in $\mathcal{D}$. If $Y$ is a final object in $\mathcal{D}$, show that $GY$ is final in $\mathcal{C}$.

*Easy exercise* 7.8. Consider the constant maps $0 : \{\star\} \to I$ and $1 : \{\star\} \to I$ in the category Top of topological spaces, where $I$ is the unit interval. Find the equaliser and coequaliser of these maps.

*Exercise* 7.9. Let $G$ be a group, and consider the category $\mathrm{B}G$ with one object.

(a) Show that a functor $F : \mathrm{B}G \to$ Set is nothing but a *$G$-set* (a set equipped with a $G$-action).

(b) Given functors $F, G : \mathrm{B}G \to$ Set, show that a natural transformation $\eta : F \Rightarrow G$ is a *$G$-equivariant map* of $G$-sets. (A map $f : S \to T$ of $G$-sets is $G$-equivariant if $f(gs) = gf(s)$ for all $s \in S$.)

(c) Describe $\lim_{\mathrm{B}G} F$ and $\mathrm{colim}_{\mathrm{B}G} F$ in terms of the group action on the corresponding $G$-set.

*Exercise* 7.10. Show that the inclusion map $\mathbf{Z} \hookrightarrow \mathbf{Q}$ is both a monomorphism and an epimorphism in the category Ring. Conclude that a morphism which is both monic and epic need not be an isomorphism.

*Exercise* 7.11. Let $\mathcal{C}$ be a category with the following properties:

- $\mathcal{C}$ is locally small, i.e., for all objects $X$ and $Y$ in $\mathcal{C}$, the class $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a set;
- all objects of $\mathcal{C}$ are isomorphic, and all morphisms in $\mathcal{C}$ are isomorphisms.

Show that there exists a group $G$ such that $\mathcal{C}$ is equivalent to $\mathrm{B}G$.

*Exercise* 7.12. Let $k$ be a field. The category $\mathsf{Mat}_k$ has as objects the non-negative integer $0, 1, 2, \cdots$. Morphisms from $m$ to $n$ by $\mathrm{Hom}_{\mathsf{Mat}_k}(m, n) = \{(n \times m)-\text{matrices over } k\}$, with composition given by

$$(A \in \mathrm{Hom}_{\mathsf{Mat}_k}(m, n), B \in \mathrm{Hom}_{\mathsf{Mat}_k}(n, p)) \mapsto BA \in \mathrm{Hom}_{\mathsf{Mat}_k}(m, p).$$

The identity morphism $1_n$ is the identity matrix of size $(n \times n)$.

(a) Show that $\mathsf{Mat}_k$ is indeed a category.

(b) Prove that $\mathsf{Mat}_k$ is equivalent to the category $\mathsf{fVec}_k$ of finite dimensional $k$-vector spaces.

*Exercise* 7.13. Let $F : \mathcal{C} \to \mathcal{D}$ be an equivalence of categories. Prove that a morphism $f$ in $\mathcal{C}$ is monic (resp. epic) if and only if the corresponding morphism $Ff$ in $\mathcal{D}$ is monic (resp. epic).

*Exercise* 7.14. Given an adjoint pair of functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$, with associated unit and counit $\eta : 1_{\mathcal{C}} \Rightarrow GF$ and $\varepsilon : FG \Rightarrow 1_{\mathcal{D}}$, show that $\eta$ and $\varepsilon$ satisfy the *unit-counit equations*

$$\varepsilon F \circ F\eta = 1_F, \;\; G\varepsilon \circ \eta G = 1_G$$

which should be read as follows: for all objects $X$ in $\mathcal{C}$ and $Y$ in $\mathcal{D}$, we have

$$\varepsilon_{FX} \circ F(\eta_X) = 1_{FX}, \;\; G(\varepsilon_Y) \circ \eta_{GY} = 1_{GY}.$$

*Exercise* 7.15. Let $\mathcal{C}$ be a category. Construct a *diagonal functor* $\Delta : \mathcal{C} \to \mathcal{C} \times \mathcal{C}$, which sends $X \in \mathrm{ob}(\mathcal{C})$ to $X \times X \in \mathrm{ob}(\mathcal{C} \times \mathcal{C})$. Find a necessary and sufficient criterion for $\Delta$ to have a right or left adjoint.

*Exercise* 7.16. Let $F : \mathcal{C} \to \mathcal{D}$ (resp. $F' : \mathcal{D} \to \mathcal{E}$) be a functor which is left adjoint to the functor $G : \mathcal{D} \to \mathcal{C}$ (resp. to $G' : \mathcal{E} \to \mathcal{D}$). Show that $F' \circ F : \mathcal{C} \to \mathcal{E}$ is left adjoint to $G \circ G' : \mathcal{E} \to \mathcal{C}$.

*Exercise* 7.17. Let $\mathcal{C}$ be a category, and let $f, g : X \to Y$ be two maps in $\mathcal{C}$. Show that the equaliser of $f$ and $g$ is monic (if it exists), and that the coequaliser of $f$ and $g$ is epic (if it exists).

*Exercise* 7.18. Let $\mathcal{C}^{\leq}_{\mathbf{R}}$ be the thin category associated to the ordered set $(\mathbf{R}, \leq)$. Define $\mathcal{C}^{\leq}_{\mathbf{Z}}$ similarly. Show that the "embedding functor" $E : \mathcal{C}^{\leq}_{\mathbf{Z}} \to \mathcal{C}^{\leq}_{\mathbf{R}}$ and "rounding up functor" $\lceil - \rceil : \mathcal{C}^{\leq}_{\mathbf{R}} \to \mathcal{C}^{\leq}_{\mathbf{Z}} : r \mapsto \lceil r \rceil$ (where $\lceil r \rceil$ is the smallest integer greater than or equal to $r$) form an adjoint pair.

*Exercise* 7.19. (a) Consider the contravariant power set functor $\mathcal{P} : \mathsf{Set}^{\mathrm{opp}} \to \mathsf{Set}$ which sends a set $S$ to its power set $\mathcal{P}(S)$, and which sends a map $f : S \to T$ to $f^\star : \mathcal{P}(T) \to \mathcal{P}(S) : X \mapsto f^{-1}(X)$. Show that $\mathcal{P}$ is representable.

(b) Consider the functor $\mathcal{T} : \mathsf{Top}^{\mathrm{opp}} \to \mathsf{Set}$ which sends a topological space $X$ to its topology $\mathcal{T}_X$, and which sends a continuous function $f : X \to Y$ to its pullback $f^\star : \mathcal{T}_Y \to \mathcal{T}_X : U \mapsto f^{-1}(U)$. Show that $\mathcal{T}$ is representable.

*Exercise* 7.20. Define, for every positive integer $n$, a functor $\mathsf{Nil}_n : \mathsf{Ring} \to \mathsf{Set}$ which sends a ring $R$ to the set $\{x \in R : x^n = 0\}$, and show that this functor is representable. Define similarly a functor $\mathsf{Nil} : \mathsf{Ring} \to \mathsf{Set}$ which sends a ring $R$ to its entire nilradical $\mathfrak{n}_R$, and show that it is *not* representable.

*Exercise* 7.21. (a) Show that the equaliser of two morphisms $f_1, f_2 : G \to H$ in $\mathsf{Grp}$ is the inclusion

$$\{g \in G : f_1(g) = f_2(g)\} \hookrightarrow G.$$

(b) Let $f : \mathbf{Z}/3 \to \mathcal{S}_3$ be an injective homomorphism. Let $g : \mathbf{Z}/3 \to \mathcal{S}_3$ be the trivial homomorphism. What is the equaliser of $f$ and $g$ in $\mathsf{Grp}$? What is the equaliser of $f^{\mathsf{ab}}$ and $g^{\mathsf{ab}}$ in $\mathsf{Ab}$?

(c) Conclude from (b) that the abelianisation functor $\mathsf{ab} : \mathsf{Grp} \to \mathsf{Ab}$ does not preserve general limits. Show also that $\mathsf{ab}$ nevertheless preserves products.

*Exercise* 7.22. Let $F : \mathcal{C} \to \mathsf{Set}$ be a functor which has a left adjoint. Show that $F$ is representable.

*Exercise* 7.23. Let $\mathcal{I}$ be an indexing category. Given diagrams $X, Y : \mathcal{I} \to \mathcal{C}$ in a category $\mathcal{C}$ together with a natural transformation $\eta : X \Rightarrow Y$, show that we have induced morphisms

$$\lim_{\mathcal{I}} \eta : \lim_{\mathcal{I}} X \to \lim_{\mathcal{I}} Y \text{ and } \operatorname{colim}_{\mathcal{I}} \eta : \operatorname{colim}_{\mathcal{I}} X \to \operatorname{colim}_{\mathcal{I}} Y$$

in $\mathcal{C}$, provided that the relevant (co)limits exist.

*Exercise* 7.24. Let $f : R \to S$ and $g : R \to T$ be morphisms in $\mathsf{Ring}$. Show that the $R$-algebra $S \otimes_R T$, equipped with the pair of morphisms $\iota_S : S \to S \otimes_R T : s \mapsto s \otimes 1$ and $\iota_T : T \to S \otimes_R T : t \mapsto 1 \otimes t$, is the pushout of $f$ and $g$ in $\mathsf{Ring}$.

*Exercise* 7.25. Let $I = (\mathbf{Z}_{\geq 0}, \leq)$, where $\leq$ is the natural order, and let $\mathcal{I}$ be the associated (thin) indexing category. Let $R$ be a ring. Let $F_R : \mathcal{I}^{\text{opp}} \to \mathsf{Ring}$ be the functor defined on objects by sending $m \in \mathbf{Z}_{\geq 0}$ to $R[X]/(X^m)$, and on morphisms by sending $m \leq n$ to the quotient map $R[X]/(X^n) \twoheadrightarrow R[X]/(X^m)$. Show that $\lim_{\mathcal{I}} F_R \cong R[\![X]\!]$.

*Exercise* 7.26. (a) Show that the forgetful functor $\operatorname{Forget} : \mathsf{Top} \to \mathsf{Set}$ admits both a left adjoint and a right adjoint, which implies in particular that it preserves arbitrary (co)limits.

(b) Show that arbitrary (co)limits exist in the category $\mathsf{Top}$ of topological spaces.
(*Construct these (co)limits by putting a suitable topology on the (co)limits of the underlying sets.*)

*Exercise* 7.27. Let $R$ be a ring. Show that arbitrary limits and colimits exist in the category ${}_R\mathsf{Mod}$. (*Construct limits as submodules of products, and colimits as quotients of coproducts.*)

*Hard exercise* 7.28. Show that the epimorphisms in $\mathsf{Grp}$ are the surjective group homomorphisms.

*Hard exercise* 7.29. Let $\mathcal{C}$ be a category in which all products and pullbacks exist. Show that equalisers exist in $\mathcal{C}$. Conversely, show that if all products and equalisers exist in $\mathcal{C}$, then pullbacks exist as well.

State and prove dual versions of these statements.

# Chapter 8

# Categories of modules

In the final chapter of this course, we combine the categorical language from Chapter 7 with our earlier adventures with commutative rings and modules over these, in order to obtain a better understanding of *categories of modules*. This will lead us to the study of (co)chain complexes and their (co)homology, and to the construction of *derived functors*, culminating in the definition of the Ext-functors.

## 8.1 Projective and injective modules

Two important technical tools for the study of categories of modules are the (dual) notions of *projective* and *injective* modules. Recall that if $R$ is a ring, and if $M$ is an $R$-module, then the functors $\operatorname{Hom}_R(M,-)$ and $\operatorname{Hom}_R(-,M)$ are *left exact*. For the covariant functor $\operatorname{Hom}_R(M,-)$, this means that any left exact sequence $0 \longrightarrow N_1 \xrightarrow{f} N_2 \xrightarrow{g} N_3$ is transformed into another left exact sequence

$$0 \longrightarrow \operatorname{Hom}_R(M,N_1) \xrightarrow{f_\star} \operatorname{Hom}_R(M,N_2) \xrightarrow{g_\star} \operatorname{Hom}_R(M,N_3)$$

where $f_\star = \operatorname{Hom}_R(M,f)$ and $g_\star = \operatorname{Hom}_R(M,g)$ are the *pushforwards* of $f$ and $g$. For the contravariant functor $\operatorname{Hom}_R(-,M)$, this means that any right (not left!) exact sequence $N_1 \xrightarrow{f} N_2 \xrightarrow{g} N_3 \longrightarrow 0$ is transformed into a left exact sequence

$$0 \longrightarrow \operatorname{Hom}_R(N_3,M) \xrightarrow{g^\star} \operatorname{Hom}_R(N_2,M) \xrightarrow{f^\star} \operatorname{Hom}_R(N_1,M)$$

where $f^\star = \operatorname{Hom}_R(f,M)$ and $g^\star = \operatorname{Hom}_R(g,M)$ are the *pullbacks* of $f$ and $g$.

Projective and injective modules make the covariant and contravariant Hom functors exact instead of only left exact, in the sense that they map arbitrary exact sequences to exact sequences. Let us spell out what this means, first for projective modules, then for injective modules.

**Proposition 8.1.1.** *Let $R$ be a ring and let $P$ be an $R$-module. The following statements are equivalent:*

*(a)* $\operatorname{Hom}_R(P,-)$ *preserves arbitrary exact sequences: if $N_1 \xrightarrow{f} N_2 \xrightarrow{g} N_3$ is exact, then so is*

$$\operatorname{Hom}_R(P,N_1) \xrightarrow{f_\star} \operatorname{Hom}_R(P,N_2) \xrightarrow{g_\star} \operatorname{Hom}_R(P,N_3);$$

*(b)* $\operatorname{Hom}_R(P,-)$ *preserves* short *exact sequences;*

*(c) if $f : M \to N$ is a surjective $R$-module homomorphism, so is $f_\star : \operatorname{Hom}_R(P,M) \to \operatorname{Hom}_R(P,N)$.*

**Definition 8.1.2.** Let $R$ be a ring and let $P$ be an $R$-module. If $P$ satisfies the equivalent conditions of Proposition 8.1.1, then $P$ is said to be a *projective* $R$-module.

Property (c) above means concretely that whenever we are given a surjective homomorphism of $R$-modules $f : M \to N$ and an arbitrary $R$-module homomorphism $g : P \to N$, then $g$ can be lifted to a map $h : P \to M$ such that $g = fh$, as in the following diagram:

$$\begin{tikzcd}
P \arrow[d, dashed, "h"] \arrow[dr, "g"] & & \\
M \arrow[r, two heads, "f"] & N \arrow[r] & 0
\end{tikzcd}$$

*Proof of Proposition 8.1.1.* The proof is almost identical to the proof of Proposition 6.4.5. □

**Example 8.1.3.** Let $R$ be a ring. If an $R$-module $P$ is free, then it is projective. Indeed, in the diagram above the desired lift can be constructed "by hand": given a basis $(x_b)_{b\in B}$ of $P$, choose for every $b \in B$ a preimage $y_b$ in $M$ of $g(x_b) \in N$, and define $h : P \to M$ by $h(x_b) = y_b$.

There are various alternative characterisations of projective $R$-modules, for example:

**Proposition 8.1.4.** *Let $R$ be a ring and let $P$ be an $R$-module. The following conditions are equivalent:*

*(a) $P$ is a projective $R$-module;*

*(b) every short exact sequence of the form $0 \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P \longrightarrow 0$ is split;*

*(c) $P$ is a direct summand of a free $R$-module, i.e., there exists an $R$-module $Q$ such that $P \oplus Q$ is free.*

*Proof.* If $P$ is projective, then Proposition 8.1.1, part (c) implies that $\mathrm{Id}_P : P \to P$ can be lifted to some homomorphism $h : P \to N$ such that $\mathrm{Id}_P = gh$. In the terminology of Proposition 2.4.8, this means that $g$ has a section, and hence that the given sequence splits. Hence (a) implies (b).

If (b) holds, choose a surjection $\pi : F \twoheadrightarrow P$ with $F$ a free $R$-module, and let $Q = \ker \pi$. Since the exact sequence $0 \to Q \hookrightarrow F \twoheadrightarrow P \to 0$ splits, we have $P \oplus Q \cong F$, hence (c) holds.

If (c) holds, choose an $R$-module $Q$ such that $P \oplus Q$ is free, and consider the obvious projection $\pi : P \oplus Q \twoheadrightarrow P$ and injection $\iota : P \hookrightarrow P \oplus Q$. Given a diagram of the form

$$\begin{tikzcd}
P \oplus Q \arrow[r, shift left, "\pi"] \arrow[dr, dashed, "\varphi"'] & P \arrow[l, shift left, hook, "\iota"] \arrow[d, dashed, "\varphi\iota"'] \arrow[dr, "g"] & & \\
& M \arrow[r, two heads, "f"] & N \arrow[r] & 0
\end{tikzcd}$$

where $f$ is surjective, we can find an $R$-module homomorphism $\varphi : P \oplus Q \to M$ such that $f\varphi = g\pi$, since $P \oplus Q$ is free. Then $\varphi\iota : P \to M$ is the desired lift of $g$, since $f(\varphi\iota) = (f\varphi)\iota = (g\pi)\iota = g$. □

**Example 8.1.5.** The (non-free) $\mathbf{Z}/6$-modules $\mathbf{Z}/2$ and $\mathbf{Z}/3$ are projective, since $\mathbf{Z}/6 \cong \mathbf{Z}/2 \oplus \mathbf{Z}/3$.

**Example 8.1.6.** Let $R$ be a ring, and assume that $R$ contains an *idempotent* element: an element $e$ for which $e^2 = e$. Then the principal ideal $(e)$ is a projective $R$-module; indeed, we have an $R$-module isomorphism $R \cong (e) \oplus (1 - e)$. This generalises the previous example: if $R = \mathbf{Z}/6$, then $-\bar{2}$ and $\bar{3}$ are idempotent elements which induce the direct sum decomposition $\mathbf{Z}/6 \cong (\bar{3}) \oplus (-\bar{2}) \cong \mathbf{Z}/2 \oplus \mathbf{Z}/3$.

**Proposition 8.1.7.** *Let $R$ be a ring and let $(P_\alpha)_{\alpha\in A}$ be a collection of $R$-modules. Then $\bigoplus_{\alpha\in A} P_\alpha$ is projective if and only if $P_\alpha$ is projective for every $\alpha \in A$.*

*Proof.* This follows from part (c) of Proposition 8.1.4: if each $P_\alpha$ is a direct summand of a free $R$-module $F_\alpha$, then $\bigoplus_{\alpha\in A} P_\alpha$ is a direct summand of the free $R$-module $\bigoplus_{\alpha\in A} F_\alpha$. Conversely, if $\bigoplus_{\alpha\in A} P_\alpha$ is a direct summand of the free $R$-module $F$, then clearly $P_\alpha$ is a direct summand of $F$ as well. □

*Quick question* 8.1.8. Give a different proof of Proposition 8.1.7, using both the universal property of the direct sum as a coproduct in ${}_R\mathsf{Mod}$ and part (c) of Proposition 8.1.1 (the lifting property).

**Proposition 8.1.9.** *Let $R$ be a ring. Then any projective $R$-module is flat.*

*Proof.* Since tensor products commute with direct sums (Proposition 6.1.9), a direct sum of $R$-modules is flat if and only if every summand is flat. Since free modules are flat (Quick question 6.4.4) and since projective $R$-modules are direct summands of free $R$-modules (Proposition 8.1.4), the result follows. □

The dual notion is the following:

**Definition 8.1.10.** Let $R$ be a ring. If an $R$-module $I$ satisfies the following (equivalent) conditions

(a) $\operatorname{Hom}_R(-,I)$ preserves arbitrary exact sequences: if $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$ is exact, then so is

$$\operatorname{Hom}_R(M_3,I) \xrightarrow{g^\star} \operatorname{Hom}_R(M_2,I) \xrightarrow{f^\star} \operatorname{Hom}_R(M_1,I);$$

(b) $\operatorname{Hom}_R(-,I)$ preserves *short* exact sequences;

(c) if $f: M \to N$ is injective, then $f^\star : \operatorname{Hom}_R(N,I) \to \operatorname{Hom}_R(M,I)$ is surjective,

then $I$ is said to be an *injective* $R$-module.

The fact that these conditions are equivalent can be proven in the same way as before. Property (c) means concretely that whenever we have an injective homomorphism of $R$-modules $f: M \to N$ and an arbitrary $R$-module homomorphism $g: M \to I$, then $g$ can be extended to an $R$-module homomorphism $h: N \to I$ such that $g = hf$, as in the following diagram:

$$\begin{tikzcd}
 & & I \\
0 \arrow[r] & M \arrow[ur, "g"] \arrow[r, hook, "f"] & N \arrow[u, dashed, "h"']
\end{tikzcd}$$

Let us collect some of the basic properties of injective modules.

**Proposition 8.1.11.** *Let $R$ be a ring, and let $I$ be an $R$-module. Then $I$ is injective if and only if every short exact sequence of the form $0 \longrightarrow I \xrightarrow{f} M \xrightarrow{g} N \longrightarrow 0$ is split.*

*Proof.* If $I$ is injective, then $\mathrm{Id}_I : I \to I$ can be extended to $h: M \to I$ such that $hf = \mathrm{Id}_I$. In the terminology of Proposition 2.4.8, this means that $f$ admits a retraction. Hence the given sequence splits.

Conversely, assume that every short exact sequence beginning with $I$ splits. Consider the diagram

$$\begin{tikzcd}
 & I \arrow[r, "\alpha"] & I \oplus_M N \\
0 \arrow[r] & M \arrow[u, "g"] \arrow[r, hook, "f"] & N \arrow[u, "\beta"'] \arrow[ul, dashed, "?"']
\end{tikzcd}$$

where $(I \oplus_M N, \alpha, \beta)$ is the pushout of $f$ and $g$ in ${}_R\mathsf{Mod}$. The description of the pushout given in Example 7.6.19 shows that $\alpha$ is injective. The short exact sequence $0 \longrightarrow I \xrightarrow{\alpha} I \oplus_M N \longrightarrow \operatorname{coker} \alpha \longrightarrow 0$ is split by assumption, which means that $\alpha$ admits a retraction $r : I \oplus_M N \to I$. Now the diagram

$$\begin{tikzcd}
 & I \arrow[r, shift right, "\alpha"'] & I \oplus_M N \arrow[l, shift right, "r"'] \\
0 \arrow[r] & M \arrow[u, "g"] \arrow[r, hook, "f"'] & N \arrow[u, "\beta"] \arrow[ul, dashed, "r\beta"]
\end{tikzcd}$$

commutes: we have $(r\beta)f = r(\beta f) = r(\alpha g) = (r\alpha)g = g$. Hence $r\beta : N \to I$ extends $g : M \to I$, showing that $I$ is injective. □

**Proposition 8.1.12.** *Let $R$ be a ring and let $(I_\alpha)_{\alpha \in A}$ be a collection of $R$-modules. Then $\prod_{\alpha \in A} I_\alpha$ is injective if and only if $I_\alpha$ is injective for every $\alpha \in A$.*

*Proof.* Let $f : M \to N$ be an injective $R$-module homomorphism, and let $g : M \to \prod_{\alpha \in A} I_\alpha$ be an arbitrary $R$-module homomorphism. Then $\pi_\alpha g : M \to I_\alpha$ can be extended to $h_\alpha : N \to I_\alpha$. It is now easy to check that the various $h_\alpha$ together define an extension $h : N \to \prod_{\alpha \in A} I_\alpha$ of $g$. □

There are various alternative characterisations of injectivity, such as [Rotman, Theorem 7.68]. There are also strong links with the notion of *divisibility* [Rotman, Lemma 7.72 and Corollary 7.73]. We will not use these results. However, we need the following fact, the proof of which is surprisingly intricate; hence we only cover the statement (interested readers may consult [Rotman, Theorem 8.104] for the proof).

**Theorem 8.1.13.** *Let $R$ be a ring. Any $R$-module can be embedded in an injective $R$-module.*

Note that the dual statement, namely that any $R$-module is a quotient of a projective $R$-module, is a triviality: any $R$-module is a quotient of a free $R$-module, and free $R$-modules are projective.

## 8.2 Complexes and their (co)homology

The basic building blocks of *homological algebra* (to which this chapter is an introduction) are *complexes of $R$-modules*, which traditionally come in two flavours: chain complexes and cochain complexes.

**Definition 8.2.1.** Let $R$ be a ring. A *chain complex* of $R$-modules is a sequence of the form

$$M_\bullet = \left( \cdots \longrightarrow M_2 \xrightarrow{d_2} M_1 \xrightarrow{d_1} M_0 \xrightarrow{d_0} M_{-1} \xrightarrow{d_{-1}} M_{-2} \longrightarrow \cdots \right)$$

where each term $M_i$ is an $R$-module (the *degree $i$ term* of the complex), and each $d_i : M_i \to M_{i-1}$ is an $R$-module homomorphism (the *degree $i$ differential*), such that $d_{i-1} \circ d_i = 0$ for all $i \in \mathbf{Z}$.

We will often write $d$ rather than $d_i$ for the various differentials, and $d_{i-1} \circ d_i = 0$ will then be abbreviated to $d^2 = 0$. A complex $M_\bullet$ is said to be *concentrated* in positive (resp. negative) degrees if $M_i = 0$ for all $i < 0$ (resp. $i > 0$); if this is the case, we will typically refrain from writing the terms in negative (resp. positive) degrees at all, since they are equal to 0 anyway. The dual notion is the following.

**Definition 8.2.2.** Let $R$ be a ring. A *cochain complex* of $R$-modules is a sequence of the form

$$M^\bullet = \left( \cdots \longrightarrow M^{-2} \xrightarrow{d^{-2}} M^{-1} \xrightarrow{d^{-1}} M^0 \xrightarrow{d^0} M^1 \xrightarrow{d^1} M^2 \longrightarrow \cdots \right)$$

where each term $M^i$ is an $R$-module (the *degree $i$ term* of the complex), and each $d^i : M^i \to M^{i+1}$ is an $R$-module homomorphism (the *degree $i$ differential*), such that $d^{i+1} \circ d^i = 0$ for all $i \in \mathbf{Z}$.

The difference between the notions of chain complex and cochain complex is purely notational: if $M_\bullet$ is a chain complex, then taking $M^i = M_{-i}$ and $d^i = d_{-i}$ defines a cochain complex, and vice versa.

For (co)chain complexes we will now define the associated (co)homology modules:

**Definition 8.2.3.** Let $M_\bullet$ be a chain complex of $R$-modules. The *i-th homology module* of $M_\bullet$ is

$$H_i(M_\bullet) = \frac{\ker(d_i : M_i \to M_{i-1})}{\operatorname{im}(d_{i+1} : M_{i+1} \to M_i)}.$$

Similarly, if $M^\bullet$ is a cochain complex of $R$-modules, the *i-th cohomology module* of $M^\bullet$ is

$$H^i(M^\bullet) = \frac{\ker(d^i : M^i \to M^{i+1})}{\operatorname{im}(d^{i-1} : M^{i-1} \to M^i)}.$$

Note that for a chain complex $M_\bullet$, we have $\operatorname{im}(d_{i+1}) \subseteq \ker(d_i)$ since $d_i \circ d_{i+1} = 0$. The equality $H_i(M_\bullet) = 0$ therefore means that $M_\bullet$ is exact in degree $i$. Hence the homology modules of $M_\bullet$ measure the failure of exactness of $M_\bullet$ in various degrees (and similarly for the cohomology modules).

**Definition 8.2.4.** A *chain map* $f_\bullet : M_\bullet \to N_\bullet$ of chain complexes of $R$-modules is a collection of $R$-module homomorphisms $(f_i : M_i \to N_i)_{i \in \mathbf{Z}}$ which makes the following diagram commute:

$$\begin{tikzcd}
\cdots \arrow[r] & M_{i+1} \arrow[r, "d_{i+1}^M"] \arrow[d, "f_{i+1}"] & M_i \arrow[r, "d_i^M"] \arrow[d, "f_i"] & M_{i-1} \arrow[r] \arrow[d, "f_{i-1}"] & \cdots \\
\cdots \arrow[r] & N_{i+1} \arrow[r, "d_{i+1}^N"] & N_i \arrow[r, "d_i^N"] & N_{i-1} \arrow[r] & \cdots
\end{tikzcd}$$

A cochain map of cochain complexes is defined similarly.

We observe that (co)chain maps of (co)chain complexes induce homomorphisms on (co)homology:

**Lemma 8.2.5.** *A chain map $f_\bullet : M_\bullet \to N_\bullet$ of chain complexes of $R$-modules induces, for every $i \in \mathbf{Z}$, an $R$-module homomorphism $H_i(f_\bullet) : H_i(M_\bullet) \to H_i(N_\bullet)$ on the $i$-th degree homology, given by*

$$x + \operatorname{im}(d_{i+1}^M) \mapsto f_i(x) + \operatorname{im}(d_{i+1}^N).$$

We leave it to the reader to check that this is indeed a well-defined $R$-module homomorphism.

**Definition 8.2.6.** Let $R$ be a ring. The category ${}_R\mathsf{Ch}$ has objects chain complexes of $R$-modules, and morphisms chain maps between chain complexes. Similarly, the category ${}_R\mathsf{CoCh}$ has objects cochain complexes of $R$-modules, and morphisms cochain maps between cochain complexes.

**Definition 8.2.7.** Let $R$ be ring. The *i-th homology functor* $H_i(-) : {}_R\mathsf{Ch} \to {}_R\mathsf{Mod}$ sends a chain complex $M_\bullet$ to its degree $i$ homology module $H_i(M_\bullet)$, and a chain map $f_\bullet : M_\bullet \to N_\bullet$ to the induced $R$-module homomorphism $H_i(f_\bullet) : H_i(M_\bullet) \to H_i(N_\bullet)$. The *i-th cohomology functor* is defined similarly.

We leave it to the reader to check that these are indeed functors.

**Definition 8.2.8.** Let $R$ be a ring. A sequence $M_\bullet \xrightarrow{f_\bullet} N_\bullet \xrightarrow{g_\bullet} P_\bullet$ of chain maps between chain complexes of $R$-modules is said to be *exact in degree i* if $M_i \xrightarrow{f_i} N_i \xrightarrow{g_i} P_i$ is exact, and *exact* if it is exact in all degrees. A *short exact sequence of chain complexes* is simply an exact sequence of the form $0 \longrightarrow M_\bullet \xrightarrow{f_\bullet} N_\bullet \xrightarrow{g_\bullet} P_\bullet \longrightarrow 0$, where 0 denotes the complex which is zero in all degrees.

For cochain complexes, the definitions are identical.

We have the following crucial result (which is essentially an application of the snake lemma. . . ):

**Theorem 8.2.9.** *Let $R$ be a ring, and let $0 \longrightarrow M_\bullet \xrightarrow{f_\bullet} N_\bullet \xrightarrow{g_\bullet} P_\bullet \longrightarrow 0$ be a short exact sequence of chain complexes of $R$-modules. Then there exist $R$-module homomorphisms $\partial_i : H_i(P_\bullet) \to H_{i-1}(M_\bullet)$ for all $i \in \mathbf{Z}$ (called* boundary maps*) such that we have a* long exact homology sequence *of the form*

$$\begin{tikzcd}
 & \cdots \arrow[r] & H_{i+1}(P_\bullet) \arrow[dll, "\partial_{i+1}"] \\
H_i(M_\bullet) \arrow[r, "H_i(f_\bullet)"] & H_i(N_\bullet) \arrow[r, "H_i(g_\bullet)"] & H_i(P_\bullet) \arrow[dll, "\partial_i"] \\
H_{i-1}(M_\bullet) \arrow[r, "H_{i-1}(f_\bullet)"] & H_{i-1}(N_\bullet) \arrow[r, "H_{i-1}(g_\bullet)"] & H_{i-1}(P_\bullet) \arrow[dll, "\partial_{i-1}"] \\
H_{i-2}(M_\bullet) \arrow[r] & \cdots &
\end{tikzcd}$$

*Moreover this long exact homology sequence is natural in the following sense: if we have a commutative diagram of short exact sequences (involving chain complexes of $R$-modules and chain maps)*

$$\begin{tikzcd}
0 \arrow[r] & M_\bullet \arrow[r, "f_\bullet"] \arrow[d, "\varphi_{M,\bullet}"] & N_\bullet \arrow[r, "g_\bullet"] \arrow[d, "\varphi_{N,\bullet}"] & P_\bullet \arrow[r] \arrow[d, "\varphi_{P,\bullet}"] & 0 \\
0 \arrow[r] & M'_\bullet \arrow[r, "f'_\bullet"] & N'_\bullet \arrow[r, "g'_\bullet"] & P'_\bullet \arrow[r] & 0
\end{tikzcd}$$

*then the maps $\partial_i : H_i(P_\bullet) \to H_{i-1}(M_\bullet)$ and $\partial'_i : H_i(P'_\bullet) \to H_{i-1}(M'_\bullet)$ induce a commutative diagram*

$$\begin{tikzcd}
\cdots \arrow[r] & H_i(M_\bullet) \arrow[r, "H_i(f_\bullet)"] \arrow[d, "H_i(\varphi_{M,\bullet})"] & H_i(N_\bullet) \arrow[r, "H_i(g_\bullet)"] \arrow[d, "H_i(\varphi_{N,\bullet})"] & H_i(P_\bullet) \arrow[r, "\partial_i"] \arrow[d, "H_i(\varphi_{P,\bullet})"] & H_{i-1}(M_\bullet) \arrow[r] \arrow[d, "H_{i-1}(\varphi_{M,\bullet})"] & \cdots \\
\cdots \arrow[r] & H_i(M'_\bullet) \arrow[r, "H_i(f'_\bullet)"] & H_i(N'_\bullet) \arrow[r, "H_i(g'_\bullet)"] & H_i(P'_\bullet) \arrow[r, "\partial'_i"] & H_{i-1}(M'_\bullet) \arrow[r] & \cdots
\end{tikzcd}$$

*Quick question* 8.2.10. State the analogous result for cochain complexes and their cohomology.

*Proof.* We provide the main steps, and leave some of the details to the reader.

We apply the snake lemma (Proposition 2.4.11) twice, first to the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & M_{i+1} \arrow[r, "f_{i+1}"] \arrow[d, "d"] & N_{i+1} \arrow[r, "g_{i+1}"] \arrow[d, "d"] & P_{i+1} \arrow[r] \arrow[d, "d"] & 0 \\
0 \arrow[r] & M_i \arrow[r, "f_i"] & N_i \arrow[r, "g_i"] & P_i \arrow[r] & 0
\end{tikzcd}$$

to obtain an exact sequence

$$\begin{tikzcd}
0 \arrow[r] & \ker(M_{i+1} \to M_i) \arrow[r] & \ker(N_{i+1} \to N_i) \arrow[r] & \ker(P_{i+1} \to P_i) \arrow[dll, rounded corners] \\
 & \operatorname{coker}(M_{i+1} \to M_i) \arrow[r] & \operatorname{coker}(N_{i+1} \to N_i) \arrow[r] & \operatorname{coker}(P_{i+1} \to P_i) \arrow[r] & 0
\end{tikzcd}$$

Pasting two such sequences together yields a commutative diagram of the form

$$\begin{tikzcd}
 & \operatorname{coker}(M_{i+1} \to M_i) \arrow[r] \arrow[d] & \operatorname{coker}(N_{i+1} \to N_i) \arrow[r] \arrow[d] & \operatorname{coker}(P_{i+1} \to P_i) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \ker(M_{i-1} \to M_{i-2}) \arrow[r] & \ker(N_{i-1} \to N_{i-2}) \arrow[r] & \ker(P_{i-1} \to P_{i-2}) &
\end{tikzcd}$$

with exact rows, and vertical maps induced by the differentials. Since the kernel and cokernel of the vertical map $\operatorname{coker}(M_{i+1} \to M_i) \to \ker(M_{i-1} \to M_{i-2})$ induced by the differential $d : M_i \to M_{i-1}$ are $H_i(M_\bullet)$ and $H_{i-1}(M_\bullet)$ respectively, and similarly for the other two vertical maps; hence a second application of the snake lemma yields an exact sequence of the form

$$\begin{tikzcd}
H_i(M_\bullet) \arrow[r, "H_i(f_\bullet)"] & H_i(N_\bullet) \arrow[r, "H_i(g_\bullet)"] & H_i(P_\bullet) \arrow[dll, rounded corners, "\partial_i"] \\
H_{i-1}(M_\bullet) \arrow[r, "H_{i-1}(f_\bullet)"] & H_{i-1}(N_\bullet) \arrow[r, "H_{i-1}(g_\bullet)"] & H_{i-1}(P_\bullet)
\end{tikzcd}$$

Pasting such sequences together yields the desired long exact homology sequence. Naturality of this sequence can be checked using the explicit description of the boundary maps $\partial_i : H_i(P_\bullet) \to H_{i-1}(M_\bullet)$, described in the proof of Proposition 2.4.11. (See also [Rotman, Theorem 10.44] for this argument.) □

Another important building block for our theory is the notion of *homotopic chain maps*.

**Definition 8.2.11.** Let $f_\bullet, g_\bullet : M_\bullet \to N_\bullet$ be chain maps between chain complexes of $R$-modules. A *chain homotopy* from $f_\bullet$ to $g_\bullet$ is a collection of $R$-module homomorphisms $(h_i : M_i \to N_{i+1})_{i \in \mathbf{Z}}$ such that $g_i - f_i = d^N_{i+1} \circ h_i + h_{i-1} \circ d^M_i$ for all $i \in \mathbf{Z}$, with maps as in the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & M_{i+1} \arrow[r, "d^M_{i+1}"] \arrow[d, shift right, "f_{i+1}"'] \arrow[d, shift left, "g_{i+1}"] & M_i \arrow[r, "d^M_i"] \arrow[d, shift right, "f_i"'] \arrow[d, shift left, "g_i"] \arrow[dl, dashed, "h_i"] & M_{i-1} \arrow[r] \arrow[d, shift right, "f_{i-1}"'] \arrow[d, shift left, "g_{i-1}"] \arrow[dl, dashed, "h_{i-1}"] & \cdots \\
\cdots \arrow[r] & N_{i+1} \arrow[r, "d^N_{i+1}"'] & N_i \arrow[r, "d^N_i"'] & N_{i-1} \arrow[r] & \cdots
\end{tikzcd}$$

If such a chain homotopy exists, then $f_\bullet$ and $g_\bullet$ are said to be *chain homotopic* (denoted by $f_\bullet \sim g_\bullet$).

As usual, essentially the same definition applies to cochain complexes.

**Proposition 8.2.12.** *With notation as above, the relation* $\sim$ *is an equivalence relation on* $\operatorname{Hom}_{{}_R\mathsf{Ch}}(M_\bullet, N_\bullet)$.

*Proof.* Reflexivity and symmetry are obvious. To prove transitivity, note that if $(h_i : M_i \to N_{i+1})_{i \in \mathbf{Z}}$ defines a chain homotopy from $f_\bullet$ to $f'_\bullet$, and if $(k_i : M_i \to N_{i+1})_{i \in \mathbf{Z}}$ defines a chain homotopy from $f'_\bullet$ to $f''_\bullet$, then $(h_i + k_i : M_i \to N_{i+1})_{i \in \mathbf{Z}}$ defines a chain homotopy from $f_\bullet$ to $f''_\bullet$. □

**Proposition 8.2.13.** *Let $f_\bullet, g_\bullet : M_\bullet \to N_\bullet$ be chain maps between chain complexes of $R$-modules. Assume that $f_\bullet \sim g_\bullet$. If $\varphi_\bullet : L_\bullet \to M_\bullet$ is a chain map of chain complexes, then $f_\bullet \circ \varphi_\bullet \sim g_\bullet \circ \varphi_\bullet$. Similarly, if $\psi_\bullet : N_\bullet \to P_\bullet$ is a chain map of chain complexes, then $\psi_\bullet \circ f_\bullet \sim \psi_\bullet \circ g_\bullet$.*

*Proof.* If $(h_i : M_i \to N_{i+1})_{i \in \mathbf{Z}}$ defines a chain homotopy from $f_\bullet$ to $g_\bullet$, then $(h_i \circ \varphi_i : L_i \to N_{i+1})_{i \in \mathbf{Z}}$ defines a chain homotopy from $f_\bullet \circ \varphi_\bullet$ to $g_\bullet \circ \varphi_\bullet$. Indeed, for all $i \in \mathbf{Z}$, we have

$$g_i \circ \varphi_i - f_i \circ \varphi_i = (d^N_{i+1} \circ h_i + h_{i-1} \circ d^M_i) \circ \varphi_i = d^N_{i+1} \circ (h_i \circ \varphi_i) + (h_{i-1} \circ \varphi_{i-1}) \circ d^L_i.$$

The second statement is proven similarly. □

We are now able to define the *homotopy category of chain complexes of $R$-modules*:

**Definition 8.2.14.** Let $R$ be a ring. The *homotopy category* of chain complexes of $R$-modules, denoted by ${}_R\mathsf{HoCh}$, has the same objects as ${}_R\mathsf{Ch}$. Morphisms in ${}_R\mathsf{HoCh}$ are chain homotopy equivalence classes of morphisms in ${}_R\mathsf{Ch}$, i.e., $\mathrm{Hom}_{{}_R\mathsf{HoCh}}(M_\bullet, N_\bullet) = \mathrm{Hom}_{{}_R\mathsf{Ch}}(M_\bullet, N_\bullet)/\!\sim$. The composition of two such equivalence classes $[f_\bullet : M_\bullet \to N_\bullet]$ and $[g_\bullet : N_\bullet \to P_\bullet]$ is simply $[g_\bullet \circ f_\bullet : M_\bullet \to P_\bullet]$.

*Quick question* 8.2.15. Check, using Proposition 8.2.13, that composition in ${}_R\mathsf{HoCh}$ is well-defined.

We have a canonical functor ${}_R\mathsf{Ch} \to {}_R\mathsf{HoCh}$ which is the identity on objects, and which sends a chain map $f_\bullet$ to the corresponding equivalence class $[f_\bullet]$. This functor is full, but not faithful, and in fact this is exactly why we care about ${}_R\mathsf{HoCh}$: it turns out that certain natural constructions are ambiguous in ${}_R\mathsf{Ch}$, but canonical in ${}_R\mathsf{HoCh}$, as we will see in §8.3. Another argument in favour of ${}_R\mathsf{HoCh}$ is the fact that (co)homology does not "see" the difference between homotopic maps, i.e., the homology functors $H_i : {}_R\mathsf{Ch} \to {}_R\mathsf{Mod}$ factor over ${}_R\mathsf{Ch} \to {}_R\mathsf{HoCh}$, and similarly for the cohomology functors:

**Proposition 8.2.16.** *Let $f_\bullet, g_\bullet : M_\bullet \to N_\bullet$ be two chain maps between chain complexes of $R$-modules. If $f_\bullet \sim g_\bullet$, then $H_i(f_\bullet) = H_i(g_\bullet)$ for all $i \in \mathbf{Z}$.*

*Proof.* If $(h_i : M_i \to N_{i+1})_{i\in\mathbf{Z}}$ defines a homotopy from $f$ to $g$ and $x \in \ker(d_i^M : M_i \to M_{i-1})$, then

$$\begin{aligned} H_i(g_\bullet)(x + \mathrm{im}(d_{i+1}^M)) &= g_i(x) + \mathrm{im}(d_{i+1}^N) \\ &= f_i(x) + d_{i+1}^N(h_i(x)) + h_{i-1}(d_i^M(x)) + \mathrm{im}(d_{i+1}^N) \\ &= f_i(x) + \mathrm{im}(d_{i+1}^N) \\ &= H_i(f_\bullet)(x + \mathrm{im}(d_{i+1}^M)). \end{aligned}$$

□

**Corollary 8.2.17.** *The homology functor $H_i(-) : {}_R\mathsf{Ch} \to {}_R\mathsf{Mod}$ factors over the canonical functor ${}_R\mathsf{Ch} \to {}_R\mathsf{HoCh}$, and therefore induces a functor ${}_R\mathsf{HoCh} \to {}_R\mathsf{Mod}$ again denoted by $H_i(-)$. Similarly, we obtain cohomology functors $H^i(-) : {}_R\mathsf{HoCoCh} \to {}_R\mathsf{Mod}$.*

## 8.3 Resolution functors

The central notion of this section is the following.

**Definition 8.3.1.** Let $R$ be a ring, and let $M$ be an $R$-module. A *resolution* of $M$ is a pair $(P_\bullet, \alpha)$, where $P_\bullet$ is an object in ${}_R\mathsf{Ch}$ concentrated in non-negative degrees, and where $\alpha : P_0 \to M$ is an $R$-module homomorphism (called the *augmentation map*), such that the sequence

$$\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \xrightarrow{\alpha} M \longrightarrow 0$$

is exact. Similarly, a *coresolution* of $M$ is a pair $(Q^\bullet, \beta)$, where $Q^\bullet$ is an object in ${}_R\mathsf{CoCh}$ concentrated in non-negative degrees, and where $\beta : M \to Q^0$ is an $R$-module homomorphism, such that the sequence

$$0 \longrightarrow M \xrightarrow{\beta} Q^0 \longrightarrow Q^1 \longrightarrow Q2 \longrightarrow \cdots$$

is exact.

A (co)resolution is said to have property P if all terms of the corresponding (co)chain complex have property P. For example, the terms appearing in a free resolution of an $R$-module $M$ are free $R$-modules, and the terms in an injective coresolution of $M$ are injective $R$-modules. Arbitrary resolutions are not very interesting, so we will restrict ourselves to the study of resolutions with favourable properties.

*Remark* 8.3.2. Let $R$ be a ring. Then every $R$-module $M$ has a free (*a fortiori* projective) resolution. Indeed, $M$ can be written as the quotient of a free $R$-module; let us therefore choose a surjective $R$-module homomorphism $\alpha : F_0 \twoheadrightarrow M$. Now the $R$-module $\ker(\alpha)$ is again a quotient of a free $R$-module; let us therefore choose a surjective $R$-module homomorphism $d_1 : F_1 \twoheadrightarrow \ker(\alpha)$. Then the sequence $F_1 \xrightarrow{d_1} F_0 \xrightarrow{\alpha} M \longrightarrow 0$ is exact (why?). Continuing this way, we get a free resolution $(F_\bullet, \alpha)$ of $M$.

*Quick question* 8.3.3. Show, using Theorem 8.1.13, that any $R$-module admits an injective coresolution.

*Remark* 8.3.4. Let $R$ be a PID, and let $M$ be a finitely generated $R$-module. Then $M$ has a *free resolution of length* 1, which means that the non-zero terms in the resolution are to be found in degree at most 1.

Indeed, choose a free $R$-module of finite rank $F_0$ which surjects onto $M$, via $\alpha : F_0 \twoheadrightarrow M$. Then the submodule $F_1 = \ker(\alpha)$ of $F_0$ is again free by Proposition 4.1.2. Therefore the exact sequence

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow F_1 \hookrightarrow F_0 \xrightarrow{\alpha} M \longrightarrow 0$$

yields a free resolution of $M$ with non-zero terms in degrees 0 and 1 only. This also shows that a resolution generalises the idea of a *presentation* using generator and relations.

If $R$ is not a PID, then a finitely generated $R$-module need not have a free resolution of finite length. Indeed, the $\mathbf{Z}/4$-module $\mathbf{Z}/2$ has a free resolution of infinite length consisting of modules of rank 1:

$$\cdots \xrightarrow{\times 2} \mathbf{Z}/4 \xrightarrow{\times 2} \mathbf{Z}/4 \xrightarrow{\times 2} \mathbf{Z}/4 \longrightarrow \mathbf{Z}/2 \longrightarrow 0$$

and one can show that *any* free resolution must have infinite length in this case (Exercise 8.3).

*Quick question* 8.3.5. Let $R = \mathbf{C}[X, Y]$ and $\mathfrak{m} = (X, Y)$. Find a free resolution of length 2 of the $R$-module $R/\mathfrak{m}$, i.e. find free $R$-modules $F_0$, $F_1$ and $F_2$ which fit into an exact sequence of the form

$$0 \longrightarrow F_2 \longrightarrow F_1 \longrightarrow F_0 \longrightarrow R/\mathfrak{m} \longrightarrow 0.$$

Given an $R$-module $M$, there are typically "many" different resolutions of $M$ in ${}_R\mathsf{Ch}$ with favourable properties, e.g., free resolutions – for the simple reason that a module can be written as the quotient of a free module in many different ways. Of course it would be desirable and practical to have a distinguished free resolution of $M$ in ${}_R\mathsf{Ch}$ *up for grabs*, for example a "minimal" one. While it is certainly possible to formalise this idea, it turns out to be equally efficient to consider resolutions up to homotopy, in other words: in ${}_R\mathsf{HoCh}$ rather than in ${}_R\mathsf{Ch}$. Indeed, many of the ambiguities in the choice of a free resolution disappear in ${}_R\mathsf{HoCh}$; to make this precise, we need the following key result.

**Proposition 8.3.6.** *Let $R$ be a ring, and let $f : M \to N$ be a homomorphism of $R$-modules. If $(P_\bullet, \alpha)$ is a* projective *resolution of $M$ and if $(Q_\bullet, \beta)$ is an arbitrary resolution of $N$, then there exists a chain map $g_\bullet : P_\bullet \to Q_\bullet$ such that $f \circ \alpha = \beta \circ g_0$ – in other words, such that the following diagram commutes:*

$$\begin{tikzcd}
\cdots \arrow[r] & P_2 \arrow[r] \arrow[d, "g_2"] & P_1 \arrow[r] \arrow[d, "g_1"] & P_0 \arrow[r, "\alpha"] \arrow[d, "g_0"] & M \arrow[r] \arrow[d, "f"] & 0 \\
\cdots \arrow[r] & Q_2 \arrow[r] & Q_1 \arrow[r] & Q_0 \arrow[r, "\beta"'] & N \arrow[r] & 0
\end{tikzcd}$$

*Moreover $g_\bullet$ is unique in ${}_R\mathsf{HoCh}$: if $g_\bullet$ and $g'_\bullet$ are chain maps which make the diagram commute, then there exist $R$-module homomorphisms $(h_i : P_i \to Q_{i+1})_{i \in \mathbf{Z}_{\geq 0}}$ defining a chain homotopy from $g_\bullet$ to $g'_\bullet$.*

*Proof.* Let us first prove the existence of $g_\bullet$ in ${}_R\mathsf{Ch}$. Since $\beta : Q_0 \to N$ is surjective and $P_0$ is projective, the composition $f \circ \alpha : P_0 \to N$ can be lifted to $g_0 : P_0 \to Q_0$ such that $f \circ \alpha = \beta \circ g_0$.

Now consider the diagram

$$\begin{tikzcd}
P_1 \arrow[r] \arrow[d, dashed, "g_1"] \arrow[dr] & \ker(\alpha) \arrow[d] \\
Q_1 \arrow[r] & \ker(\beta)
\end{tikzcd}$$

in which the horizontal maps (induced by the differentials $P_1 \to P_0$ and $Q_1 \to Q_0$) are surjective, and the right vertical arrow is the restriction of $g_0 : P_0 \to Q_0$ (why are these maps well-defined?). Now the diagonal composition $P_1 \to \ker(\alpha) \to \ker(\beta)$ can be lifted to $g_1 : P_1 \to Q_1$, since $P_1$ is projective. Repeating this argument yields a chain map $g_\bullet : P_\bullet \to Q_\bullet$ with the desired properties.

Let us now prove that the morphism $g_\bullet$ is unique in ${}_R\mathsf{HoCh}$. Assume that we are given two morphisms $g_\bullet, g'_\bullet : M_\bullet \to N_\bullet$ in ${}_R\mathsf{Ch}$ which make the following diagram commute:

$$\begin{tikzcd}
\cdots \arrow[r] & P_2 \arrow[r, "d_2^P"] \arrow[d, shift right, "g_2'"'] \arrow[d, shift left, "g_2"] & P_1 \arrow[r, "d_1^P"] \arrow[d, shift right, "g_1'"'] \arrow[d, shift left, "g_1"] \arrow[dl, dashed, blue, "h_1"'] & P_0 \arrow[r, "\alpha"] \arrow[d, shift right, "g_0'"'] \arrow[d, shift left, "g_0"] \arrow[dl, dashed, blue, "h_0"'] & M \arrow[r] \arrow[d, "f"] & 0 \\
\cdots \arrow[r] & Q_2 \arrow[r, "d_2^Q"'] & Q_1 \arrow[r, "d_1^Q"'] & Q_0 \arrow[r, "\beta"'] & N \arrow[r] & 0
\end{tikzcd}$$

The challenge is to construct a homotopy $(h_i : P_i \to Q_{i+1})_{i \in \mathbf{Z}_{\geq 0}}$ from $g_\bullet$ to $g'_\bullet$. This will again be done inductively. Set $\delta_\bullet = g'_\bullet - g_\bullet$. The image of $\delta_0 : P_0 \to Q_0$ is contained in $\ker(\beta)$ (why?). Since the map $Q_1 \to \ker(\beta)$ (induced by the differential $Q_1 \to Q_0$) is surjective and since $P_0$ is projective, we can lift the map $P_0 \to \ker(\beta)$ (induced by $\delta_0$) to $h_0 : P_0 \to Q_1$.

Let us now construct $h_1$. The image of the map $\delta_1 - h_0 \circ d_1^P : P_1 \to Q_1$ lies in $\ker(d_1^Q)$ (why?). Since the map $Q_2 \to \ker(d_1^Q)$ (induced by $d_2^Q$) is surjective and since $P_1$ is projective, the map $P_1 \to \ker(d_1^Q)$ induced by $\delta_1 - h_0 \circ d_1^P$ can be lifted to $h_1 : P_1 \to Q_2$ satisfying $d_2^Q \circ h_1 = \delta_1 - h_0 \circ d_1^P$, or equivalently,

$$g_1' - g_1 = d_2^Q \circ h_1 + h_0 \circ d_1^P.$$

Repeating this argument yields the desired homotopy $(h_i : P_i \to Q_{i+1})_{i \in \mathbf{Z}_{\geq 0}}$ from $g_\bullet$ to $g'_\bullet$. □

Proposition 8.3.6 implies that projective resolutions are "essentially unique up to homotopy". To make this statement rigorous, we need the following piece of terminology.

**Definition 8.3.7.** Let $R$ be a ring. A chain map $f_\bullet : M_\bullet \to N_\bullet$ is said to be a *homotopy equivalence* if (the equivalence class of) $f_\bullet$ becomes an isomorphism in ${}_R\mathsf{HoCh}$. Two chain complexes $M_\bullet$ and $N_\bullet$ are said to be *homotopy equivalent* if there exists a homotopy equivalence $f_\bullet : M_\bullet \to N_\bullet$.

*Quick question* 8.3.8. Check that $f_\bullet : M_\bullet \to N_\bullet$ is a homotopy equivalence if and only if there exists a chain map $g_\bullet : N_\bullet \to M_\bullet$ such that $g_\bullet \circ f_\bullet \sim 1_{M_\bullet}$ and $f_\bullet \circ g_\bullet \sim 1_{N_\bullet}$.

**Theorem 8.3.9.** *Let $R$ be ring, and let $M$ be an arbitrary $R$-module. If $(P_\bullet, \alpha)$ and $(Q_\bullet, \beta)$ are projective resolutions of $M$, then $P_\bullet$ and $Q_\bullet$ are homotopy equivalent.*

*Proof.* By Proposition 8.3.6, there exist $f_\bullet : P_\bullet \to Q_\bullet$ and $g_\bullet : Q_\bullet \to P_\bullet$ such that $1_M \circ \alpha = \beta \circ f_0$ and $1_M \circ \beta = \alpha \circ g_0$. Now $g_\bullet \circ f_\bullet : P_\bullet \to P_\bullet$ is a chain map such that $1_M \circ \alpha = \alpha \circ (g_0 \circ f_0)$. Since $1_{P_\bullet} : P_\bullet \to P_\bullet$ is another chain map with the same property ($1_M \circ \alpha = \alpha \circ 1_{P_0}$), we conclude from the second part of Proposition 8.3.6 that $g_\bullet \circ f_\bullet \sim 1_{P_\bullet}$. The same arguments show that $f_\bullet \circ g_\bullet \sim 1_{Q_\bullet}$. □

**Corollary 8.3.10.** *There exists a* projective resolution functor $P_\bullet(-) : {}_R\mathsf{Mod} \to {}_R\mathsf{HoCh}$*, which is defined as follows. On objects, $P_\bullet(-)$ sends an $R$-module $M$ to a projective resolution $P_\bullet(M)$ of $M$. On morphisms, $P_\bullet(-)$ sends a homomorphism of $R$-modules $f : M \to N$ to (the equivalence class of) a chain map $g_\bullet : P_\bullet(M) \to P_\bullet(N)$ which makes the following diagram commute:*

$$\begin{tikzcd}
\cdots \arrow[r] & P_2(M) \arrow[r] \arrow[d, "g_2"] & P_1(M) \arrow[r] \arrow[d, "g_1"] & P_0(M) \arrow[r, "\alpha_M"] \arrow[d, "g_0"] & M \arrow[r] \arrow[d, "f"] & 0 \\
\cdots \arrow[r] & P_2(N) \arrow[r] & P_1(N) \arrow[r] & P_0(N) \arrow[r, "\alpha_N"'] & N \arrow[r] & 0
\end{tikzcd}$$

*Proof.* We leave the proof to the reader. The point is that given $M$, there are many potential choices for $(P_\bullet(M), \alpha_M)$, but once such a choice has been made for every $R$-module $M$, then all other ambiguities essentially disappear because of Proposition 8.3.6 and Theorem 8.3.9. □

Of course these results admit a "dual version", which we summarise now, leaving all proofs for the reader. Let $R$ be a ring, and let $f : M \to N$ be a homomorphism of $R$-modules. If $(I^\bullet, \alpha)$ is an arbitrary coresolution of $M$ and if $(J^\bullet, \beta)$ is an *injective* coresolution of $N$, then there exists a cochain map $g^\bullet : I^\bullet \to J^\bullet$ such that $g^0 \circ \alpha = \beta \circ f$ – in other words, such that the following diagram commutes:

$$\begin{tikzcd}
\cdots & I^2 \arrow[l] \arrow[d, "g^2"] & I^1 \arrow[l] \arrow[d, "g^1"] & I^0 \arrow[l] \arrow[d, "g^0"] & M \arrow[l, "\alpha"'] \arrow[d, "f"] & 0 \arrow[l] \\
\cdots & J^2 \arrow[l] & J^1 \arrow[l] & J^0 \arrow[l] & N \arrow[l, "\beta"] & 0 \arrow[l]
\end{tikzcd}$$

Moreover $g^\bullet$ is unique in ${}_R\mathsf{HoCoCh}$: if $g^\bullet$ and $g^{\bullet\prime}$ are cochain maps which make this diagram commute, there exist $R$-module homomorphisms $(h^i : I^{i+1} \to J^i)_{i \in \mathbf{Z}_{\geq 0}}$ which define a cochain homotopy from $g^\bullet$ to $g^{\bullet\prime}$. This implies in particular that any two injective coresolutions of a given $R$-module are homotopy equivalent, and that there exists an *injective coresolution functor* $I^\bullet(-) : {}_R\mathsf{Mod} \to {}_R\mathsf{HoCoCh}$.

## 8.4 Derived functors: the example of Ext

In this final section, we will cover an example of *derived functors* defined on categories of modules, namely the Ext-functors (derived from the left exact Hom-functor). While it is certainly possible to construct derived functors in a much more general setting – namely for arbitrary abelian categories rather than for categories of $R$-modules, and for arbitrary left/right exact functors rather than for Hom – the example of Ext covers all of the main ideas already. Let us start right away with the definition.

**Definition 8.4.1.** Let $R$ be a ring, and let $N$ be an $R$-module. The functor $\mathrm{Ext}^n_R(-, N)$ is the composition

$${}_R\mathsf{Mod} \xrightarrow{P_\bullet(-)} {}_R\mathsf{HoCh} \xrightarrow{\mathrm{Hom}_R(-,N)} {}_R\mathsf{HoCoCh} \xrightarrow{H^n(-)} {}_R\mathsf{Mod}.$$

We already know $P_\bullet(-) : {}_R\mathsf{Mod} \to {}_R\mathsf{HoCh}$ and $H^n(-) : {}_R\mathsf{HoCoCh} \to {}_R\mathsf{Mod}$, but how is $\mathrm{Hom}_R(-, N)$ a functor from ${}_R\mathsf{HoCh}$ to ${}_R\mathsf{HoCoCh}$? We first note that $\mathrm{Hom}_R(-, N) : {}_R\mathsf{Mod} \to {}_R\mathsf{Mod}$ induces a functor $\mathrm{Hom}_R(-, N) : {}_R\mathsf{Ch} \to {}_R\mathsf{CoCh}$, defined on objects by sending a chain complex

$$\cdots \longrightarrow M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \longrightarrow \cdots$$

to the cochain (!) complex with $\mathrm{Hom}_R(M_i, N)$ in degree $i$:

$$\cdots \longrightarrow \mathrm{Hom}_R(M_{i-1}, N) \xrightarrow{d_i^\star} \mathrm{Hom}_R(M_i, N) \xrightarrow{d_{i+1}^\star} \mathrm{Hom}_R(M_{i+1}, N) \longrightarrow \cdots .$$

Next, we observe that if two chain maps $f_\bullet, g_\bullet : M_\bullet \to M'_\bullet$ are chain homotopic, then so are the induced cochain maps $\mathrm{Hom}_R(f_\bullet, N)$ and $\mathrm{Hom}_R(g_\bullet, N)$. Indeed, if $(h_i : M_i \to M'_{i+1})_{i \in \mathbf{Z}_{\geq 0}}$ defines a homotopy from $f_\bullet$ to $g_\bullet$, as in the diagram

$$\begin{tikzcd}
\cdots \arrow[r] & M_{i+1} \arrow[r, "d^M_{i+1}"] \arrow[d, shift right, "f_{i+1}"'] \arrow[d, shift left, "g_{i+1}"] & M_i \arrow[r, "d^M_i"] \arrow[dl, "h_i"] \arrow[d, shift right, "f_i"'] \arrow[d, shift left, "g_i"] & M_{i-1} \arrow[r] \arrow[dl, "h_{i-1}"] \arrow[d, shift right, "f_{i-1}"'] \arrow[d, shift left, "g_{i-1}"] & \cdots \\
\cdots \arrow[r] & M'_{i+1} \arrow[r, "d^{M'}_{i+1}"'] & M'_i \arrow[r, "d^{M'}_i"'] & M'_{i-1} \arrow[r] & \cdots
\end{tikzcd},$$

then the reader can (and should!) check that the maps $(h_i^\star : \mathrm{Hom}_R(M'_{i+1}, N) \to \mathrm{Hom}_R(M_i, N))_{i \in \mathbf{Z}_{\geq 0}}$ yield a homotopy from $\mathrm{Hom}_R(f_\bullet, N)$ to $\mathrm{Hom}_R(g_\bullet, N)$, as in the diagram

$$\begin{tikzcd}
\cdots & \mathrm{Hom}_R(M_{i+1}, N) \arrow[l] & \mathrm{Hom}_R(M_i, N) \arrow[l, "(d^M_{i+1})^\star"'] & \mathrm{Hom}_R(M_{i-1}, N) \arrow[l, "(d^M_i)^\star"'] & \cdots \arrow[l] \\
\cdots & \mathrm{Hom}_R(M'_{i+1}, N) \arrow[l] \arrow[u, shift left, "g^\star_{i+1}"] \arrow[u, shift right, "f^\star_{i+1}"'] \arrow[ur, "h_i^\star"] & \mathrm{Hom}_R(M'_i, N) \arrow[l, "(d^{M'}_{i+1})^\star"] \arrow[u, shift left, "g^\star_i"] \arrow[u, shift right, "f^\star_i"'] \arrow[ur, "h_{i-1}^\star"] & \mathrm{Hom}_R(M'_{i-1}, N) \arrow[l, "(d^{M'}_i)^\star"] \arrow[u, shift left, "g^\star_{i-1}"] \arrow[u, shift right, "f^\star_{i-1}"'] & \cdots \arrow[l]
\end{tikzcd}.$$

Since $\mathrm{Hom}_R(-, N) : {}_R\mathsf{Ch} \to {}_R\mathsf{CoCh}$ sends homotopic chain maps to homotopic cochain maps, we get an induced functor $\mathrm{Hom}_R(-, N) : {}_R\mathsf{HoCh} \to {}_R\mathsf{HoCoCh}$ on homotopy categories.

Let us go back to the definition of $\mathrm{Ext}^n_R(-, N) : {}_R\mathsf{Mod}^{\mathrm{opp}} \to {}_R\mathsf{Mod}$ as being the composition

$${}_R\mathsf{Mod} \xrightarrow{P_\bullet(-)} {}_R\mathsf{HoCh} \xrightarrow{\mathrm{Hom}_R(-,N)} {}_R\mathsf{HoCoCh} \xrightarrow{H^n(-)} {}_R\mathsf{Mod}$$

and unravel this definition further. First of all, it is not hard to see that $\mathrm{Ext}^n_R(-, N)$ is *contravariant* – we have omitted the superscripts opp above to simplify notation. Indeed, $P_\bullet(-)$ and $H^n(-)$ are covariant, but $\mathrm{Hom}_R(-, N)$ is contravariant; hence the composition of these functors is again contravariant.

The definition tells us that we should proceed as follows to compute $\mathrm{Ext}^n_R(M, N) := \mathrm{Ext}n_R(-, N)(M)$:

- choose a projective resolution $\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \xrightarrow{\alpha} M \longrightarrow 0$ of $M$,
- apply $\mathrm{Hom}_R(-, N)$ to the complex $\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \longrightarrow 0$ (with $M$ deleted!) to obtain
$$0 \longrightarrow \mathrm{Hom}_R(P_0, N) \longrightarrow \mathrm{Hom}_R(P_1, N) \longrightarrow \mathrm{Hom}_R(P_2, N) \longrightarrow \cdots,$$
- take cohomology in degree $n$ to obtain
$$\mathrm{Ext}^n_R(M, N) = \frac{\ker(\mathrm{Hom}_R(P_n, N) \longrightarrow \mathrm{Hom}_R(P_{n+1}, N))}{\mathrm{im}(\mathrm{Hom}_R(P_{n-1}, N) \longrightarrow \mathrm{Hom}_R(P_n, N))}.$$

At this point it is not clear why the functors $\mathrm{Ext}^n_R(-, N)$ should be useful or interesting. The fundamental reason for studying these functors is that they are "derivatives" or "offspring" of the left exact functor in $\mathrm{Hom}_R(-, N)$, in the sense that they measure the extent to which $\mathrm{Hom}_R(-, N)$ is not exact. The name Ext is an abbreviation for *extension* (but secretly also for *extremely cool functor*); we will see that $\mathrm{Ext}^1_R(-, N)$ in particular has lots of things to say about extensions of modules.

**Lemma 8.4.2.** *Let $R$ be a ring and let $N$ be an arbitrary $R$-module. Then the functors* $\mathrm{Hom}_R(-, N)$ *and* $\mathrm{Ext}^0_R(-, N)$ *are naturally isomorphic.*

*Proof.* Let $M$ be an $R$-module, and let $\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \xrightarrow{\alpha} M \longrightarrow 0$ be a projective resolution of $M$. Since $\mathrm{Hom}_R(-,N)$ is left exact, we certainly get an exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(M,N) \xrightarrow{\alpha^\star} \mathrm{Hom}_R(P_0,N) \longrightarrow \mathrm{Hom}_R(P_1,N).$$

On the other hand, $\mathrm{Ext}^n_R(M,N)$ can be computed as the degree $n$ cohomology of the cochain complex

$$0 \longrightarrow \mathrm{Hom}_R(P_0,N) \longrightarrow \mathrm{Hom}_R(P_1,N) \longrightarrow \mathrm{Hom}_R(P_2,N) \longrightarrow \cdots$$

In particular $\mathrm{Ext}^0_R(M,N)$ is the kernel of $\mathrm{Hom}_R(P_0,N) \longrightarrow \mathrm{Hom}_R(P_1,N)$; using the previous exact sequence, we see that this kernel can be identified to $\mathrm{Hom}_R(M,N)$ (via the map $\alpha^\star$). Moreover this identification is easily checked to be functorial in $M$, which finishes the proof. □

Hence $\mathrm{Ext}^0_R(-,N)$ is really the same things as $\mathrm{Hom}_R(-,N)$; the functors $\mathrm{Ext}^1_R(-,N)$, $\mathrm{Ext}^2_R(-,N)$, et cetera should be thought as belonging to successive generations of offspring of $\mathrm{Hom}_R(-,N)$. To get a better feeling for what these functors "mean", we will study a few examples first.

**Example 8.4.3.** Let $R$ be a ring, and let $P$ be a projective $R$-module. Then for an arbitrary $R$-module $N$, we have $\mathrm{Ext}^n_R(P,N) = 0$ whenever $n \geq 1$. Indeed, this follows from the fact that there is a trivial projective resolution of $P$, namely $\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow P \xrightarrow{\mathrm{Id}} P \longrightarrow 0$.

**Example 8.4.4.** Let $R$ be a ring, and let $c \in R$ be an element which is not a zero divisor. Our goal is to compute $\mathrm{Ext}^n_R(R/(c),N)$, for an arbitrary choice of the $R$-module $N$. We then have the following free (a fortiori projective) resolution for the $R$-module $R/(c)$:

$$\cdots \longrightarrow 0 \longrightarrow R \xrightarrow{\cdot c} R \longrightarrow R/(c) \longrightarrow 0.$$

Hence $\mathrm{Ext}^n_R(R/(c),N)$ is the degree $n$ cohomology of the cochain complex

$$0 \longrightarrow \mathrm{Hom}_R(R,N) \xrightarrow{(\cdot c)^\star} \mathrm{Hom}_R(R,N) \longrightarrow \mathrm{Hom}_R(0,N) \longrightarrow \cdots$$

which can be identified with the cochain complex

$$0 \longrightarrow N \xrightarrow{\cdot c} N \longrightarrow 0 \longrightarrow \cdots,$$

non-zero terms in degrees 0 and 1. Hence $\mathrm{Ext}^0_R(R/(c),N) = \{y \in N : cy = 0\} \cong \mathrm{Hom}_R(R/(c),N)$ (why?), $\mathrm{Ext}^1_R(R/(c),N) \cong N/cN$ and $\mathrm{Ext}^n_R(R/(c),N) = 0$ whenever $n \geq 2$.

To go a bit further, we need the following result:

**Lemma 8.4.5.** *Let $R$ be a ring, and let $(M_i)_{i\in I}$ and $N$ be $R$-modules. Then*

$$\mathrm{Ext}^n_R\left(\bigoplus_{i\in I} M_i, N\right) \cong \prod_{i\in I} \mathrm{Ext}^n_R(M_i,N).$$

*Proof.* We sketch the proof, and leave the details for the reader. If $(\mathrm{P}_{\bullet,i},\alpha_i)$ is a projective resolution of $M_i$, then $\left(\bigoplus_{i\in I} P_{\bullet,i}, \bigoplus_{i\in I}\alpha_i\right)$ is a projective resolution of $\bigoplus_{i\in I} M_i$ (make this precise!). Now $\mathrm{Ext}^n_R\left(\bigoplus_{i\in I} M_i, N\right)$ is the degree $n$ cohomology of the cochain complex

$$0 \longrightarrow \mathrm{Hom}_R\left(\bigoplus_{i\in I} P_{0,i}, N\right) \longrightarrow \mathrm{Hom}_R\left(\bigoplus_{i\in I} P_{1,i}, N\right) \longrightarrow \mathrm{Hom}_R\left(\bigoplus_{i\in I} P_{2,i}, N\right) \longrightarrow \cdots$$

which can be identified with

$$0 \longrightarrow \prod_{i\in I} \mathrm{Hom}_R(P_{0,i}, N) \longrightarrow \prod_{i\in I} \mathrm{Hom}_R(P_{1,i}, N) \longrightarrow \prod_{i\in I} \mathrm{Hom}_R(P_{2,i}, N) \longrightarrow \cdots$$

Since cohomology commutes with direct products (why?), the result follows. □

We are now able to prove the following basic result:

**Proposition 8.4.6.** *Let $R$ be a PID. If $M$ and $N$ are $R$-modules, if $M$ is finitely generated, and if $n \geq 2$, then*

$$\mathrm{Ext}^n_R(M, N) = 0.$$

*Proof.* This follows from Theorem 4.1.5, together with Lemma 8.4.5 and Example 8.4.4. □

Let us now introduce the *long exact sequence* involving the various $\mathrm{Ext}^n$.

**Theorem 8.4.7.** *Let $R$ be a ring. Let $N$ be an $R$-module, and let*

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

*be a short exact sequence of $R$-modules. Then we have a long exact sequence of the form*

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Hom}_R(M'', N) \arrow[r, "g^\star"] & \mathrm{Hom}_R(M, N) \arrow[r, "f^\star"] & \mathrm{Hom}_R(M', N) \arrow[dll, rounded corners, "\partial"] \\
& \mathrm{Ext}^1_R(M'', N) \arrow[r, "g^\star"] & \mathrm{Ext}^1_R(M, N) \arrow[r, "f^\star"] & \mathrm{Ext}^1_R(M', N) \arrow[dll, rounded corners, "\partial"] \\
& \mathrm{Ext}^2_R(M'', N) \arrow[r, "g^\star"] & \mathrm{Ext}^2_R(M, N) \arrow[r, "f^\star"] & \mathrm{Ext}^2_R(M', N) \arrow[dll, rounded corners, "\partial"] \\
& \cdots & &
\end{tikzcd}$$

*Proof.* Let $(P'_\bullet, \alpha')$ and $(P''_\bullet, \alpha'')$ be projective resolutions of $M'$ and $M''$ respectively. Set $P_n = P'_n \oplus P''_n$ for all $n \geq 0$; let $\iota_n : P'_n \to P_n$ be the inclusion, and let $\pi_n : P_n \to P''_n$ be the projection.

We claim that the diagram

$$\begin{tikzcd}
& 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\cdots \arrow[r] & P'_2 \arrow[r] \arrow[d, "\iota_2"] & P'_1 \arrow[r] \arrow[d, "\iota_1"] & P'_0 \arrow[r, "\alpha'"] \arrow[d, "\iota_0"] & M' \arrow[r] \arrow[d, "f"] & 0 \\
\cdots \arrow[r, dashed] & P_2 \arrow[r, dashed, "d_2"] \arrow[d, "\pi_2"] & P_1 \arrow[r, dashed, "d_1"] \arrow[d, "\pi_1"] & P_0 \arrow[r, dashed, "\alpha"] \arrow[d, "\pi_0"] & M \arrow[r] \arrow[d, "g"] & 0 \\
\cdots \arrow[r] & P''_2 \arrow[r] \arrow[d] & P''_1 \arrow[r] \arrow[d] & P''_0 \arrow[r, "\alpha''"] \arrow[d] & M'' \arrow[r] \arrow[d] & 0 \\
& 0 & 0 & 0 & 0 &
\end{tikzcd}$$

can be completed using the dashed red arrows in such a way that the middle row is a resolution of $M$. This is the so-called *horseshoe lemma*; here we sketch the construction of $\alpha$, for further details we refer to [Rotman, Lemma 10.53]. To construct $\alpha$, note that since $P''_0$ is projective, the map $\alpha'' : P''_0 \to M''$ can be lifted to some $\widetilde{\alpha''} : P''_0 \to M$. We now choose $\alpha : P_0 = P'_0 \oplus P''_0 \to M : (x, y) \mapsto (f \circ \alpha')(x) + \widetilde{\alpha''}(y)$; this is easily seen to be surjective. A similar method allows us to construct $d_1, d_2, \cdots$.

We now apply $\mathrm{Hom}_R(-, N)$ to the split exact sequence of chain complexes

$$0 \longrightarrow P'_\bullet \longrightarrow P_\bullet \longrightarrow P''_\bullet \longrightarrow 0$$

to get a complex of cochain complexes

$$0 \longrightarrow \mathrm{Hom}_R(P''_\bullet, N) \xrightarrow{\pi^\star_\bullet} \mathrm{Hom}_R(P_\bullet, N) \xrightarrow{\iota^\star_\bullet} \mathrm{Hom}_R(P'_\bullet, N) \longrightarrow 0$$

which is still exact (check! – this is a special case of part (b) of Exercise 7.3).

The result now follows immediately from the long exact sequence in cohomology associated to this exact sequence (see Theorem 8.2.9), together with the definition of $\mathrm{Ext}^n_R$. □

*Remark* 8.4.8. The long exact sequence from Theorem 8.4.7 shows that $\mathrm{Ext}^1_R(-, N)$ measures in a way how far $\mathrm{Hom}_R(-, N)$ is away from being exact, i.e., how far $N$ is away from being injective. An alternative interpretation (which we will not prove) is the following: $\mathrm{Ext}^1_R(M, N) = 0$ if and only if every short exact sequence of the form $0 \to N \to E \to M \to 0$ is split.

# Exercises

*Easy exercise* 8.1. Let $R$ be a non-zero ring and let $S = R \times R$. Show that $I = \{(0, r) : r \in R\}$ is a projective $S$-module which is not free.

*Easy exercise* 8.2. Let $R$ be a ring, and let $P_1$ and $P_2$ be projective $R$-modules. Show that $P_1 \otimes_R P_2$ is a projective $R$-module. Does the result remain true when the adjective "projective" is replaced by "flat"?

*Easy exercise* 8.3. Show that the $\mathbf{Z}/4$-module $\mathbf{Z}/2$ does not have a free resolution of finite length.

*Easy exercise* 8.4. Compute $\mathrm{Ext}^n_{\mathbf{Z}/4}(\mathbf{Z}/2, \mathbf{Z}/2)$ for all $n \geq 0$, via the free resolution from Remark 8.3.4.

*Exercise* 8.5. Let $R = \mathbf{Z}[\sqrt{-5}]$ (see Example 1.3.13), $I = (2, 1 + \sqrt{-5})$ and $J = (3, 1 - \sqrt{-5})$. Show that $I \cong J$ as $R$-modules, that $I + J = R$ and that $IJ = (1 - \sqrt{-5})$. Next, use Exercise 2.14 to show that $I \oplus J \cong R^2$ as $R$-modules. Conclude that $I$ and $J$ are projective $R$-modules which are not free.

*Exercise* 8.6. Let $R = \mathbf{Q}[X, Y, Z]/(X^2 + Y^2 + Z^2 - 1)$ and let $\varphi : R^3 \to R$ be the homomorphism of $R$-modules defined by the equalities $\varphi((1, 0, 0)) = \overline{X}$, $\varphi((0, 1, 0)) = \overline{Y}$ and $\varphi((0, 0, 1)) = \overline{Z}$. Show that $\ker \varphi$ is a projective $R$-module.

*Exercise* 8.7. Let $R$ be a PID. If $M$ is finitely generated over $R$, show that the following are equivalent: $M$ is free; $M$ is projective; $M$ is flat. Does this remain true without the finite generation assumption?

*Exercise* 8.8. Let $R$ be an integral domain. Show that if $M$ is an injective $R$-module, then $M$ is *divisible*: this means that for every $r \in R$, the $R$-module homomorphism $M \xrightarrow{\times r} M$ is surjective. Deduce from this that if $R$ is not a field and $M$ is both injective and projective, then $M$ must be the zero module.

*Exercise* 8.9. Let $R$ and $S$ be rings. Assume that the functor $F : {}_R\mathsf{Mod} \to {}_S\mathsf{Mod}$ is left adjoint to $G : {}_S\mathsf{Mod} \to {}_R\mathsf{Mod}$. If $F$ is exact, show that $G$ sends injective $S$-modules to injective $R$-modules. If $G$ is exact, show that $F$ sends projective $R$-modules to projective $S$-modules.

*Exercise* 8.10. Let $R$ be a ring. Consider the following diagram of $R$-modules, with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & K_1 \arrow[r, "i"] \arrow[d, dashed, "\alpha"] & P_1 \arrow[r, "p"] \arrow[d, dashed, "\beta"] & B \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & K_2 \arrow[r, "j"] & P_2 \arrow[r, "q"] & B \arrow[r] & 0
\end{tikzcd}$$

Assume that both $P_1$ and $P_2$ are projective $R$-modules.

(a) Show that there exists an $R$-module homomorphism $\beta : P_1 \to P_2$ such that $p = q \circ \beta$.

(b) Show that there exists an $R$-module homomorphism $\alpha : K_1 \to K_2$ such that $\beta \circ i = j \circ \alpha$.

(c) Define $\psi : P_1 \oplus K_2 \to P_2$ by $\psi(x, k) = \beta(x) - j(k)$. Show that $\ker \psi \cong K_1$.

(d) Conclude that $P_1 \oplus K_2 \cong P_2 \oplus K_1$: this is *Schanuel's lemma.*

State and prove the dual version of this result as well.

*Exercise* 8.11. Let $R$ be a non-zero ring and let $x, y \in R$. Consider the following objects in ${}_R\mathsf{CoCh}$:

– $K(x)$, non-zero in degrees 0 and 1 only, given by

$$0 \to R \xrightarrow{\varphi} R \to 0, \text{ where } \varphi : R \to R : r \mapsto xr;$$

– $K(x)[1]$, non-zero in degrees 1 and 2 only, obtained by "shifting $K(x)$ to the right by one degree";

– $K(x, y)$, non-zero in degrees 0, 1 and 2 only, given by

$$0 \to R \xrightarrow{\chi} R \oplus R \xrightarrow{\psi} R \to 0, \quad \text{where } \begin{cases} \chi : R \to R \oplus R : r \mapsto (yr, -xr), \\ \psi : R \oplus R \to R : (r, s) \mapsto xr + ys \end{cases}$$

Construct a short exact sequence in ${}_R\mathsf{CoCh}$ of the form $0 \to K(x)[1] \to K(x, y) \to K(x) \to 0$ and compute the associated long exact cohomology sequence explicitly (including the boundary maps).

*Exercise* 8.12. Let $R = \mathbf{C}[X, Y]/(X^2 - Y^3)$, and let $\mathfrak{m}$ be the maximal ideal of $R$ generated by the images of $X$ and $Y$. Write down an explicit free resolution of the $R$-module $R/\mathfrak{m}$.

*Exercise* 8.13. Let $R = \mathbf{C}[X, Y]$, and let $\mathfrak{m} = (X, Y)$. Using appropriate resolutions, compute the $R$-modules $\mathrm{Ext}^n_R(R/\mathfrak{m}, R/\mathfrak{m})$ and $\mathrm{Ext}^n_R(R/(X), R/(Y))$ for all integers $n \geq 0$.

*Exercise* 8.14. Let $n \geq 2$ be an integer, let $R = \mathbf{Z}[T]/(T^n - 1)$ and let $t$ be the image of $T$ in $R$. Let $M$ be an $R$-module and consider $\mathbf{Z}$ as an $R$-module on which $t$ acts trivially, in other words: scalar multiplication satisfies $t \cdot m = m$ for all $m \in \mathbf{Z}$. Let $N = 1 + t + t^2 + \cdots + t^{n-1}$.

Show that we have a projective resolution for $\mathbf{Z}$ given by

$$\cdots \longrightarrow R \longrightarrow R \longrightarrow R \longrightarrow R \longrightarrow \mathbb{Z} \to 0,$$

where the differentials are alternately multiplication by $t - 1$ and by $N$. Compute $\mathrm{Ext}^n_R(\mathbf{Z}, M)$ for all $n$.

*Exercise* 8.15. Let $R$ be a ring, and let $n \geq 1$ be an integer. Assume that $N$ is an $R$-module such that $\mathrm{Ext}^n_R(-, N)$ vanishes, in other words: $\mathrm{Ext}^n_R(M, N) = 0$ for all $R$-modules $M$.

(a) Show that $\mathrm{Ext}^{n+1}_R(-, N)$ vanishes as well.

(b) Show that $\mathrm{Ext}^{n-1}_R(-, N)$ is a right exact functor.

*Hard exercise* 8.16. Let $R$ be a ring, and let $M$ and $N$ be $R$-modules. Assume that $r \in R$ is not a zero divisor on $N$ (in other words: the $R$-module homomorphism $N \to N : n \mapsto rn$ is injective) and $r$ annihilates $M$. Show that $\mathrm{Hom}_R(M, N) = 0$ and $\mathrm{Ext}^1_R(M, N) \cong \mathrm{Hom}_R(M, N/rN)$.