Advanced Algebra

Alex Torzewski
King's College London

Semester 1, 2024–25, version of September 22, 2024

Notes by Fred Diamond, revised by Alex Torzewski

Contents

Part 1

Group Theory

We will start with a review of basic concepts and results in group theory, then go deeper into topics including group actions and free groups, introducing some of the language of category theory along the way.

1 Basic definitions, properties and examples

DEFINITION 1.1 A *group* $(G,*)$ is a pair consisting of a set G together with a a binary operation $*$ on G, i.e. a function

$$\begin{array}{ccc} G\times G & \to & G \\ (g,h) & \mapsto & g*h, \end{array}$$

such that the binary operation satisfies:

(i) $*$ is associative, i.e. $g*(h*k)=(g*h)*k$ for all $g,h,k\in G$;

(ii) $*$ has an identity element, i.e. there is an element $e\in G$ such that $e*g=g*e=g$ for all $g\in G$;

(iii) every element $g\in G$ has an inverse with respect to $*$, i.e. an element $h\in G$ such that $g*h=h*g=e$.

When the binary operation is clear from the context, we often denote a group $(G,*)$ simply by its underlying set G. The binary operation is also referred to variously as "product", "multiplication" or "composition".

EXAMPLE 1.2 The most familiar groups are the sets $\mathbb{Z}$ (the integers), $\mathbb{Q}$ (the rational numbers), $\mathbb{R}$ (the real numbers) and $\mathbb{C}$ (the complex numbers) with the usual addition operation $+$.

EXAMPLE 1.3 The sets of non-zero rational numbers $\mathbb{Q}^\times$, real numbers $\mathbb{R}^\times$ and complex numbers $\mathbb{C}^\times$ are groups under the usual multiplication operation $\times$; the set of non-zero integers[1] is *not* a group under multiplication (for lack of inverses).

EXAMPLE 1.4 The set $\mathrm{GL}_n(\mathbb{R})$ of invertible real $n\times n$ matrices (for any positive integer n) is a group under multiplication, as is $\mathrm{SL}_n(\mathbb{R})$ (the set of such matrices of determinant 1). Similarly we have other matrix groups, such as $\mathrm{SL}_n(\mathbb{Z})$ or $\mathrm{GL}_n(\mathbb{C})$.

EXAMPLE 1.5 For any set X let S_X denote the set of permutations of X:

$$S_X=\{f:X\to X \mid f \text{ is bijective}\}.$$

Then composition of functions $\circ$ defines a binary operation on S_X satisfying properties (i),(ii), (iii) of Definition 1.1, and we call S_X the *group of permutations of* X (sometimes also denoted Σ_X). Note we do not stipulate that X is finite! In the case when $X=\{1,2,\dots,n\}$, we denote $S_{\{1,2,\dots,n\}}$ by S_n.

EXERCISE 1.6 Let $(G,*)$ be a group. Prove that the identity element asserted to exist in (ii) of Definition 1.1 is unique. Prove that for every $g\in G$, the inverse element required to exist in (iii) of Definition 1.1 is also unique. In other words, the identity and inverses are not additional data over $(G,*)$. We denote the identity by e (or occasionally 1) and the inverse of g by g^{-1}.

EXAMPLE 1.7 We let D_n denote the dihedral group of order[2] $2n$, i.e. the group of symmetries of a regular n-gon (for $n\geq 3$).

[1] Note that this set is *not* denoted $\mathbb{Z}^\times$. The more general meaning of the notation $^\times$ notation will be recalled later, giving $\mathbb{Z}^\times=\{\pm 1\}$, which is a group under multiplication.

[2] Note there are different conventions: this is often denoted D_{2n}.

EXAMPLE 1.8 If G and H are groups, with operations $*_G$ and $*_H$, then[3] $G \times H$ is a group under the operation defined by

$$(g,h) * (g',h') = (g *_G g', h *_H h').$$

DEFINITION 1.9 We say a group G (with operation $*$) is *abelian* if $*$ is commutative, i.e. $g * h = h * g$ for all $g,h \in G$. For abelian groups it is common to refer to the binary operation as "addition" and to use the symbol $+$.

So Examples 1.2 and 1.3 are abelian, but Example 1.4 is not (if $n > 1$), nor are 1.5 and 1.7 (if $n > 2$). As for Example 1.8, $G \times H$ is abelian if and only if G and H are both abelian.

REMARK 1.10 Note that associativity ensures that there is no ambiguity arising from writing arbitrary product

$$g_1 * g_2 * \ldots * g_n$$

without brackets. For example, $g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3$.

We often omit the symbol for the operation entirely and simply write gh instead of $g * h$. Similarly, $g_1 * (g_2 * g_3)$ becomes $g_1(g_2 g_3)$, which by the previous remark is equal to $(g_1 g_2)g_3$ and so we can and shall write unambiguously as $g_1 g_2 g_3$ (and likewise for products of arbitrarily many elements).

Recall that g^{-1} denotes the inverse of g. We combine this and the previous notation by setting

$$g^n = \begin{cases} e & n = 0 \\ \underbrace{g * g * \cdots * g}_{n \text{ times}} & n > 0 \\ (g^{-n})^{-1} & n < 0 \end{cases}.$$

In the case of abelian groups, when the binary operation is denoted $+$, we instead define the identity by 0, the inverse by $-g$ and powers by ng.

EXERCISE 1.11 Prove that if G is a group then the following "Cancellation Law" holds: for any $g,h,k \in G$, if $gh = gk$, then $h = k$. Verify that when $n < 0$, then g^n is equal to $(g^{-1})^{-n}$. More generally, check that $g^a g^b = g^{a+b}$ for any $a, b \in \mathbb{Z}$.

EXERCISE 1.12 Let G be a group. Show that $g^n h^n = (gh)^n$ for all $g,h \in G$ and $n \in \mathbb{Z}$ if and only if G is abelian. On the other hand, show that $(gh)^{-1} = h^{-1}g^{-1}$ always.

DEFINITION 1.13 If G is a group and $H \subset G$, then H is a *subgroup* of G if the following[4] hold:

(i) if $h,h' \in H$, then $hh' \in H$;

(ii) $e \in H$ (where e is the identity element of G);

(iii) if $h \in H$, then $h^{-1} \in H$.

It is common to use the notation $H \leq G$ to denote subgroups.

If H is a subgroup of G, then H is itself a group, with the same binary operation restricted to H, i.e. the function $G \times G \to G$ defined by $(g,g') \mapsto gg'$ restricts (by (i)) to define a binary operation $H \times H \to H$, under which H is a group.

[3] Recall that if X and Y are any sets, then $X \times Y$ is the set of ordered pairs (x,y), where $x \in X$ and $y \in Y$.

[4] A standard exercise shows that if H is not empty, then the three properties together are equivalent to the single condition that $h^{-1}h' \in H$ for all $h,h' \in H$.

EXAMPLE 1.14 We have that $\mathbb{Z}$ is a subgroup of $\mathbb{Q}$, which is a subgroup of $\mathbb{R}$, which is a subgroup of $\mathbb{C}$ (all under addition). Note that if H is a subgroup of G, and G itself is a subgroup of a group K, then H is a subgroup of K.

EXAMPLE 1.15 We have the chain of subgroups

$$\mathrm{SL}_n(\mathbb{Z}) \subset \mathrm{SL}_n(\mathbb{R}) \subset \mathrm{GL}_n(\mathbb{R}) \subset \mathrm{GL}_n(\mathbb{C}).$$

Another example is the chain of subgroups

$$Z \subset D \subset B \subset \mathrm{GL}_n(\mathbb{R}),$$

where

- Z is the set of scalar matrices in $\mathrm{GL}_n(\mathbb{R})$, i.e. matrices of the form rI_n, where $r \in \mathbb{R}^\times$ and I_n is the $n \times n$ identity matrix;
- D is the set of diagonal matrices in $\mathrm{GL}_n(\mathbb{R})$;
- and B is the set of upper-triangular matrices in $\mathrm{GL}_n(\mathbb{R})$.

DEFINITION 1.16 For any element g of a group G, define the *subgroup*[5] *of G generated by g*, denoted $\langle g \rangle$ to be $\{ g^n \mid n \in \mathbb{Z} \}$. We say the group G is *cyclic* if $G = \langle g \rangle$ for some $g \in G$.

EXAMPLE 1.17 Let $G = \mathbb{Z}$, and let $m \in \mathbb{Z}$. Then

$$\langle m \rangle = \{ nm \mid n \in \mathbb{Z} \}$$

is the set of integer multiples of m. (Recall that we're writing nm instead of m^n since the operation is $+$.) In particular $\langle 1 \rangle = \mathbb{Z}$, so $\mathbb{Z}$ is cyclic.

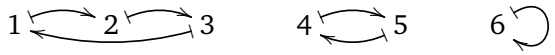
EXAMPLE 1.18 Let $G = \mathbb{Q}^\times$ and $r \in \mathbb{Q}^\times$. Then $\langle r \rangle = \{ r^n \mid n \in \mathbb{Z} \}$, so now for example $\langle 1 \rangle = \{1\}$.

EXERCISE 1.19 Show that $\mathbb{Q}^\times$ is not cyclic.

EXAMPLE 1.20 Let $G = S_6$. Recall that in "cycle notation," (123)(45) denotes the permutation σ such that

$$\sigma(1) = 2,\ \sigma(2) = 3,\ \sigma(3) = 1,\ \sigma(4) = 5,\ \sigma(5) = 4 \text{ and } \sigma(6) = 6,$$

which can be visualised as:



Iterating this gives $\sigma^2 = (132)$, $\sigma^3 = (45)$, $\sigma^4 = (123)$, $\sigma^5 = (132)(45)$ and $\sigma^6 = e$, $\sigma^7 = \sigma$, etc. Note also that $\sigma^{-1} = \sigma^5$, $\sigma^{-2} = \sigma^4$, etc., so

$$\langle \sigma \rangle = \{ e, \sigma, \sigma^2, \sigma^3, \sigma^4, \sigma^5 \}.$$

DEFINITION 1.21 For any subset S of a group G, the *subgroup of G generated by S* is

$$\langle S \rangle = \bigcap_{H \in \mathcal{A}} H,$$

where $\mathcal{A}$ is the set of subgroups H of G such that $S \subset H$. Then $\langle S \rangle$ is the smallest subgroup of G containing S (in a sense made precise in the exercises). So this does generalise Definition 1.16!

EXAMPLE 1.22 Let $G = \mathbb{Z}$ and $a, b \in \mathbb{Z}$, not both 0. Then

$$\langle \{a, b\} \rangle = \{ ma + nb \mid m, n \in \mathbb{Z} \} = \langle d \rangle,$$

where $d = \gcd(a, b)$. Note that $\langle g \rangle$ is shorthand for $\langle \{g\} \rangle$. More generally if $S = \{g_1, \ldots, g_k\}$ is finite, then we just write $\langle g_1, \ldots, g_k \rangle$ for $\langle S \rangle$, so for example $\langle 6, 9 \rangle = \langle 3 \rangle$ as subgroups of $\mathbb{Z}$.

[5] It is easy to check that it is indeed a subgroup.

EXAMPLE 1.23 Let $G = \mathbb{Q}^\times$ and $a, b \in \mathbb{Q}^\times$. Now

$$\langle\{a, b\}\rangle = \{\, a^m b^n \mid m, n \in \mathbb{Z}\,\},$$

so for example

$$\langle 6, 9\rangle = \{\, 6^m 9^n \mid m, n \in \mathbb{Z}\,\} = \{\, 2^m 3^k \mid m, k \in \mathbb{Z},\ m \equiv k \bmod 2\,\}.$$

(For the last equality, let $k = m + 2n$ for $\subset$, and $n = (k - m)/2$ for $\supset$.) Letting $\mathcal{P} =$ {prime numbers}, we have

$$\langle\mathcal{P}\rangle = \{\, r \in \mathbb{Q}^\times \mid r > 0\,\}$$

(as a consequence of the Fundamental Theorem of Arithmetic), and so

$$\mathbb{Q}^\times = \langle\mathcal{P} \cup \{-1\}\rangle.$$

EXAMPLE 1.24 If $G = S_3$, then $\langle(12)\rangle = \{e, (12)\}$ and similarly $\langle(13)\rangle = \{e, (13)\}$, but $\langle(12), (13)\rangle = G$.

DEFINITION 1.25 If G is a group, then its cardinality $\#G$ (or $|G|$) is called the *order* of G. Note that the order of G may be infinite (and we say G is *finite* if it has finite order). If $g \in G$, we define the *order* of g to be the order of $\langle g\rangle$.

EXAMPLE 1.26 The order of S_n is $n!$, so for example the order of S_6 is 720 and the order of the element $(123)(45)$ is 6 (see Example 1.20).

EXAMPLE 1.27 The order of $\mathbb{Q}^\times$ is infinite, but it has some elements of finite order, namely 1, which has order 1 (as does the identity element of any group), and -1, which has order 2.

2 Cosets and quotient groups

DEFINITION 2.1 Suppose that H is a subgroup of a group G. For each $g \in G$, the set

$$gH = \{\, gh \mid h \in H\,\}$$

is called a *left coset* of H in G (or where G is clear from the context simply a *left coset* of H). Similarly $Hg = \{\, hg \mid h \in H\,\}$ is called a *right coset* of H in G.

EXERCISE 2.2 Fix $g \in G$. Show that the map of sets $H \to G$ given by $h \mapsto gh$ is injective and has image gH. So left (and similarly right) cosets of H always have the same cardinality as H.

EXERCISE 2.3 Show that for $g, g' \in G$, we have

$$gH \cap g'H \neq \emptyset \iff g^{-1}g' \in H \iff gH = g'H.$$

In other words, left cosets are either disjoint or equal. Since every element of g is in some left coset of H in G, namely gH, the left cosets gH partition the group into disjoint subsets. Similar statements hold for right cosets.

We call any choice of $g' \in G$ for which $g'H = gH$ a choice of *coset representative* of the left coset gH.

EXAMPLE 2.4 Let $G = S_3$, and let $H = \langle(123)\rangle = \{e, (123), (132)\}$. Then $eH = (123)H = (132)H = H$ and

$$(12)H = \{(12), (23), (13)\} = (23)H = (13)H.$$

Similarly, the left cosets of $H' = \{e, (12)\}$ are

$$eH' = H',\ (13)H' = \{(13), (123)\} \text{ and } (23)H' = \{(23), (132)\}.$$

EXAMPLE 2.5 Let $G = \mathbb{Z}$ and let $H = \langle n \rangle$, where n is any positive integer. If $a \in \mathbb{Z}$, then the resulting left coset[6] of H in G is

$$a + H = \{ a + kn \mid k \in \mathbb{Z} \},$$

is the *residue class* of a modulo n, sometimes denoted $[a]_n$.

DEFINITION 2.6 We let G/H (said as "G mod H") denote the set of left cosets of H in G. The *index* of H in G, denoted $[G : H]$, is defined to be the cardinality of G/H (which may be infinite).

EXAMPLE 2.7 In Example 2.4, we have $G/H = \{H, (12)H\}$[7], so $[G : H] = 2$, and similarly since $G/H' = \{H', (13)H', (23)H'\}$, we have $[G : H'] = 3$.

EXAMPLE 2.8 In Example 2.5, the left cosets of $\langle n \rangle$ in $\mathbb{Z}$ are precisely the n distinct residue classes modulo n, i.e.

$$\mathbb{Z}/\langle n \rangle = \{ [0]_n, [1]_n, \ldots, [n-1]_n \}.$$

Therefore $[\mathbb{Z} : \langle n \rangle] = n$. In particular, an infinite group can have subgroups with finite index. It also has at least one subgroup of infinite index: the index of $\{e\}$ in G is the same as the order of G.

EXERCISE 2.9 A slightly different, but ultimately equivalent, way to introduce G/H is to consider it as the set of equivalence classes of G under the equivalence relation $g \sim g' \iff g^{-1}g' \in H$. Convince yourself of this!

If the group G is finite, then so are the order and index of every subgroup. Since each element of G is in exactly one left coset gH, and there are $[G : H]$ such left cosets, each having cardinality $\#H$, by Exercise 2.2, it follows that

$$\#G = [G : H] \cdot \#H.$$

This proves Lagrange's Theorem:

THEOREM 2.10 *Let G be a finite group. If H is a subgroup of G, then the order of H divides the order of G. Therefore if $g \in G$, then the order of g divides the order of G.*

DEFINITION 2.11 Suppose that H is a subgroup of a group G. We say H is a *normal* subgroup of G if $ghg^{-1} \in H$ for all $g \in G$, $h \in H$. Normal subgroups are often denoted by $H \trianglelefteq G$. The element ghg^{-1} of G is called the *conjugate* of h by g, and

$$gHg^{-1} = \{ ghg^{-1} \mid h \in H \}$$

is called the *conjugate* of H by g.

EXAMPLE 2.12 If G is abelian, then $ghg^{-1} = h \in H$ for all $g \in G$, $h \in H$, so every subgroup of G is normal.

EXERCISE 2.13 If $G = S_3$, show that $\langle (123) \rangle$ is normal, but $\langle (12) \rangle$ is not.

EXAMPLE 2.14 Consider the subgroups

$$Z \subset D \subset B \subset G = \mathrm{GL}_n(\mathbb{R})$$

defined in Example 1.15. Then Z is a normal subgroup of G, but D and B are not (if $n > 1$).

[6] The notation $g + H$ clashes slightly with Definition 2.1, but if we think of the notation in Definition 2.1 abbreviating $g * H$, then makes sense.

[7] Note sets don't allow duplicates. So $\{H, (123)H, (132)H\} = \{H\}$ and also $\{(123)H\}$ since these are the same!

REMARK 2.15 Suppose we have $K \leq H \leq G$, that is K is a subgroup of a group H and H is also a subgroup of G. Then it is easy to check that K is a subgroup of G. Now suppose that K is normal in H and H is normal in G. Then it is not necessarily true that K is normal as a subgroup of G! The condition that $gkg^{-1} \in K$ for all $k \in K$ and $g \in G$ is genuinely stronger than the condition that $hkh^{-1} \in K$ for all $k \in K$ and $h \in H$. "Being a subgroup is a transitive condition, being a normal subgroup is not". On the other hand, if we instead assume that K is a normal subgroup of G, then K is normal as a subgroup of H (a weaker condition).

EXERCISE 2.16 Find a chain of subgroups $K \leq H \leq D_4$, such that K is normal in H and H is normal in D_4, but K is not normal in D_4.

It is easy to check that if H is a subgroup of G, then gHg^{-1} is also a subgroup of G (for any $g \in G$). It is immediate from the definition that H is a normal subgroup of G if and only if $gHg^{-1} \subset H$ for all $g \in H$. In fact if $H \trianglelefteq G$, then we also have $g^{-1}Hg \subset H$, so

$$H = g(g^{-1}Hg)g^{-1} \subset gHg^{-1},$$

which combined with $gHg^{-1} \subset H$ implies that equality holds, i.e.

$$H \trianglelefteq G \iff gHg^{-1} = H \text{ for all } g \in G.$$

We claim also that $gHg^{-1} = H$ if and only if $gH = Hg$. Indeed if $gHg^{-1} = H$, then

$$gH = (gHg^{-1})g = Hg.$$

On the other hand if $gH = Hg$, then

$$gHg^{-1} = \{\, kg^{-1} \mid k \in gH \,\} = \{\, kg^{-1} \mid k \in Hg \,\} = H.$$

We therefore conclude also that

$$H \trianglelefteq G \iff gH = Hg \text{ for all } g \in G.$$

Suppose now that H is a normal subgroup of G, and consider the set G/H of left cosets of H in G. We define a binary operation $*$ on G/H as follows: for $gH, kH \in G/H$ (where $g, k \in G$), we let

$$gH * kH = (gk)H.$$

Note that $(gk)H$ is an element of G/H, but the operation is not obviously well-defined: we need to check that the output $(gk)H$ depends only on the left cosets gH and kH and not the choice of g and k, i.e. that if $gH = g'H$ and $kH = k'H$, then $(gk)H = (g'k')H$. To that end recall that if $gH = g'H$ and $kH = k'H$, then $g^{-1}g' \in H$ and $k^{-1}k' \in H$. Since we assumed that $H \trianglelefteq G$, it follows that $k^{-1}(g^{-1}g')k \in H$, and therefore

$$(gk)^{-1}g'k' = k^{-1}g^{-1}g'k' = (k^{-1}(g^{-1}g')k)(k^{-1}k') \in H,$$

which implies that $(gk)H = (g'k')H$ as required.

Having defined the binary operation $*$ on G/H, we now claim that G/H is a group with respect to this operation. Indeed if gH, kH and ℓH are elements of G/H, then

$$\begin{aligned}(gH * kH) * \ell H &= (gk)H * \ell H = (gk\ell)H \\ &= gH * (k\ell)H = gH * (kH * \ell H),\end{aligned}$$

so the operation is associative. The left coset $H = eH$ is an identity element, since

$$eH * gH = gH = gH * eH$$

for all $gH \in G/H$. Finally if $gH \in G/H$, then

$$gH * g^{-1}H = eH = g^{-1}H * gH,$$

so $g^{-1}H$ is an inverse of gH.

We've been writing the symbol $*$ for emphasis on the definition of the operation, but now we'll start omitting it (along with some of the parentheses), and just write

$$(gH)(kH) = gkH.$$

Recall also that since H was assumed to be normal in G, there is no difference between left and right cosets of H in G: we could just as well write the preceding equation as $(Hg)(Hk) = Hgk$. Consequently, it is common to denote $gH \in G/H$ simply by $[g]$ (and speak of an equivalence class of coset representatives).

DEFINITION 2.17 Suppose that H is a normal subgroup of a group G. Then the group G/H, with the binary operation $(gH)(kH) = gkH$, is called the *quotient group* of G by H.

EXAMPLE 2.18 Returning to Example 2.8, we have that $\langle n \rangle$ is normal in $\mathbb{Z}$ (since $\mathbb{Z}$ is abelian). We can therefore define the quotient group $\mathbb{Z}/\langle n \rangle$, which is just the familiar group of integers modulo n under the addition operation on residue classes:

$$[a] + [b] = [a + b].$$

This group is commonly denoted by $\mathbb{Z}/n\mathbb{Z}$ (which notationally comes with a canonical generator $[1]$).

EXAMPLE 2.19 Returning to Example 1.15 (and 2.14), recall that $Z = \{ rI_n \mid r \in \mathbb{R}^\times \}$ is a normal subgroup of $\mathrm{GL}_n(\mathbb{R})$. We can therefore define the *projective general linear group* $\mathrm{PGL}_n(\mathbb{R})$ as the quotient group $\mathrm{GL}_n(\mathbb{R})/Z$. Its elements are sets of matrices

$$AZ = \{ rA \mid r \in \mathbb{R}^\times \},$$

where $A \in \mathrm{GL}_n(\mathbb{R})$.

EXERCISE 2.20 Let $N \trianglelefteq G$. Prove that there is a canonical bijection

$$\{\text{subgroups of } G \text{ containing } N\} \longleftrightarrow \{\text{subgroups of } G/N\}.$$

3 Homomorphisms and isomorphisms

DEFINITION 3.1 Suppose that G and H are groups, with operations $*_G$ and $*_H$ respectively. We say that a function $f : G \to H$ is a *(group) homomorphism* (from G to H) if

$$f(g *_G g') = f(g) *_H f(g') \quad \text{for all } g, g' \in G.$$

EXAMPLE 3.2 Let $G = H = \mathbb{R}$ (under addition). Then the function $f : G \to H$ defined by $f(x) = 2x$ is a homomorphism, but the function defined by $f(x) = x^2$ is not.

EXAMPLE 3.3 Let $G = H = \mathbb{R}^\times$ (under multiplication). Then the function $f : G \to H$ defined by $f(x) = x^2$ is a homomorphism, but the function defined by $f(x) = 2x$ is not.

EXAMPLE 3.4 Let $G = \mathbb{R}_{>0} = \{ x \in \mathbb{R} \mid x > 0 \}$ (under multiplication). Then the function $f : G \to \mathbb{R}$ defined by $f(x) = \log(x)$ is a homomorphism (since $\log(xy) = \log(x) + \log(y)$ for all $x, y \in G$), but the function defined by $f(x) = x$ is not.

EXAMPLE 3.5 Let $G = H = \mathrm{GL}_n(\mathbb{R})$. Then the function $f : G \to H$ defined by $f(A) = ({}^tA)^{-1}$ is a homomorphism (where ${}^tA$ denotes the transpose of A), but the function defined by $f(A) = A^{-1}$ is not (unless $n = 1$).

EXAMPLE 3.6 The function $f : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$ defined by $f(A) = \det(A)$ is a homomorphism (since $\det(AB) = \det(A)\det(B)$ for all $A, B \in \mathrm{GL}_n(\mathbb{R})$).

EXAMPLE 3.7 Let G be any group and g any element of G. Then the function $f : \mathbb{Z} \to G$ defined by $f(n) = g^n$ is a homomorphism (using Exercise 1.11).

EXAMPLE 3.8 If N is a normal subgroup of a group G, then the function $\pi : G \to G/N$ defined by $\pi(g) = gN$ (called a *quotient map*) is a homomorphism (since $\pi(gh) = ghN = (gN)(hN)$ for all $g, h \in N$ by definition of the operation on G/N).

EXAMPLE 3.9 If $\varphi : G \to H$ and $\theta : H \to K$ are homomorphisms, then so is the composite $\theta \circ \varphi : G \to K$.

EXERCISE 3.10 Suppose that a subset $S \subseteq G$ generates G, i.e. $\langle S \rangle = G$. Prove that a homomorphism $\varphi : G \to H$ is uniquely determined by its restriction $\varphi|_S : S \to H$ as a map of sets. On the other hand, not every map of sets extends to a group homomorphism. We shall use this and its analogues for other objects later in the course often and without comment!

EXERCISE 3.11 In Exercise 1.6, we saw that for any group there is a well-defined map $(-)^{-1} : G \to G$ sending $g \mapsto g^{-1}$. Show that this need not be a group homomorphism. Moreover, show this is a group homomorphism if and only if G is abelian.

For any homomorphism $f : G \to H$, we have

$$e_H f(e_G) = f(e_G) = f(e_G e_G) = f(e_G) f(e_G)$$

(where e_G is the identity of G, e_H is the identity of H, and we omit the symbols for the operations), so the cancellation law implies that $f(e_G) = e_H$. It follows that for any $g \in G$, we have

$$f(g) f(g^{-1}) = f(g g^{-1}) = f(e_G) = e_H,$$

so $f(g^{-1}) = f(g)^{-1}$. Furthermore a straightforward induction argument shows that in fact $f(g^n) = f(g)^n$ for all $n \in \mathbb{Z}$.

DEFINITION 3.12 If $f : G \to H$ is a homomorphism, then the *image* of f is

$$\mathrm{im}(f) = f(G) = \{ f(g) \mid g \in G \},$$

and the *kernel* of f is

$$\ker(f) = \{ g \in G \mid f(g) = e_H \}.$$

It is easy to check that $\mathrm{im}(f)$ is a subgroup of H. Note that by definition, f is surjective if and only $\mathrm{im}(f) = H$. It is also easy to check that $\ker(f)$ is a subgroup of G.

EXERCISE 3.13 Show that a homomorphism $f : G \to H$ is injective if and only if $\ker(f) = \{e_G\}$, i.e. "the kernel is trivial". Show that $\ker(f)$ is in fact a normal subgroup of G.

DEFINITION 3.14 We say that a function $f : G \to H$ is a *(group) isomorphism* if f is a bijective homomorphism. We say that G is *isomorphic* to H if there is an isomorphism $f : G \to H$.

Returning to some of the examples:

EXAMPLE 3.15 The homomorphism $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$ is surjective. Furthermore $\ker(f) = \{0\}$, or equivalently f is injective, so f is an isomorphism.

EXAMPLE 3.16 The homomorphism $f : \mathbb{R}^\times \to \mathbb{R}^\times$ defined by $f(x) = x^2$ is not surjective (since $\mathrm{im}(f) = \mathbb{R}_{>0}$), nor is it injective (since $\ker(f) = \{\pm 1\}$).

EXAMPLE 3.17 The homomorphism $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is an isomorphism, so $\mathbb{R}_{>0}$ is isomorphic to $\mathbb{R}$. We often write $f : G \xrightarrow{\sim} H$ to indicate that f is an isomorphism, and $G \cong H$ to indicate that G is isomorphic to H (without specifying the isomorphism), so for example $\ln : \mathbb{R}_{>0} \xrightarrow{\sim} \mathbb{R}$, and $\mathbb{R}_{>0} \cong \mathbb{R}$.

EXAMPLE 3.18 The homomorphism $f : \mathrm{GL}_n(\mathbb{R}) \to \mathrm{GL}_n(\mathbb{R})$ defined by $f(A) = ({}^tA)^{-1}$ is an isomorphism. An isomorphism from a group G to itself is called an *automorphism* (of G), so f is an automorphism of $\mathrm{GL}_n(\mathbb{R})$, and the isomorphism defined in Example 3.14 is an automorphism of $\mathbb{R}$.

EXAMPLE 3.19 The homomorphism $\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$ is surjective, but not injective unless $n = 1$: its kernel is (by definition) $\mathrm{SL}_n(\mathbb{R})$.

EXAMPLE 3.20 The quotient map $\pi : G \to G/N$ of Example 3.8 is surjective, with kernel N.

EXAMPLE 3.21 If $\varphi : G \to H$ and $\theta : H \to K$ are isomorphisms, then so is $\theta \circ \varphi : G \to K$ (since composites of homomorphisms are homomorphisms, and composites of bijective functions are bijective). Note also that since φ is bijective, it has an inverse as a map of sets $\psi : H \to G$, which is also bijective. Furthermore since $\varphi \circ \psi$ is the identity on H and φ is a homomorphism, it follows that

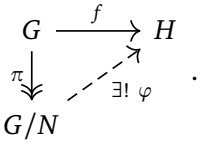
$$\varphi(\psi(hh')) = hh' = \varphi(\psi(h))\varphi(\psi(h')) = \varphi(\psi(h)\psi(h'))$$

for all $h, h' \in H$, and since φ is injective, this implies that $\psi(hh') = \psi(h)\psi(h')$, i.e. ψ is a homomorphism, hence an isomorphism.

For example, the inverse of the isomorphism $\log : \mathbb{R}_{>0} \xrightarrow{\sim} \mathbb{R}$ is the isomorphism $\psi : \mathbb{R} \xrightarrow{\sim} \mathbb{R}_{>0}$ defined by $\psi(x) = e^x$. What is the inverse of the automorphism of $\mathrm{GL}_n(\mathbb{R})$ defined in Example 3.17?

Recall from Exercise 3.12 that if $f : G \to H$ is a homomorphism, then $\ker(f)$ is a normal subgroup of G. We can therefore consider the quotient group G/N, where $N = \ker(f)$.

We claim that there exists a unique map φ such that "the following diagram commutes"



Here π denotes the quotient map of Example 3.8. That the following diagram commutes is taken to mean that for any pair of directed paths in the diagram with the same start and end points, the corresponding compositions of functions agrees. As such, the picture is nothing more than a convenient way of recording algebraic statements which must be satisfied. In this case, there is only one non-trivial condition to check, namely we are asserting that $f = \varphi \circ \pi$. Such a diagram with this property is called a *commutative diagram*.

Note that since π is surjective, if φ exists it is uniquely defined by f. Explicitly, for any $[g] \in G/N$, the condition that $f = \varphi \circ \pi$ forces that

$$\varphi([g]) = \varphi(\pi(g)) = f(g).$$

We claim that this formula, $\varphi([g]) = f(g)$, does indeed a homomorphism $\varphi : G/N \to H$. We first must check that this is a well-defined map of sets. We need to show that if $g, g' \in G$ and $gN = g'N$, then $f(g) = f(g')$. This is true since

$$\begin{aligned} gN = g'N \iff g^{-1}g' \in N \quad &\iff f(g^{-1}g') = e_H \\ &\iff f(g^{-1})f(g') = e_H \iff f(g') = f(g). \end{aligned}$$

Note that this not only shows that φ is well-defined, but injective. Furthermore φ is a homomorphism since

$$\varphi((gN)(g'N)) = \varphi((gg')N) = f(gg') = f(g)f(g') = \varphi(gN)\varphi(g'N)$$

for all $gN, g'N \in G/N$. Note that $\mathrm{im}(\varphi)$ is precisely $\mathrm{im}(f)$. So we may consider φ as a homomorphism $\varphi : G/N \to \mathrm{im}(f)$ which is bijective and therefore an isomorphism. Putting this altogether we have proven the "First Isomorphism Theorem":

THEOREM 3.22 *Let $f : G \to H$ be a group homomorphism. Then f factors*[8] *as a composite of the quotient map $\pi : G \to G/\ker(f)$ and a unique isomorphism $G/\ker(f) \xrightarrow{\varphi} \mathrm{im}(f) \leq H$.*

EXAMPLE 3.23 Recall from Example 3.15 that the homomorphism $f : \mathbb{R}^\times \to \mathbb{R}^\times$ defined by $f(x) = x^2$ has kernel ± 1 and image $\mathbb{R}_{>0}$. We thus have an isomorphism $\mathbb{R}^\times/\{\pm 1\} \xrightarrow{\sim} \mathbb{R}_{>0}$ defined by $\{\pm x\} \mapsto x^2$.

EXERCISE 3.24 Let $f : G \to H$ be a group homomorphism. Show that if N is a normal subgroup contained in $\ker(f)$, then f also factors as the composite of a map $\pi : G \to G/N$ and a unique surjective homomorphism $\varphi : G/N \twoheadrightarrow \mathrm{im}(f)$ that need not be an isomorphism. We often say that f *descends* to a map on G/N (the map φ). Show that $\ker(f)$ is the unique largest subgroup of G with this property for the homomorphism f.

The rest of this section was not lectured

Suppose now that G is a group and that H and K are subgroups of G. Consider the subset of G defined by

$$HK := \{hk \mid h \in H, k \in K\}.$$

This need not be a subgroup of G.

EXERCISE 3.25 Let $G = S_3$, $H = \langle(12)\rangle$ and $K = \langle(13)\rangle$, then

$$HK = \{e, (12), (13), (132)\}$$

is not a subgroup of S_3. Prove this directly from Lagrange's theorem!

On the other hand, when one of the subgroups is normal in G then we always obtain a group:

LEMMA 3.26 *Suppose that H and N are subgroups of a group G and that N is a normal in G. Then $HN = NH$ is a subgroup of G.*

Proof. Note first that

$$HN = \bigcup_{h \in H} hN = \bigcup_{h \in H} Nh = NH,$$

where the middle equality holds since $N \trianglelefteq G$.

Next note that $e = ee \in HN$.

Suppose now that $g, g' \in HN$, so $g = hn$ and $g' = h'n'$ for some $h, h' \in H$, $n, n' \in N$. Then $nh' \in NH = HN$, so $nh' = h''n''$ for some $h'' \in H$, $n'' \in N$. Therefore

$$gg' = hnh'n' = hh''n''n' \in HN,$$

as required.

[8] That is to say f is a composite $f = \varphi \circ \pi$

Finally note that if $g = hn \in HN$ (where $h \in H$, $n \in N$), then $g^{-1} = n^{-1}h^{-1} \in NH = HN$. □

Note that since $N \subset HN$ and $N \trianglelefteq G$, we have $N \trianglelefteq HN$, so we can form the quotient group HN/N. Now consider the composite homomorphism

$$f : H \longrightarrow HN \longrightarrow HN/N,$$

where the first map is the inclusion (sending h to h) and the second is the quotient map of Example 3.8.

We have $h \in \ker(f)$ if and only if $hN = N$, i.e. $h \in N$, so $\ker(f) = H \cap N$. In particular $H \cap N$ is a normal subgroup of H. Furthermore f is surjective since every element of HN/N is of the form $hnN = hN = f(h)$ for some $h \in H$ (and $n \in N$). Therefore it follows from the First Isomorphism Theorem 3.21 that we have an isomorphism

$$\varphi : H/(H \cap N) \xrightarrow{\sim} HN/N$$

defined by $\varphi(h(H \cap N)) = hN$. This is called the "Second Isomorphism Theorem."

EXAMPLE 3.27 Let $G = \mathrm{GL}_n(\mathbb{R})$, $H = \mathrm{SL}_n(\mathbb{R})$, and $N = Z = \{ rI_n \mid r \in \mathbb{R}^\times \}$, as in Example 2.19, so $G/N = \mathrm{PGL}_n(\mathbb{R})$. Since $\det(rI_n) = r^n$, we have $rI_n \in H \cap N$ if and only if $r^n = 1$, so $H \cap N = \{I_n\}$ if n is odd, and $H \cap N = \{\pm I_n\}$ if n is even. The group $H/(H \cap N)$ is denoted $\mathrm{PSL}_n(\mathbb{R})$.

Thus if n is odd, then $\mathrm{PSL}_n(\mathbb{R})$ can be identified with $\mathrm{SL}_n(\mathbb{R})$ (via the isomorphism sending the coset $\{A\}$ to the matrix A). Note also that in this case we can write any $A \in \mathrm{GL}_n(\mathbb{R})$ as $A = rB$, where $r = (\det A)^{1/n}$ and $B = r^{-1}A \in \mathrm{SL}_n(\mathbb{R})$. Therefore $A \in NH$, so $HN = NH = G$, and we conclude that if n is odd, then the map

$$H \xrightarrow{\sim} H/(H \cap N) \xrightarrow{\sim} HN/N = G/N$$

sending A to AN defines an isomorphism $\mathrm{SL}_n(\mathbb{R}) \xrightarrow{\sim} \mathrm{PGL}_n(\mathbb{R})$.

On the other hand if n is even, then we find that

$$HN = \{ A \in \mathrm{GL}_n(\mathbb{R}) \mid \det(A) > 0 \}$$

is a subgroup of index 2 in G (the two cosets being the set of matrices with positive determinant and the set with negative determinant). It follows that $[G/N : HN/N] = [G : HN] = 2$ (the first equality being in the exercise on the "Third Isomorphism Theorem"), so in this case the image of the isomorphism

$$\mathrm{PSL}_n(\mathbb{R}) = \mathrm{SL}_n(\mathbb{R})/\{\pm I_n\} = H/(H \cap N) \xrightarrow{\sim} HN/N$$

is a subgroup of $G/N = \mathrm{PGL}_n(\mathbb{R})$ of index 2.

4 Group actions

DEFINITION 4.1 Let G be a group and let X be a set. An *action* of G on X is a map of sets

$$G \times X \longrightarrow X$$
$$(g, x) \longrightarrow g \cdot x$$

such that

(i) $e \cdot x = x$ for all $x \in X$;

(ii) $g \cdot (h \cdot x) = (gh) \cdot x$ for all $g, h \in G$, $x \in X$.

The choice of an action of G on X makes X into a *G-set*. That is, for a fixed group G, a G-set is a pair $(X, \cdot)$ consisting of a set X and an action "$\cdot$" of G on X. As for groups, it is common to denote a G-set simply by its underlying set when the action is clear.

Note that (ii) ensures that group actions are in some sense associative when considered together with group multiplication. So, as in Remark 1.10, brackets are unnecessary when considering chains of group multiplications and an action.

EXAMPLE 4.2 Let $G = S_n$ and let $X = \{1, 2, \ldots, n\}$. Recall that an element of S_n is a bijective function $\sigma : X \to X$. We can therefore define an action of S_n on X by $\sigma \cdot i = \sigma(i)$ for $\sigma \in S_n$ and $i \in X$. This satisfies the conditions in the definition since (i) $e : X \to X$ is the identity function, so $e \cdot i = e(i) = i$ for all $i \in X$, and (ii) for any $\sigma, \tau \in X$, we have $\sigma\tau = \sigma \circ \tau$, so

$$\sigma \cdot (\tau \cdot i) = \sigma(\tau(i)) = (\sigma \circ \tau)(i) = (\sigma\tau) \cdot i$$

for all $i \in X$. Similarly for any set X, there is an action of the corresponding symmetric group S_X on X defined in exactly the same way.

EXAMPLE 4.3 Similarly to the preceding example, D_n acts[9] on the set of vertices $V = \{v_1, \ldots, v_n\}$ of a regular n-gon. Indeed we may view each symmetry $\sigma \in D_n$ as a function from the n-gon to itself, and its restriction to V defines a bijection function $V \to V$. We thus obtain an action of D_n on V.

EXAMPLE 4.4 $\mathrm{GL}_n(\mathbb{R})$ acts on $\mathbb{R}^n$ by matrix multiplication, i.e. $A \cdot \mathbf{x} = A\mathbf{x}$ for $A \in \mathrm{GL}_n(\mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^n$ (viewed as column vectors). Indeed $I_n\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, and $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all $A, B \in \mathrm{GL}_n(\mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^n$ by associativity of matrix multiplication.

EXAMPLE 4.5 For any group G and set X, there is a "trivial action" of G on X defined by $g \cdot x = x$ for all $g \in G$, $x \in X$.

EXAMPLE 4.6 If a group G acts on two sets X and Y, then it acts on $X \times Y$ by $g \cdot (x, y) = (g \cdot x, g \cdot y)$. Given two group actions, G on X and H on Y, there is an action of $G \times H$ on $X \times Y$ defined by $(g, h) \cdot (x, y) = (g \cdot x, h \cdot y)$.

EXAMPLE 4.7 If G acts on X and H is a subgroup of G, then H acts on Y by restricting the function $G \times X \to X$ to $H \times X \subset G \times X$.

EXAMPLE 4.8 If G acts on X and $f : H \to G$ is any homomorphism of groups, then there is an action of H on X defined by $h \cdot x = f(h) \cdot x$ for $h \in H$, $x \in X$. (Why is this an action, and how is the preceding example a special case of this one?)

There are many natural examples of actions arising directly from constructions in group theory.

EXAMPLE 4.9 Any group G acts on $X = G$ by left multiplication, i.e. $g \cdot x = gx$ for $g \in G$, $x \in G$ (since $ex = x$ and $g(hx) = (gh)x$ for all $g, h, x \in G$).

EXAMPLE 4.10 Right multiplication, i.e. $g \cdot x = xg$ does *not* define an action[10] of G on G (unless G is abelian), since $g \cdot (h \cdot x) = g \cdot (xh) = xhg$ need not be the same as $(gh) \cdot x = xgh$. On the other hand, $g \cdot x = xg^{-1}$ does define an action of G on G.

EXAMPLE 4.11 In the preceding two examples, we saw two (usually different) actions of a group G on itself: $(g, x) \mapsto gx$ and $(g, x) \mapsto xg^{-1}$. Additionally, G acts on itself by *conjugation*, i.e. $g \cdot x = gxg^{-1}$ for $g, x \in G$ (since $exe^{-1} = e$, and $g(hxh^{-1})g^{-1} = (gh)x(gh)^{-1}$ for all $g, h, x \in G$).

[9] We sometimes say "G acts on V (by…)" to mean there is an action of G on V (defined by…)

[10] This would however give an example of a *right action* of G on the set $X = G$, but we will only be discussing *left actions* for now, and just calling them *actions*.

EXAMPLE 4.12 Let H be a (not necessarily normal) subgroup of a group G, and let $X = G/H$ be the set of left cosets of H in G. Then G acts on X by left multiplication, i.e. $g \cdot (kH) = gkH$ for $g \in G$, $x = kH \in X$ (where $k \in G$). First note that this formula is well-defined, i.e. if $x = kH = k'H$ (where $k, k' \in G$), then $gkH = gk'H$ (since $(gk)^{-1}gk' = k^{-1}k' \in H$), and this is an action since $e \cdot kH = kH$, and

$$g \cdot (g' \cdot (kH)) = gg'kH = (gg') \cdot (kH)$$

for all $g, g' \in G$, $kH \in X$.

DEFINITION 4.13 Suppose that G acts on X and $x \in X$. The *orbit* of x (under the action of G) is the subset $G \cdot x := \{ g \cdot x \mid g \in G \}$ of X.

LEMMA 4.14 *If G acts on X, then the set of orbits defines a partition of X into disjoint subsets, i.e. every element of X is in a unique orbit.*

Proof. First note that every element of X is in an orbit. Indeed if $x \in X$, then $x = e \cdot x \in G \cdot x$ is in the orbit of x.

Now we must show that this is the unique orbit containing x, i.e. if $x \in G \cdot y$ for some $y \in X$, then $G \cdot y = G \cdot x$. To see this, note that if $x \in G \cdot y$, then $x = h \cdot y$ for some $h \in G$, so

$$G \cdot x = \{ g \cdot x \mid g \in G \} = \{ g \cdot (h \cdot y) \mid g \in G \} = \{ (gh) \cdot y \mid g \in G \} \subseteq G \cdot y.$$

By symmetry, we also have that $G \cdot y \subseteq G \cdot x$ and so equality. □

EXAMPLE 4.15 Consider the action of $G = S_n$ on $X = \{1, 2, \ldots, n\}$, as in Example 4.2. For any $i, j \in X$, there is an element $\sigma \in G$ such that $\sigma(i) = j$; for example, let $\sigma = e$ if $i = j$ and $\sigma = (i, j)$ if $i \neq j$. Therefore $G \cdot i = X$, for any $i \in X$, i.e. X is the only orbit. This is in fact true for the action of S_X on any set X: for any $x, y \in X$, let $\sigma \in S_X$ be the element defined by $\sigma(x) = y$, $\sigma(y) = x$ and $\sigma(z) = z$ otherwise.

DEFINITION 4.16 We say that an action of G on a set X is *transitive* (or that G *acts transitively* on X) if X is the only orbit; i.e. for every $x, y \in X$, there is an element $g \in G$ such that $g \cdot x = y$.

So for example the action of S_X on X is transitive.

EXAMPLE 4.17 The action of D_n on $V = \{v_1, \ldots, v_n\}$ in Example 4.3 is transitive. Indeed for any pair of vertices, there is a rotation bringing one to the other.

EXAMPLE 4.18 The action of $\mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$ in Example 4.4 is *not* transitive. Indeed $A \cdot \mathbf{0} = \mathbf{0}$ for every $A \in \mathrm{GL}_n(\mathbb{R})$, so the orbit of $\mathbf{0}$ is the set $\{\mathbf{0}\}$. On the other hand if $\mathbf{x}$ is any non-zero element of $\mathbb{R}^n$, then there is an invertible matrix A with first column $\mathbf{x}$ (extend $\mathbf{x} = \mathbf{x}_1$ to any basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ for $\mathbb{R}^n$, and use these as the columns of A). Therefore there an element $A \in \mathrm{GL}_n(\mathbb{R})$ such that $A \cdot \mathbf{e}_1 = \mathbf{x}$, where $\mathbf{e}_1 = {}^t(1, 0, \ldots, 0)$ (again using ^{t} to denote transpose). Therefore every $\mathbf{x} \in \mathbb{R}^n \smallsetminus \{\mathbf{0}\}$ is in the orbit of $\mathbf{e}_1$. It follows that there are precisely two orbits in $\mathbb{R}^n$ under the action of $\mathrm{GL}_n(\mathbb{R})$: $\{\mathbf{0}\}$ and $\mathbb{R}^n \smallsetminus \{\mathbf{0}\}$.

EXAMPLE 4.19 For the trivial action of G on X, as in Example 4.5, the orbits are the singleton sets $\{x\}$ for $x \in X$.

EXAMPLE 4.20 Suppose that G acts on X and on Y and consider the action of G on $X \times Y$ as in Example 4.6. Note that $G \cdot (x, y)$ might *not* be the same as $(G \cdot x) \times (G \cdot y)$: the former is $\{ (g \cdot x, g \cdot y) \mid g \in G \}$, and the latter is $\{ (g \cdot x, h \cdot y) \mid g, h \in G \}$. On the other hand if G acts on X and H acts on Y, then

$$(G \times H) \cdot (x, y) = \{ (g \cdot x, h \cdot y) \mid g \in G, h \in H \} = (G \cdot x) \times (H \cdot y).$$

EXAMPLE 4.21 The actions of G on itself defined by $g \cdot x = gx$ (Example 4.9) and $g \cdot x = xg^{-1}$ (Example 4.10) are both transitive.

EXAMPLE 4.22 Consider the action of G on itself by conjugation, as in Example 4.11. The orbit of h is then its *conjugacy class*

$$G \cdot h = \{ ghg^{-1} \mid g \in G \}.$$

Suppose for example that $G = \mathrm{GL}_n(\mathbb{C})$, and that A is a diagonal matrix with distinct diagonal entries $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}^\times$. The conjugacy class of A is then the set of matrices in $\mathrm{GL}_n(\mathbb{C})$ with characteristic polynomial $\prod_{i=1}^{n}(X - \alpha_i)$. (Why is this true, and why do we need to assume the α_i are distinct?)

EXERCISE 4.23 Show that a subgroup $H \leq G$ is a normal subgroup if and only if H is a union of conjugacy classes.

EXAMPLE 4.24 Suppose that H is a subgroup of G. Then the action of G on G/H by left mutliplication (Example 4.12) is transitive. On the other hand, consider the action of H on G by left multiplication (i.e. $h \cdot g = hg$ for $h \in H$, $g \in G$; see Examples 4.9 and 4.7). The orbit of an element $g \in G$ under the action of H is then

$$H \cdot g = \{ hg \mid g \in G \} = Hg.$$

Therefore the orbits are precisely the *right* cosets of H in G. Similarly, under the action of H on G defined by $h \cdot g = gh^{-1}$ for $h \in H$, $g \in G$, the orbits are the *left* cosets of H in G.

EXERCISE 4.25 Convince yourself that it makes sense to consider "sub-G-sets" of a G-set X. Show that for any $x \in X$, the orbit $G \cdot x$ is a sub-G-set of X. Show that a subset $Y \subseteq X$ is a sub-G-set if any only if it is a union of orbits.

DEFINITION 4.26 Suppose that G acts on X and $x \in X$. The *stabiliser* (or *isotropy group*) of x (under the action of G) is

$$\mathrm{Stab}_G(x) = \{ g \in G \mid g \cdot x = x \}.$$

This is sometimes also denoted G_x.

It is straightforward to check that $\mathrm{Stab}_G(x)$ is indeed a subgroup of G.

EXAMPLE 4.27 Consider the action of $G = S_X$ on X (as in Example 4.2). Then, by definition, we have

$$\mathrm{Stab}_G(x) = \{ \sigma \in S_X \mid \sigma(x) = x \}.$$

So if $X = \{1, 2, 3, 4\}$ (so $G = S_4$), we have

$$\mathrm{Stab}_G(4) = \{ e, (12), (13), (23), (123), (132) \}.$$

EXAMPLE 4.28 Consider the action on $G = D_n$ on $V = \{v_1, \ldots, v_n\}$ as in Example 4.3. Then for any $v \in V$, we have $G_v = \{e, \varphi\}$, where φ is the reflection in the axis of symmetry through v.

EXAMPLE 4.29 For the trivial action of G on X (Example 4.5) we have $\mathrm{Stab}_G(x) = G$ for all $x \in X$.

EXAMPLE 4.30 If G acts on X and on Y, then the stabiliser of an element $(x, y) \in X \times Y$ under the resulting action of G (as in Example 4.6) is $\mathrm{Stab}_G((x, y)) = \mathrm{Stab}_G(x) \cap \mathrm{Stab}_G(y)$. If G acts on X and H acts on Y, then the stabiliser of (x, y) under the resulting action of $G \times H$ on $X \times Y$ is $\mathrm{Stab}_{G \times H}((x, y)) = \mathrm{Stab}_G(x) \times \mathrm{Stab}_H(y)$.

EXAMPLE 4.31 The stabiliser of an element $h \in G$ under the action of G on G by left multiplication is $\{e\}$. (This was the action in Example 4.9, and the same holds for the action defined by $g \cdot h = hg^{-1}$ in Example 4.10.)

EXAMPLE 4.32 Consider the action of G on itself by conjugation (Example 4.11). The stabiliser of $h \in G$ is then

$$\{ g \in G \mid ghg^{-1} = h \} = \{ g \in G \mid gh = hg \} = Z_G(h),$$

i.e. the centralizer of h in G.

We have the following fundamental relation between orbits and stabilisers.

PROPOSITION 4.33 (Orbit–Stabiliser theorem) *Suppose that G is a group acting on a set X, and $x \in X$. Then for any $x \in X$, there is a bijection*

$$G/\mathrm{Stab}_G(x) \to G \cdot x$$

defined by $g\mathrm{Stab}_G(x) \mapsto g \cdot x$ for $g \in G$.

Proof. We first show that this is well-defined; i.e. the output is independent of how the coset $g\mathrm{Stab}_G(x)$ in $G/\mathrm{Stab}_G(x)$ was written. Specifically, we want to show that if $g\mathrm{Stab}_G(x) = h\mathrm{Stab}_G(x)$, then $g \cdot x = h \cdot x$. Indeed we have

$$\begin{aligned} g\mathrm{Stab}_G(x) = h\mathrm{Stab}_G(x) &\iff h^{-1}g \in \mathrm{Stab}_G(x) \iff (h^{-1}g) \cdot x = x \\ &\iff g \cdot x = (hh^{-1}g) \cdot x = h \cdot (h^{-1}g \cdot x) = h \cdot x, \end{aligned}$$

as required. So there is a well-defined map of sets $f : G/\mathrm{Stab}_G(x) \to G \cdot x$ given by $f(g\mathrm{Stab}_G(x)) = g \cdot x$.

Running the above implications in reverse shows that f is injective, i.e. if $f(g\mathrm{Stab}_G(x)) = g \cdot x = h \cdot x = f(h\mathrm{Stab}_G(x))$, then $g\mathrm{Stab}_G(x) = h\mathrm{Stab}_G(x)$. Finally f is surjective by the definition of $G \cdot x$: if $y \in G \cdot x$, then $y = g \cdot x = f(g\mathrm{Stab}_G(x))$ for some $g \in G$. □

REMARK 4.34 Note that $G/\mathrm{Stab}_G(x)$ is also a G-set via Example 4.12. We did not introduce the notion of a map of G-sets $X \to X'$. This is a map of the underlying sets that "commutes" with the action of G. In this terminology, we strengthen Proposition 4.33 to saying that canonical map $G/\mathrm{Stab}_G(x) \to G \cdot x$ is an "isomorphism" of G-sets.

COROLLARY 4.35 *If G is a group acting on a (possibly infinite) set X, and $x \in X$, then $[G : \mathrm{Stab}_G(x)] = \#(G \cdot x)$.*

EXAMPLE 4.36 Consider the action of $G = S_n$ on $X = \{1, 2, \ldots, n\}$, and suppose $x \in X$. Since action is transitive (Example 4.15), the orbit $G \cdot x = X$ has cardinality n. So by Corollary 4.35, we have $[G : \mathrm{Stab}_G(x)] = n$. Since $G = S_n$ has order $n!$, it follows that $\mathrm{Stab}_G(x)$ has order $n!/n = (n-1)!$. In fact $\mathrm{Stab}_G(x)$ is isomorphic to the symmetric group on the set $Y = X \smallsetminus \{x\}$ with $n-1$ elements: if $\sigma \in \mathrm{Stab}_G(x)$, then $\sigma(x) = x$, so if $y \neq x$, then $\sigma(y) \neq x$, i.e. σ restricts to a function $\sigma' = \sigma|_Y : Y \to Y$. Furthermore σ' is bijective: its inverse is the restriction of σ^{-1} to Y. We thus obtain a function $\varphi : \mathrm{Stab}_G(x) \to S_Y$, defined by $\sigma \mapsto \sigma|_Y$, and φ is clearly a homomorphism since $(\sigma \circ \tau)|_Y = \sigma|_Y \circ \tau|_Y$ for all $\sigma, \tau \in \mathrm{Stab}_G(x)$. Finally φ is bijective since it has an inverse function: every $\sigma' \in S_Y$ extends uniquely to a permutation $\sigma : X \to X$ such that $\sigma(x) = x$, i.e. an element $\sigma \in \mathrm{Stab}_G(x)$ such that $\sigma|_Y = \sigma'$.

EXAMPLE 4.37 Consider the action of $G = D_n$ on $V = \{v_1, \ldots, v_n\}$, and suppose $v \in V$. Recall from Example 4.2 that the action is transitive, so $\#(G \cdot v) = \#V = n$. Therefore $[G : G_v] = n$, so G_v has order $\#G/n = 2$, as we already saw in Example 4.37.

EXAMPLE 4.38 Consider the action of $G = \mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$ as in Example 4.4. Recall from Example 4.18 that the two orbits are $G \cdot \mathbf{0} = \{\mathbf{0}\}$ and $G \cdot \mathbf{e}_1 = \mathbb{R}^n \smallsetminus \{\mathbf{0}\}$, where

$\mathbf{e}_1 = {}^t(1,0,\dots,0)$. To describe the stabilisers, note that $\mathrm{Stab}_G(\mathbf{0}) = G$ and that $\mathrm{Stab}_G(\mathbf{e}_1)$ is the set of $A \in \mathrm{GL}_n(\mathbb{R})$ such that $A\mathbf{e}_1 = \mathbf{e}_1$, i.e. the first column of A is $\mathbf{e}_1$, so

$$G_{\mathbf{0}} = \left\{ \begin{pmatrix} 1 & {}^t\mathbf{y} \\ \mathbf{0} & B \end{pmatrix} \middle| \, \mathbf{y} \in \mathbb{R}^{n-1}, B \in \mathrm{GL}_{n-1}(\mathbb{R}) \right\}.$$

We can describe the stabiliser of an arbitrary element of $\mathbb{R}^n \setminus \{\mathbf{0}\}$ using the following lemma.

LEMMA 4.39 *Suppose that G acts on X, and that $g \in G$, $x, y \in X$ are such that $y = g \cdot x$. Then $\mathrm{Stab}_G(y) = g\mathrm{Stab}_G(x)g^{-1}$.*

Proof. We have

$$\begin{aligned} h \in G_y &\iff h \cdot y = y \iff (hg) \cdot x = g \cdot x \\ &\iff (g^{-1}hg) \cdot x = x \iff g^{-1}hg \in G_x \iff h \in gG_xg^{-1}, \end{aligned}$$

as required. □

EXAMPLE 4.40 Returning to Example 4.38, recall that the stabilizer of $\mathbf{e}_1$ is the set of matrices in $G = \mathrm{GL}_n(\mathbb{R})$ with first column $\mathbf{e}_1$. Recall also from 4.18 that if $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, then $\mathbf{x} = A\mathbf{e}_1$, where A is any matrix in $\mathrm{GL}_n(\mathbb{R})$ whose first column is $\mathbf{x}$. It follows from Lemma 4.39 that $\mathrm{Stab}_G(\mathbf{x}) = AG_{\mathbf{0}}A^{-1}$; in particular that $\mathrm{Stab}_G(\mathbf{x})$ is conjugate to $\mathrm{Stab}_G(\mathbf{e}_1)$.

EXAMPLE 4.41 Let H be a subgroup of G, and consider the (transitive) action of G on $X = G/H$ as in Example 4.12. The stabiliser of the coset $H \in X$ is the set of $g \in G$ such that $gH = H$, i.e. the stabilizer of H is H. Now consider an arbitrary coset gH. According to Lemma 4.39, the stabiliser of gH is gHg^{-1}.

DEFINITION 4.42 We say that a group G is *solvable* if there is a chain of subgroups

$$\{e\} = G_0 \leq G_1 \leq \cdots \leq G_{n-1} \leq G_n = G$$

such that the following hold for $i = 1, \dots, n$:

- G_{i-1} is normal in G_i;
- G_i/G_{i-1} is abelian.

Note that we do not insist that G_i is normal in G (cf. Remark 2.15). Note that any abelian group G is solvable: just take $G_1 = G$.

EXAMPLE 4.43 The group A_5 is not solvable. (Recall that A_n denotes the subgroup of S_n consisting of *even* permutations, i.e. those that can be written as the product of an even number of transpositions (i, j).) To see this we first claim that its conjugacy classes are represented by

- $[e] = \{e\}$ having 1 element;
- $[(12)(34)]$ having 15 elements;
- $[(123)]$ having 20 elements;
- $[(12345)]$ having 12 elements;
- $[(12354)]$ having 12 elements.

(Note that the conjugacy class of (12345) in S_5 splits into two classes in A_5, whose elements are only conjugate to each other S_5 via odd permutations.) We leave this to the exercises.

Now we know that any normal subgroup N is a union of conjugacy classes by Exercise 4.23, so of order a sum of the above numbers. In fact, as a subgroup it necessarily contains e its order is of the form $1 + (\dots)$, where the bracket runs over the sizes of the

some subset of the remaining non-trivial conjugacy classes. But by Lagrange's Theorem 2.10, $\#N$ divides the order of A_5, i.e. 60. By considering the specific numbers above, we find that the only normal subgroups of A_5 are $\{e\}$ and A_5 itself. Since $A_5/\{e\} = A_5$ is not abelian, it cannot be solvable. In fact, A_5 is *simple*, that is it is a nontrivial group whose only normal subgroups are $\{e\}$ and G.

Class equation and solubility of prime order groups (not lectured)

Returning now to the action of a group G on itself by conjugation (Example 4.11), recall that the orbit of $h \in G$ is its conjugacy class $\{ ghg^{-1} \mid g \in G \}$ (see Example 4.22), and its stabiliser is the centralizer of h in G, i.e. $Z_G(h)$ (see Example 4.32). Suppose now that G is a finite group, and let $[h]$ denote the conjugacy class of h in G. Note that it follows from Corollary 4.35 that its cardinality

$$\#[h] = [G : Z_G(h)] = \#G/\#Z_G(h)$$

divides the order of the group G. Furthermore

$$\#[h] = 1 \iff [h] = \{h\} \iff Z_G(h) = G \iff h \in Z(G),$$

where $Z(G) = \{ h \in G \mid gh = hg \text{ for all } g \in G \}$ is the *center* of the group G. Since the set of orbits defines a partition of G (Lemma 4.14), we obtain the following formula, called the *Class Equation* for G:

THEOREM 4.44 *Let G be a finite group, and let C denote the set of conjugacy classes in G. Then*

$$\#G \;=\; \sum_{[g]\in C} \#[g] \;=\; \#Z(G) + \sum_{[g]\in C,\, \#[g]>1} \#[g]\,,$$

and each term $\#[g] = [G : Z_G(g)]$ is a divisor of $\#G$.

EXAMPLE 4.45 Let us determine the Class Equation for the dihedral group

$$D_n = \{ e, \rho, \rho^2, \ldots, \rho^{n-1}, \varphi, \varphi\rho, \varphi\rho^2, \ldots, \varphi\rho^{n-1} \},$$

where ρ is any rotation of order n, and φ is any reflection.

Suppose first that n is odd. Using that $\varphi\rho^i\varphi^{-1} = \varphi\rho^i\varphi = \rho^{-i}$ for all i, we find that $[\rho^i] = \{\rho^i, \rho^{n-i}\}$ has two elements for $i = 1, \ldots, (n-1)/2$. (Note that accordingly $Z_G(\rho^i)$ has order n, and since it contains $\langle\rho\rangle$, it follows that $Z_G(\rho^i) = \langle\rho\rangle$ for $i = 1, \ldots, n-1$.) Furthermore since $\rho^{-i}\varphi\rho^i = \varphi\rho^{2i}$, we find that $[\varphi]$ contains the n elements of the form $\varphi\rho^j$, but no elements of the form ρ^j (and accordingly $Z_G(\varphi\rho^j)$ has order 2, and hence equals $\langle\varphi\rho^j\rangle$). Of course $[e] = \{e\}$ and $Z_G(e) = G$. Therefore the Class Equation is:

$$2n = 1 + \underbrace{2 + 2 + \cdots + 2}_{(n-1)/2 \text{ times}} + n.$$

On the other hand if n is even, then $\rho^{n/2} \in Z(G)$, whereas $[\rho^i] = \{\rho^i, \rho^{n-i}\}$ has two elements for $i = 1, \ldots, (n/2)-1$ (in which case again $Z_G(\rho^i) = \langle\rho\rangle$). Now however

$$[\varphi] = \{ \varphi, \varphi\rho^2, \ldots, \varphi\rho^{n-2} \} \quad \text{and} \quad [\varphi\rho] = \{ \varphi\rho, \varphi\rho^3, \ldots, \varphi\rho^{n-1} \}$$

each have $n/2$ elements (and accordingly $Z_G(\varphi\rho^j) = \langle\varphi\rho^j, \rho^{n/2}\rangle$ has order 4). Therefore in this case the Class Equation is:

$$2n = \underbrace{1 + 1}_{\#Z(G)} + \underbrace{2 + 2 + \cdots + 2}_{(n/2)-1 \text{ times}} + n/2 + n/2.$$

EXAMPLE 4.46 The conjugacy class of an element $\sigma \in S_n$ is determined by its "cycle structure" as follows. We can write σ as a product of disjoint cycles[11]

$$(i_1^{(1)}, \ldots, i_{n_1}^{(1)})(i_1^{(2)}, i_2^{(2)}, \ldots, i_{n_2}^{(2)}) \cdots (i_1^{(k)}, i_2^{(k)}, \ldots, i_{n_k}^{(k)}),$$

where $1 \leq n_1 \leq n_2 \leq \cdots \leq n_k$ and $n_1 + n_2 + \cdots + n_k = n$. If $\sigma' = \tau\sigma\tau^{-1}$, then σ' has the same cycle structure, with each $i_\ell^{(j)}$ replaced by $\tau(i_\ell^{(j)})$, and conversely any element with the same cycle structure as σ is in the conjugacy class of σ. It follows that the conjugacy classes of S_n are in bijection with the set of partitions of n. For example if $n = 5$, then we have the following conjugacy classes (counting the size of each with some elementary combinatorics):

- $[e] = \{e\}$ (partition $1+1+1+1+1$);
- $[(12)]$ has 10 elements (partition $1+1+1+2$);
- $[(12)(34)]$ has 15 elements (partition $1+2+2$);
- $[(123)]$ has 20 elements (partition $1+1+3$);
- $[(12)(345)]$ has 20 elements (partition $2+3$);
- $[(1234)]$ has 30 elements (partition $1+4$);
- $[(12345)]$ has 24 elements (partition 5).

So the Class Equation for S_5 is

$$120 = 1 + 10 + 15 + 20 + 20 + 30 + 24.$$

Theorem 4.44 has the following immediate consequence:

COROLLARY 4.47 *Suppose that p is prime and G has order p^r for some $r \geq 1$. Then $\#Z(G) > 1$.*

Proof. If $g \in G$ and $g \notin Z(G)$, then $\#[g] > 1$, and $\#[g] = [G : Z_G(g)]$ divides $\#G = p^r$, so $\#[g] = p^s$ for some $s \geq 1$. In particular $\#[g]$ is divisible by p, and therefore so is

$$\#Z(G) \;=\; \#G \;-\; \sum_{[g] \in C,\, \#[g] > 1} \#[g].$$

□

This in turn can be used to prove that groups of prime-power are order are solvable. First we recall the definition of solvability, and some examples and properties.

DEFINITION 4.48 We say that a group G is *solvable* if there is a chain of subgroups

$$\{e\} = G_0 \subset G_1 \subset \cdots \subset G_{n-1} \subset G_n = G$$

such that the following hold for $i = 1, \ldots, n$:

- G_{i-1} is normal in G_i;
- G_i / G_{i-1} is abelian.

Note that any abelian group G is solvable: just take $G_1 = G$.

EXAMPLE 4.49 The group S_4 is solvable. Indeed consider the chain of subgroups

$$\{e\} = G_0 \subset G_1 \subset G_2 \subset G_3 = S_4,$$

where $G_1 = \{e, (12)(34), (13)(24), (14)(23)\}$ and $G_2 = A_4$. We then have that $[S_4 : A_4] = 2$, so $A_4 \trianglelefteq S_4$ and S_4/A_4 has order 2, hence is abelian; $G_1 \trianglelefteq A_4$ (in fact $G_1 \trianglelefteq S_4$) and $\#(G_1/A_4) = 3$ is prime, so G_1/A_4 is cyclic, hence abelian; finally $\{e\} \trianglelefteq G_1$ and $G_1/\{e\}$ is isomorphic to G_1, which is abelian.

[11] We view a 1-cycle as the identity, so for example $(123) = (4)(123)$ in S_4 has $k = 2$, $n_1 = 1$, $n_2 = 3$.

PROPOSITION 4.50 *Suppose that G is a group and H is a subgroup of G. If G is solvable, then so is H. Furthermore if H is normal in G, then*

$$G \text{ is solvable} \iff H \text{ and } G/H \text{ are both solvable.}$$

Proof. Suppose that G is solvable, so there is a chain of subgroups

$$\{e\} = G_0 \subset G_1 \subset \cdots \subset G_{n-1} \subset G_n = G$$

such that each $G_{i-1} \trianglelefteq G_i$ and G_i/G_{i-1} is abelian. Letting $H_i = H \cap G_i$ for $i = 0, \ldots, n$ gives a chain of subgroups

$$\{e\} = H_0 \subset H_1 \subset \cdots \subset H_{n-1} \subset H_n = H.$$

For $i = 1, \ldots, n$, we have that H_{i-1} is the kernel of the composite homomorphism $H_i \hookrightarrow G_i \to G_i/G_{i-1}$, so $H_{i-1} \trianglelefteq H_i$ and H_i/H_{i-1} is isomorphic to a subgroup of the abelian group G_i/G_{i-1}, hence is abelian. Therefore H is solvable.

Suppose now that H is normal in G, let $\overline{G} = G/H$, and write $\overline{g} = gH \in \overline{G}$ for $g \in G$. If G is solvable, then we have already seen that H is solvable. Considering the same chain of subgroups of G as above, let $\overline{G}_i = \{\, \overline{g} \mid g \in G_i \,\}$ be the image of G_i under the homomorphism to $\overline{G}$ defined by $g \mapsto \overline{g}$. This gives a chain of subgroups

$$\{\overline{e}\} = \overline{G}_0 \subset \overline{G}_1 \subset \cdots \subset \overline{G}_{n-1} \subset \overline{G}_n = \overline{G},$$

and each $\overline{G}_{i-i}$ is normal in $\overline{G}_i$ since if $g \in G_{i-1}$ and $k \in G_i$, then $\overline{g}\overline{k}\overline{g}^{-1} = \overline{gkg^{-1}} \in \overline{G}_i$. Furthermore since G_{i-1} is contained in the kernel of the composite of surjective homomorphisms

$$G_i \longrightarrow \overline{G}_i \longrightarrow \overline{G}_i/\overline{G}_{i-1},$$

there is a well-defined surjective homomorphism $G_i/G_{i-1} \to \overline{G}_i/\overline{G}_{i-1}$, and since G_i/G_{i-1} is abelian, it follows that so is $\overline{G}_i/\overline{G}_{i-1}$. This proves that $\overline{G}$ is solvable.

Finally, continue to assume that H is normal in G, but now that H and $\overline{G} = G/H$ are solvable. We thus have a chain of subgroups

$$\{e\} = H_0 \subset H_1 \subset \cdots \subset H_{m-1} \subset H_m = H$$

such that each $H_{i-1} \trianglelefteq H_i$ and H_i/H_{i-1} is abelian, and similarly a chain

$$\{\overline{e}\} = \overline{G}_0 \subset \overline{G}_1 \subset \cdots \subset \overline{G}_{m-1} \subset \overline{G}_n = \overline{G}$$

such that each $\overline{G}_{j-1} \trianglelefteq \overline{G}_j$ and $\overline{G}_j/\overline{G}_{j-1}$ is abelian. We then let $G_j = \{\, g \in G \mid \overline{g} \in \overline{G}_j \,\}$ to get a chain of subgroups

$$H = G_0 \subset G_1 \subset \cdots \subset G_{n-1} \subset G_n = G$$

such that each $G_{j-1} \trianglelefteq G_j$ and $G_j/G_{j-1} \xrightarrow{\sim} \overline{G}_j/\overline{G}_{j-1}$ is abelian. Therefore the chain

$$\{e\} = H_0 \subset H_1 \subset \cdots \subset H_m = G_0 \subset G_1 \subset \cdots \subset G_n = G$$

shows that G is solvable. □

EXAMPLE 4.51 If $n \geq 5$, then S_n is not solvable. Indeed S_n has a subgroup isomorphic to S_5, and hence a subgroup isomorphic to A_5. So if S_n were solvable, then it would follow from Proposition 4.50 that A_5 were solvable, contradicting the conclusion of Example 4.43.

PROPOSITION 4.52 *Suppose that p is prime and G has order p^r for some $r \geq 0$. Then G is solvable.*

Proof. We prove the proposition by induction on r. If $r = 0$, then $G = \{e\}$ is clearly solvable.

Suppose now that $r > 1$, and every group of order p^s for $s < r$ is solvable. By Corollary 4.47, $\#Z(G) > 1$, and since $\#Z(G)$ divides $\#G = p^r$, we have $\#Z(G) = p^t$ with $1 \leq t \leq r$. Note that it follows from the definition of $Z(G)$ that $Z(G)$ is abelian and $Z(G) \trianglelefteq G$. Furthermore $\#(G/Z(G)) = p^s$, where $s = r - t < r$, so the induction hypothesis implies that $G/Z(G)$ is solvable. Since $Z(G)$ and $G/Z(G)$ are both solvable, Proposition 4.50 implies that G is solvable. □

5 Categories and actions

DEFINITION 5.1 A *category* $\mathcal{C}$ consists of

- a class[12] of *objects* of $\mathcal{C}$ (sometimes denoted $\mathrm{Ob}(\mathcal{C})$);
- for each pair of objects A, B of $\mathcal{C}$, a set of *morphisms* from A to B, denoted $\mathrm{Hom}_{\mathcal{C}}(A, B)$;
- for each triple of objects A, B, C of $\mathcal{C}$, a *composition* map

$$\begin{array}{ccccc} \mathrm{Hom}_{\mathcal{C}}(A,B) & \times & \mathrm{Hom}_{\mathcal{C}}(B,C) & \to & \mathrm{Hom}_{\mathcal{C}}(A,C) \\ (\,f & , & g\,) & \mapsto & g \circ f \end{array}$$

such that the following hold:

(i) for each object B of $\mathcal{C}$, there is a morphism $\mathrm{id}_B \in \mathrm{Hom}_{\mathcal{C}}(B, B)$ with the property that for all objects A of $\mathcal{C}$ and morphisms $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$, $g \in \mathrm{Hom}_{\mathcal{C}}(B, A)$, we have

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

(ii) for all objects A, B, C, D of $\mathcal{C}$ and morphisms $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$, $g \in \mathrm{Hom}_{\mathcal{C}}(B, C)$, $h \in \mathrm{Hom}_{\mathcal{C}}(C, D)$, we have

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

Before giving examples, we note that there can only be one morphism id_B satisfying the condition in (ii): if id'_B were another, then we would have

$$\mathrm{id}'_B = \mathrm{id}'_B \circ \mathrm{id}_B = \mathrm{id}_B$$

(the first equality by (i) for id_B, and the second by (i) for id'_B). Note that this was the same argument as for uniqueness of the identity element of a group. We call id_B the *identity morphism* on B (in $\mathcal{C}$), and naturally we refer to (ii) as "associativity of composition."

The following is the prototypical example of a category:

EXAMPLE 5.2 Define Sets to be the category whose

- objects are sets[13];

[12] We do not define here what we mean by class! Classes are generalisations of sets (note we have not defined this either) that are allowed to be "extremely large" in some sense. For our purposes, not much intuition will be lost by reading "set" in place of "class". In fact, if we modify Definition 5.1 by insisting the objects form a set, then we obtain the perfectly good definition of a *small category*. The reason do not do this is that many key examples of categories do not have this property (for example Sets, Gps). In fact, some authors use a more general definition of category where the morphisms are also allowed to be classes rather than sets. To them our definition is of a *locally small category*. This is much less important as all the examples we give are locally small.

[13] There is no set of all sets. This is genuinely a class!

- morphisms are maps of sets, i.e. for each pair of sets A, B, let $\mathrm{Hom}_{\mathsf{Sets}}(A,B) = \{\text{functions } f : A \to B\}$;
- composition is the usual composition of functions.

Condition (i) is satisfied by the identity function on B for every set B. Finally condition (ii) is satisfied since composition of functions is associative.

EXAMPLE 5.3 Define Gps to be the category whose

- objects are groups;
- morphisms are homomorphisms of groups, i.e. for each pair of sets G, H, $\mathrm{Hom}_{\mathsf{Gps}}(G,H)$ is the set of group homomorphisms $f : G \to H$;
- composition is the usual composition of functions. (Recall that composites of homomorphisms are homomorphisms, so if G, H and K are any groups, then composition does indeed define a map

$$\mathrm{Hom}_{\mathsf{Gps}}(G,H) \times \mathrm{Hom}_{\mathsf{Gps}}(H,K) \to \mathrm{Hom}_{\mathsf{Gps}}(G,K).)$$

The same reasoning as in Example 5.2 shows that the conditions in the definition of a category are satisfied (using also the fact that if H is a group, then id_H is a homomorphism). If we only consider abelian groups, then we also obtain a category denoted Ab.

EXAMPLE 5.4 Similarly we define $\mathbb{R}$-vec to be the category whose

- objects are vector spaces over $\mathbb{R}$;
- morphisms are $\mathbb{R}$-linear functions, i.e. if U and V are vector spaces over $\mathbb{R}$, then $\mathrm{Hom}_{\mathbb{R}\text{-}\mathsf{vec}}(U,V)$ is the set of $\mathbb{R}$-linear functions $f : U \to V$;
- composition is the usual composition of functions.

Again this forms a category since identity maps are $\mathbb{R}$-linear, and composites of $\mathbb{R}$-linear functions are $\mathbb{R}$-linear.

Before giving another example, let us introduce some standard shorthand notation for morphisms: we write $f : A \to B$ or $A \xrightarrow{f} B$ to denote an element $f \in \mathrm{Hom}_{\mathcal{C}}(A,B)$ (where A and B are objects of $\mathcal{C}$). Thus given $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$, we have the element $g \circ f \in \mathrm{Hom}_{\mathcal{C}}(A,C)$, which we denote $A \xrightarrow{f} B \xrightarrow{g} C$. Thus condition (iii) means that $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$ unambiguously denotes $h \circ (g \circ f) = (h \circ g) \circ f$.

EXAMPLE 5.5 Let $\mathcal{C}$ be the category with two objects A, B and three morphisms $\mathrm{id}_A : A \to A, \mathrm{id}_B : B \to B$ and one denoted $f : A \to B$. We can represent this pictorially[14] as:

$$\mathrm{id}_A \circlearrowright A \xrightarrow{\quad f \quad} B \circlearrowright \mathrm{id}_B \,.$$

Note that a category consists of three pieces of data: the objects, the morphisms and the composition maps. The reason we have not specified the composition rules is that they are completely forced in this case! For example, we know that $f \circ \mathrm{id}_A = f$ by the condition (i) in Definition 5.1. When writing categories pictorially like this, it is common to omit forced information. For example, one may not include all the identity maps. Note the category has no knowledge of "what" the morphism f is, just that it exists and how it composes with other morphisms.

Note that in Examples 5.2, 5.3 and 5.4, the objects were sets (with some additional structure in Examples 5.3 and 5.4), and the morphisms were maps of the underlying sets

[14] As we shall see later, this is an example of a directed graph (Definition 17.10)

(possibly) preserving additional structure. This was not the case for Example 5.5, we have no concept of what objects A, B "are" this is not contained in the data of a category. Consequently, the morphisms are not just maps of the "underlying sets" of A, B. There are many interesting categories that don't admit a "concrete" description in terms of sets.

EXAMPLE 5.6 Let G be any group. Let A be any set; for example $A = \emptyset$. Let $\mathcal{C}_G$ be the category whose only object is A, $\mathrm{Hom}_{\mathcal{C}_G}(A,A) = G$, and $g \circ h = gh$ for all $g, h \in G$. Then (i) holds since A is the only object, (ii) is satisfied by the identity e_G, and (iii) holds since the operation on G is associative.

EXAMPLE 5.7 We now define another category $\mathcal{C}$ whose objects are vector spaces over $\mathbb{R}$ (as in Example 5.4), but now let $\mathrm{Hom}_{\mathcal{C}}(U,V)$ be the set of equivalence classes in $\mathrm{Hom}_{\mathbb{R}\text{-vec}}(U,V)$ under the equivalence relation defined by $f \sim f'$ if $f' = rf$ for some $r \in \mathbb{R}^\times$. It is easy to check that this in fact an equivalence relation, and we let $[f]$ denote the equivalence class of f. We still need to define the composition map

$$\mathrm{Hom}_{\mathcal{C}}(U,V) \times \mathrm{Hom}_{\mathcal{C}}(V,W) \to \mathrm{Hom}_{\mathcal{C}}(U,W)$$

for all objects U, V and W of $\mathcal{C}$. We wish to do this by the formula

$$[g] \circ [f] = [g \circ f]$$

for $[f] \in \mathrm{Hom}_{\mathcal{C}}(U,V)$ and $[g] \in \mathrm{Hom}_{\mathcal{C}}(V,W)$ (where $f \in \mathrm{Hom}_{\mathbb{R}\text{-vec}}(U,V)$ and $g \in \mathrm{Hom}_{\mathbb{R}\text{-vec}}(V,W)$, so $g \circ f \in \mathrm{Hom}_{\mathbb{R}\text{-vec}}(U,W)$), but we need to check that this is well-defined on conjugacy classes, i.e. if $f \sim f'$ and $g \sim g'$, then $g \circ f \sim g' \circ f'$. This holds since if $f' = rf$ for some $r \in \mathbb{R}^\times$ and $g' = sg$ for some $s \in \mathbb{R}^\times$, then $g' \circ f' = (sg) \circ (rf) = (rs)(g \circ f)$.

This notion of objects, morphisms and composition does indeed satsify the conditions in Definition 5.1: (i) by definition, (ii) by $[\mathrm{id}_V] \in \mathrm{Hom}_{\mathcal{C}}(V,V)$ and (iii) since

$$\begin{aligned} [h] \circ ([g] \circ [f]) &= [h] \circ [g \circ f] &&= [h \circ (g \circ f)] \\ &= [(h \circ g) \circ f] &&= [h \circ g] \circ [f] &&= ([h] \circ [g]) \circ [f] \end{aligned}$$

for all objects U, V, W, X of $\mathcal{C}$ and morphisms $[f] \in \mathrm{Hom}_{\mathcal{C}}(U,V)$, $[g] \in \mathrm{Hom}_{\mathcal{C}}(V,W)$, $[h] \in \mathrm{Hom}_{\mathcal{C}}(W,X)$.

DEFINITION 5.8 Let $\mathcal{C}$ be a category and $A \xrightarrow{f} B$ a morphism in $\mathcal{C}$. We say that f is an *isomorphism* (in $\mathcal{C}$) if there is a morphism $B \xrightarrow{g} A$ in $\mathcal{C}$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$, and we call g the *inverse* of f in $\mathcal{C}$.

Thus for example an isomorphism in Sets is a bijection, and an isomorphism in Gps is an isomorphism of groups. (Recall that if $G \xrightarrow{f} H$ is a bijective homomorphism of groups, then its inverse map is also a homomorphism. We'll see later though that there are categories whose morphisms are structure-preserving functions, but not every bijection is an isomorphism.)

DEFINITION 5.9 If A is an object of $\mathcal{C}$, then an *automorphism* of A (in $\mathcal{C}$) is an isomorphism $A \xrightarrow{f} A$ (in $\mathcal{C}$). We let $\mathrm{Aut}_{\mathcal{C}}(A)$ denote the set of automorphisms of A.

The following proposition is **the** reason groups are ubiquitous across mathematics!

PROPOSITION 5.10 *If A is an object of a category $\mathcal{C}$, then $\mathrm{Aut}_{\mathcal{C}}(A)$ is a group under composition.*

Proof. We first prove that if $f, g \in \mathrm{Aut}_{\mathcal{C}}(A)$, then $f \circ g \in \mathrm{Aut}_{\mathcal{C}}(A)$. By assumption $A \xrightarrow{f} A$ and $A \xrightarrow{g} A$ are isomorphisms, so there are morphisms $A \xrightarrow{f'} A$ and $A \xrightarrow{g'} A$ such that $f' \circ f = f \circ f' = \mathrm{id}_A$ and $g' \circ g = g \circ g' = \mathrm{id}_A$. It follows that

$$(g' \circ f') \circ (f \circ g) = (g' \circ (f' \circ f)) \circ g = (g' \circ \mathrm{id}_A) \circ g = g' \circ g = \mathrm{id}_A$$

(using associativity of composition twice for the first equality), and similarly $(f \circ g) \circ (g' \circ f') = \mathrm{id}_A$, so $A \xrightarrow{f \circ g} A$ is indeed an isomorphism, i.e. $f \circ g \in \mathrm{Aut}_{\mathcal{C}}(A)$.

We already know that composition is associative (by the definition of a category). We have that id_A is an automorphism (its inverse being id_A), and it is an identity element for the composition operation. Finally if $f \in \mathrm{Aut}_{\mathcal{C}}(A)$, then its inverse morphism f' is also an automorphism (since it has an inverse, namely f), so f has an inverse in $\mathrm{Aut}_{\mathcal{C}}(A)$ with respect to the composition operation. Therefore $\mathrm{Aut}_{\mathcal{C}}(A)$ is a group. □

EXAMPLE 5.11 If X is any set, then $\mathrm{Aut}_{\mathsf{Sets}}(X)$ is the set of bijective functions $f : X \to X$, i.e. the set of permutations of X. Furthermore the operation is composition, so $\mathrm{Aut}_{\mathsf{Sets}}(X) = S_X$.

EXAMPLE 5.12 If G is a group, then $\mathrm{Aut}_{\mathsf{Gps}}(G)$ is the group of automorphisms of the group G (usually just denoted $\mathrm{Aut}(G)$).

EXAMPLE 5.13 If V is a vector space over $\mathbb{R}$, then $\mathrm{Aut}_{\mathbb{R}\text{-}\mathsf{vec}}(V)$ is the group of invertible (i.e. bijective) $\mathbb{R}$-linear maps $f : V \to V$ (since its inverse map is also necessarily $\mathbb{R}$-linear). In particular the elements of $\mathrm{Aut}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n)$ are precisely the maps

$$\begin{array}{rccc} f_A : & \mathbb{R}^n & \to & \mathbb{R}^n \\ & \mathbf{x} & \mapsto & A\mathbf{x} \end{array}$$

for $A \in \mathrm{GL}_n(\mathbb{R})$. Furthermore since $f_A \circ f_B = f_{AB}$, the bijection $\mathrm{GL}_n(\mathbb{R}) \to \mathrm{Aut}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n)$ defined by $A \mapsto f_A$ is an isomorphism of groups.

EXAMPLE 5.14 Let $\mathcal{C}$ be the category defined in Example 5.7. If $f \in \mathrm{Hom}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n, \mathbb{R}^n)$, then the morphism $[f] \in \mathrm{Hom}_{\mathcal{C}}(\mathbb{R}^n, \mathbb{R}^n)$ has an inverse in $\mathcal{C}$ if and only if $[f] \circ [g] = [\mathrm{id}_{\mathbb{R}^n}] = [g] \circ [f]$ for some $g \in \mathrm{Hom}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n, \mathbb{R}^n)$. This certainly holds if f is itself is invertible (take g to be the inverse of f). On the other hand if $[g] \circ [f] = [\mathrm{id}_{\mathbb{R}^n}]$, then $rg \circ f = \mathrm{id}_{\mathbb{R}^n}$ for some $r \in \mathbb{R}^\times$, so f is invertible. Therefore the function

$$\begin{array}{ccc} \mathrm{Aut}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n) & \to & \mathrm{Aut}_{\mathcal{C}}(\mathbb{R}^n) \\ f & \mapsto & [f] \end{array}$$

is surjective, and also clearly a homomorphism. In view of Example 5.13, we thus have a surjective homomorphism $\mathrm{GL}_n(\mathbb{R}) \to \mathrm{Aut}_{\mathcal{C}}(\mathbb{R}^n)$ defined by $A \mapsto [f_A]$. Its kernel is the set of matrices of the form rI_n for $r \in \mathbb{R}^\times$, i.e. the subgroup denoted Z in Example 2.19. Therefore $\mathrm{Aut}_{\mathcal{C}}(R^n)$ is isomorphic to the quotient group $\mathrm{PGL}_n(\mathbb{R}) = \mathrm{GL}_n(\mathbb{R})/Z$.

EXAMPLE 5.15 Let G be a group and let A be the object of the category $\mathcal{C}_G$ (as defined in Example 5.6). Then $\mathrm{Aut}_{\mathcal{C}_G}(A) = G$, as a group.

REMARK 5.16 This new language allows us to give a new perspective on group actions. From the data of an action

$$G \times X \to X,$$

we obtain for each $g \in G$ a map of sets $\varphi_g : X \to X$ defined by $x \xmapsto{\varphi_g} g \cdot x$. It is easy to check that each φ_g is bijective. So this assignment of group elements to maps $X \to X$ in

fact defines a map

$$\Phi\colon G \longrightarrow \mathrm{Aut}_{\mathsf{Sets}}(X)\ (= S_X)$$
$$g \longmapsto (\varphi_g : X \to X).$$

We claim that this map is in fact a group homomorphism (where $\mathrm{Aut}_{\mathsf{Sets}}(X)$ is considered as a group as in Proposition 5.10). To do so we must show that $\Phi(gh) := \varphi_{gh} = \varphi_g \circ \varphi_h =: \Phi(g) \circ \Phi(h)$. But for any $x \in X$,

$$\varphi_{gh}(x) = (gh) \cdot x = g \cdot (h \cdot x) = \varphi_g(\varphi_h(x)),$$

as desired.

This perspective generalises to any category!

DEFINITION 5.17 Let G be a group and let A be an object of a category $\mathcal{C}$. An *action* of G on A (as an object of $\mathcal{C}$) is a homomorphism

$$f : G \to \mathrm{Aut}_{\mathcal{C}}(A).$$

Note this returns the original definition when $\mathcal{C} = \mathsf{Sets}$.

EXAMPLE 5.18 We can strengthen Example 4.11. Let G be any group, and define

$$f : G \to \mathrm{Aut}_{\mathsf{Gps}}(G)$$

by $f(g) = \varphi_g$, where $\varphi_g : G \to G$ is conjugation by g, i.e. $\varphi_g(h) = ghg^{-1}$. Thus conjugation defines an action of G on G as a group, i.e. as an object of Gps, in the sense of Definition 5.17, not just an action on G as a set.

EXAMPLE 5.19 Revisiting Example 4.4, the isomorphism

$$\mathrm{GL}_n(\mathbb{R}) \xrightarrow{\sim} \mathrm{Aut}_{\mathbb{R}\text{-}\mathsf{vec}}(\mathbb{R}^n)$$

defines an action of $\mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$ as a vector space over $\mathbb{R}$ (not just as a set).

If G is any group, then giving an action of G on $\mathbb{R}^n$ (as a real vector space) is equivalent (via composition with the above isomorphism) to giving a homomorphism $f : G \longrightarrow \mathrm{GL}_n(\mathbb{R})$. Group actions on vector spaces are called *representations* and are a very important area of study.

EXAMPLE 5.20 If $\mathcal{C}$ is the category defined in Example 5.7, then the isomorphism

$$\mathrm{PGL}_n(\mathbb{R}) \xrightarrow{\sim} \mathrm{Aut}_{\mathcal{C}}(\mathbb{R}^n)$$

in Example 5.14 gives an action $\mathrm{PGL}_n(\mathbb{R})$ on $\mathbb{R}^n$ as an object of $\mathcal{C}$, and precomposing with the homomorphism $\mathrm{GL}_n(\mathbb{R}) \to \mathrm{PGL}_n(\mathbb{R})$ defines an action of $\mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$ as an object of $\mathcal{C}$.

EXAMPLE 5.21 For any object A of any category $\mathcal{C}$, there is an action of the group $G = \mathrm{Aut}_{\mathcal{C}}(A)$ on A (as an object of $\mathcal{C}$).

EXAMPLE 5.22 Given any category $\mathcal{C}$, define the category $\mathcal{C}^{\mathrm{op}}$ as follows: its objects are the same as those for $\mathcal{C}$, but for any objects A and B, we let

$$\mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(A, B) = \mathrm{Hom}_{\mathcal{C}}(B, A).$$

So for each morphism $B \xrightarrow{f} A$ in $\mathcal{C}$, we get a morphism $A \xrightarrow{f^{\mathrm{op}}} B$ in $\mathcal{C}^{\mathrm{op}}$, where $f^{\mathrm{op}} = f$, but we write $^{\mathrm{op}}$ to indicate when we are viewing it as a morphism in $\mathcal{C}^{\mathrm{op}}$. Given morphisms $A \xrightarrow{f^{\mathrm{op}}} B$ and $B \xrightarrow{g^{\mathrm{op}}} C$ in $\mathcal{C}^{\mathrm{op}}$, we have $B \xrightarrow{f} A$ and $C \xrightarrow{g} B$ in $\mathcal{C}$. We therefore have their composite $C \xrightarrow{f \circ g} A$ in $\mathcal{C}$, and we define composition in the category $\mathcal{C}^{\mathrm{op}}$ (denoting it $\circ^{\mathrm{op}}$)

by $g^{\mathrm{op}} \circ^{\mathrm{op}} f^{\mathrm{op}} = (f \circ g)^{\mathrm{op}}$, i.e. $C \xrightarrow{g} B \xrightarrow{f} A$ in $\mathcal{C}$ defines $C \xleftarrow{g^{\mathrm{op}}} B \xleftarrow{f^{\mathrm{op}}} A$ in $\mathcal{C}^{\mathrm{op}}$. We can thus think of the process of getting $\mathcal{C}^{\mathrm{op}}$ from $\mathcal{C}$ as "reversing all arrows."

For the conditions in Definition 5.1, note that (i) for $\mathcal{C}^{\mathrm{op}}$ is the same as (i) for $\mathcal{C}$, (ii) is satisfied by $\mathrm{id}_A^{\mathrm{op}}$, and (iii) holds since given any morphisms $A \xrightarrow{f^{\mathrm{op}}} B$, $B \xrightarrow{g^{\mathrm{op}}} C$ and $C \xrightarrow{h^{\mathrm{op}}} D$ in $\mathcal{C}^{\mathrm{op}}$, we have

$$\begin{aligned} h^{\mathrm{op}} \circ^{\mathrm{op}} (g^{\mathrm{op}} \circ^{\mathrm{op}} f^{\mathrm{op}}) &= h^{\mathrm{op}} \circ^{\mathrm{op}} (f \circ g)^{\mathrm{op}} = ((f \circ g) \circ h)^{\mathrm{op}} \\ &= (f \circ (g \circ h))^{\mathrm{op}} = (g \circ h)^{\mathrm{op}} \circ^{\mathrm{op}} f^{\mathrm{op}} = (h^{\mathrm{op}} \circ^{\mathrm{op}} g^{\mathrm{op}}) \circ f^{\mathrm{op}}. \end{aligned}$$

6 Free groups and presentations

We now introduce the notion of a free group; roughly speaking it's the largest group that can be generated by a particular set of elements.

DEFINITION 6.1 Let S be any set, we are going to construct the free group "generated by S". First let S^+ and S^- also be two copies of S. So given an element $s \in S$, we obtain an element "s"of S^+. Of course, given an element of S, we also obtain an element of S^-, but we use s^{-1} to denote this element. For example, if $S = \{x, y\}$, then

$$S^+ = \{x, y\}, \quad S^- = \{x^{-1}, y^{-1}\}.$$

For reasons that will become clear below, we say that $s^{-1} \in S^-$ is the inverse of $s \in S^+$ and vice versa. We refer to $S^{\pm} := S^+ \cup S^-$ as *letters*.

A *reduced word* on S is a sequence $(t_1, t_2, \ldots, t_n)$ of any (finite) length $n \geq 0$ such that

- $t_i \in S^{\pm}$ for $i = 1, \ldots, n$;
- t_{i+1} is not the inverse of t_i for $i = 1, \ldots, n-1$.

EXAMPLE 6.2 Note that n can be 0, the only reduced word of length 0 being the empty word ( ). Furthermore S can be empty, in which case the only reduced word on S is the empty word.

EXAMPLE 6.3 Suppose that $S = \{x, y\}$ has exactly two elements.

- There is only one reduced word of length 0, namely ( );
- there are 4 reduced words of length 1, corresponding to the 4 elements of $S^{\pm}$: (x), (x^-), $(y)$and (y^-);
- 12 of length 2, given by

$$\begin{array}{cccc} (x,x), & \cancel{(x,x^-)}, & (x,y), & (x,y^-), \\ \cancel{(x^-,x)}, & (x^-,x^-), & (x^-,y), & (x^-,y^-), \\ (y,x), & (y,x^-), & (y,y), & \cancel{(y,y^-)}, \\ (y^-,x), & (y^-,x^-), & \cancel{(y^-,y)}, & (y^-,y^-), \end{array}$$

  the ones crossed out failing to be reduced[15];
- $36 = 12 \cdot 3$ of length 3, which we won't list, and similarly $108 = 36 \cdot 3$ of length 4, etc.

Note that in the example, S was a finite set, but it needn't be in general.

[15]The word *reduced* refers to the second condition in the definition; a sequence which doesn't necessarily satisfy this may be called a *word*. I'll sometimes just say "word" to mean "reduced word" implicitly in lectures, but I'll be more careful in the typed notes.

We define F_S to be the set of all reduced words on S, and we define a binary operation $*$ on F_S by the formula

$$(t_1, t_2, \ldots, t_n) * (t'_1, t'_2, \ldots, t'_{n'}) = (t_1, t_2, \ldots, t_{n-k}, t'_{k+1}, \ldots, t_{n'}),$$

where k is the least integer $0 \leq i \leq \min\{n, n'\}$ such that t'_{i+1} is not the inverse of t_{n-i}. Often, we will have $k = 0$. This is when the "word" given by concatenation is a reduced word, but otherwise we "cancel" neighbouring terms until we obtain a reduced word. We could even have $k = n$ or n'. For example,

$$(t_1, t_2, ..., t_n) * (t_n^{-1}, t_{n-1}^{-1} ..., t_1^{-1}) = ().$$

EXAMPLE 6.4 If $S = \{x, y\}$ as in Example 6.3, then

$$(x, y, x^-, x^-, y) * (y^-, x, y) = (x, y, x^-, \cancel{x^-, y, y^-, x}, y) = (x, y, x^-, y).$$

(In this case we had $k = 2$ in the formula defining the operation.)

We leave it as an exercise to check that the operation $*$ on F_S is associative. We clearly have that the empty word ( ) is an identity element, and we have seen that the inverse of $(t_1, t_2, \ldots, t_n)$ is $(t_n^{-1}, t_{n-1}^{-1} \ldots, t_1^{-1})$, so F_S is a group under this operation.

DEFINITION 6.5 The group $(F_S, *)$ is called the *free group* on S.

For $s \in S$, we usually just write s for the word (s) of length 1 and s^{-1} for the word (s^{-1}). Similarly, we can extend this to write arbitrary elements of F_S in the form

$$s_1^{\epsilon_1} s_2^{\epsilon_2} \cdots s_n^{\epsilon_n},$$

where $\epsilon \in \{\pm 1\}$. We can take this one step further by additionally writing s^n for the reduced word $\underbrace{ss...s}_{n \text{ times}}$ and similarly for s^{-n}.

As usual we also omit the symbol $*$ for the operation, so we could write the equation in Example 6.4 as

$$(xyx^{-2}y)(y^{-1}xy) = xyx^{-1}y.$$

EXAMPLE 6.6 If $S = \emptyset$, then $F_S = \{e\}$, where $e = ( )$.

EXAMPLE 6.7 If $S = \{s\}$, then there are just two reduced words of each length $n > 0$, namely $(s, \ldots, s) = s^n$ and $(s^-, \ldots, s^-) = s^{-n}$. Thus F_S is generated by s, and we have an isomorphism $\mathbb{Z} \xrightarrow{\sim} F_S$ defined by $n \mapsto s^n$.

EXAMPLE 6.8 If $S = \{x, y\}$, then we have $xy \neq yx$ (or written out more fully, $(x)*(y) = (x, y) \neq (y, x) = (y) * (x)$), so F_S is not abelian. The same argument shows that if $\#S > 1$, then F_S is not abelian. (Recall also that S could be infinite.)

There is a canonical map of sets $i_S : S \to F_S$ given by $s \mapsto (s)$. The following property can be viewed as characterizing a free group.

PROPOSITION 6.9 *Let S be a set, G a group and $f : S \to G$ a map of sets. Then there is a unique homomorphism of groups $\widetilde{f} : F_S \to G$ such that the following diagram commutes (as maps of sets):*

(1)
$$\begin{array}{ccc} S & \xrightarrow{f} & G \\ \downarrow{\scriptstyle i_S} & \nearrow_{\exists!\ \widetilde{f}} & \\ F_S & & \end{array}$$

given by $\widetilde{f}((s)) = f(s)$ for all $s \in S$, i.e. identifying each $s \in S$ with the corresponding element $(s) \in F_S$, the function $f : S \to G$ extends uniquely to a homomorphism $F_S \to G$.

Proof. By the definition of F_S, each element can be written uniquely in the form $g = t_1 t_2 \cdots t_n$, where $n \geq 0$, each $t_i = s_i^{\epsilon_i}$ for some $s_i \in S$, $\epsilon_i \in \{\pm 1\}$, and $t_{i+1} \neq t_i^{-1}$ for $i = 1, \ldots, n-1$. We define $\widetilde{f} : F_S \to G$ by

$$\widetilde{f}(g) = f(s_1)^{\epsilon_1} f(s_2)^{\epsilon_2} \cdots f(s_n)^{\epsilon_n}$$

for g as above.[16]

Then $\widetilde{f}$ clearly has the property that $\widetilde{f}(s) = f(s)$ for all $s \in S$. Furthermore it is the only possible function $F_S \to G$ that can be a homomorphism, since we must have $\widetilde{f}(s_i^{\pm 1}) = \widetilde{f}(s_i)^{\pm 1} = f(s_i)^{\pm 1}$.

It remains to prove that $\widetilde{f}$ is a homomorphism, i.e. that $\widetilde{f}(gg') = \widetilde{f}(g)\widetilde{f}(g')$ for all $g, g' \in F_S$. Writing $g = t_1 t_2 \cdots t_n$ and $g' = t_1' t_2' \cdots t_{n'}'$ as above, we have

$$gg' = t_1 \cdots t_{n-k} t_{k+1}' \cdots t_{n'}',$$

where $k \geq 0$ is such that $t_i' = t_{n-i+1}^{-1}$, for $i = 1, \ldots, k$ and $t_{k+1}' \neq t_{n-k}^{-1}$. The definition of $\widetilde{f}$ gives

$$\begin{aligned} \widetilde{f}(g) &= f(s_1)^{\epsilon_1} \cdots f(s_{n-k})^{\epsilon_{n-k}} f(s_{n-k+1})^{\epsilon_{n-k+1}} \cdots f(s_n)^{\epsilon_n} \\ \text{and} \quad \widetilde{f}(g') &= f(s_1')^{\epsilon_1'} \cdots f(s_k')^{\epsilon_k'} f(s_{k+1}')^{\epsilon_{k+1}'} \cdots f(s_{n'}')^{\epsilon_{n'}'}, \end{aligned}$$

but $s_1' = s_n$ and $\epsilon_1' = -\epsilon_n$, so $f(s_n)^{\epsilon_n} f(s_1')^{\epsilon_1'} = e$, and similarly $f(s_{n-1})^{\epsilon_{n-1}} f(s_2')^{\epsilon_2'} = \cdots = f(s_{n-k+1})^{\epsilon_{n-k+1}} f(s_k')^{\epsilon_k'} = e$. Therefore

$$\widetilde{f}(g)\widetilde{f}(g') = f(s_1)^{\epsilon_1} \cdots f(s_{n-k})^{\epsilon_{n-k}} f(s_{k+1}')^{\epsilon_{k+1}'} \cdots f(s_{n'}')^{\epsilon_{n'}'} = \widetilde{f}(gg'),$$

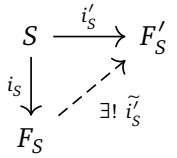
as required. □

EXAMPLE 6.10 Let $S = \{x, y\}$. There is a unique homomorphism $\widetilde{f} : F_S \to F_S$ such that $\widetilde{f}(x) = y$ and $\widetilde{f}(y) = x$. Furthermore it is an automorphism since $\widetilde{f} \circ \widetilde{f}$ is the unique homomorphism $F_S \to F_S$ such that $x \mapsto x$ and $y \mapsto y$. Since id_{F_S} also has this property, $\widetilde{f} \circ \widetilde{f} = \mathrm{id}_{F_S}$, i.e. $\widetilde{f}$ is its own inverse.

REMARK 6.11 The property (1) of Proposition 6.9 is an example of a *universal property*, with this being the *universal property of a free group*. We won't explain what exactly the phrase universal property means, but we will explain why it "characterises free groups".

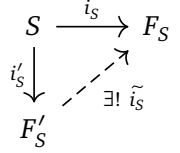
Suppose we had another pair (F_S', i_S') that satisfies the property given in Proposition 6.9, i.e. F_S' is a group and $i_S' : S \to F_S$ is a map of sets (we don't assume injective, but this will be forced by the following) such that any map of sets $f : S \to G$ extends uniquely to a map of groups $i_S' : F_S' \to G$ (that is such that $f = \tilde{f} \circ i_S'$).

Now $i_S' : S \to F_S$ is a perfectly good map of sets to apply the universal property of F_S to. That is, we can take $f : S \to G$ to be $i_S' : S \to F_S'$ in diagram (1):

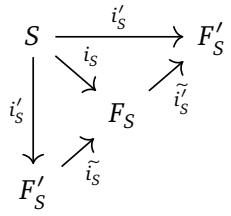


[16] Recall the convention that $g = e_{F_S}$ if $n = 0$; similarly the expression for $\widetilde{f}(g)$ is interpreted as meaning e_G in this case.

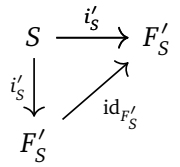
to obtain a unique group homomorphism $\widetilde{i'_S}$ making the diagram commute. On the other hand, we are assuming $i'_S : S \to F'_S$ also has the ability to extend any map of sets. So we can apply the same argument with the roles swapped to obtain a reverse group homomorphism $\widetilde{i_S} : F'_S \to F_S$:



We claim that $\widetilde{i'_S}, \widetilde{i_S}$ define mutually inverse group homomorphisms, so that in particular F_S and F'_S are isomorphic. To show this is to show that $\widetilde{i'_S} \circ \widetilde{i_S} = \mathrm{id}_{F'_S}$ and $\widetilde{i_S} \circ \widetilde{i'_S} = \mathrm{id}_{F_S}$. This once again follows from the universal property! From the above two diagrams, we know that the following diagram commutes:



Focus only on the outer triangle, we already know one map $F'_S \to F'_S$ that makes this triangle commute, namely $\mathrm{id}_{F'_S} : F'_S \to F'_S$:



But now look again at the universal property, it says that there is a unique map making this triangle commute, so both $\mathrm{id}_{F'_S}$ and $\widetilde{i'_S} \circ \widetilde{i_S}$ must be equal! The same is of course true for id_{F_S} and $\widetilde{i_S} \circ \widetilde{i'_S}$ and they really are inverse isomorphisms.

We have proven that (F'_S, i'_S) is as close as possible to (F_S, i_S) (we can't say they are equal). Objects satisfying universal properties are "unique up to unique isomorphism". This is the closest possible thing to them being equal. Note that this relies on the data of the group and the map from S. Note also, how this would have failed without the existence and the uniqueness requirements of the universal property.

This argument can be repeated verbatim for other universal properties to show "uniqueness" of such objects. Note however, this does nothing to show that an object exists that actually satisfies the universal property. We cannot use the universal property to define free groups until we know they exist.

EXERCISE 6.12 Let $\mathbb{Z}_S$ denote the free abelian group with basis S. So if $S = \{x\}$ then $\mathbb{Z}_S = \mathbb{Z}$ with generator labelled by x. Then $\mathbb{Z}_S$ satisfies Proposition 6.9 with every instance of "group" replaced by "abelian group". Try and prove as much of this and the analogous version of Remark 6.11 as you can.

We leave the proof of the following as an exercise:

LEMMA 6.13 *Let S be a set, G a group and $f : S \to G$ a function. Then the homomorphism $\widetilde{f} : F_S \to G$ extending f is surjective if and only if $G = \langle f(S) \rangle$ (i.e. G is generated by the image of S).*

If R is a subset of F_S, then we write $\langle S \,|\, R \rangle$ for the quotient group F_S/N, where N is the subgroup of F_S generated by the set of conjugates of elements of R, i.e. $N = \langle T \rangle$, where $T = \{ grg^{-1} \,|\, g \in F_S, r \in R \}$. Here we have implicitly used that N is a normal subgroup; in fact the following lemma says that N is the smallest normal subgroup containing R (sometimes referred to as the subgroup *normally generated* by R or *normal closure* of R):

LEMMA 6.14 *Suppose that R is a subset of a group G. Let $T = \{ grg^{-1} \,|\, g \in G, r \in R \}$, and let $N = \langle T \rangle$. Then $R \subset N$ and $N \trianglelefteq G$; furthermore if N' is any normal subgroup of G such that $R \subset N'$, then $N \subset N'$.*

Proof. We clearly have $R \subset T \subset \langle T \rangle = N$.

To see that N is normal in G, note that $h^{-1}Th = T$ for all $h \in G$ (where $h^{-1}Th$ denotes $\{ h^{-1}th \,|\, t \in T \}$). Since $T = h^{-1}Th \subset h^{-1}Nh$, it follows that $N = \langle T \rangle \subset h^{-1}Nh$, and hence that $hNh^{-1} \subset N$ for all $h \in G$.

Suppose now that N' is a normal subgroup of G and $R \subset N'$. Since $N' \trianglelefteq G$, we have $grg^{-1} \in N'$ for all $g \in G$, $r \in R$. Therefore $T \subset N'$, and hence $N = \langle T \rangle \subset N'$. □

DEFINITION 6.15 If G is a group, then a *presentation* of G is an isomorphism $\langle S \,|\, R \rangle \xrightarrow{\sim} G$ (where S is a set, and R is a subset of F_S).

Note that if $F_S/N = \langle S \,|\, R \rangle \xrightarrow{\sim} G$ is a presentation of G, then the composite $F_S \to F_S/N \to G$ is surjective (where the first homomorphism is $g \mapsto gN$). Therefore by Lemma 6.13, G is generated by the image of S; we can think of the presentation as describing G in terms of a set of "generators" (corresponding to elements of S) and "relations" among them (corresponding to elements of R).

EXAMPLE 6.16 Let G be any group, let $S = G$, and let $f : S \to G$ be the identity function. By Proposition 6.9 this extends to a homomorphism $\widetilde{f} : F_S \to G$, which is clearly surjective. Letting $R = N = \ker(\widetilde{f})$, we have that N is the smallest normal subgroup of F_S contained $R = N$, so by the First Isomorphism Theorem, we have $\langle S \,|\, R \rangle = F_S/N \xrightarrow{\sim} G$.

The preceding example shows that every group has at least one presentation, but the one we just gave was constructed very brutally. Presentations with fewer generators and relations provide more practical descriptions of the group.

EXAMPLE 6.17 Let $G = \mathbb{Z}/n\mathbb{Z}$. Recall from Example 6.7 that $F_{\{s\}}$ is isomorphic to $\mathbb{Z}$ (via $s^m \longleftrightarrow m$ for $m \in \mathbb{Z}$). We therefore have the surjection $F_{\{s\}} \to G$, with kernel $\langle s^n \rangle$. This gives a presentation $\langle S \,|\, R \rangle \xrightarrow{\sim} G$ with $S = \{s\}$ and $R = \{s^n\}$. When the sets S and R are finite, we usually just list the elements instead of writing $\langle S \,|\, R \rangle$, so for example, we write $\langle s \,|\, s^n \rangle$ instead of $\langle \{s\} \,|\, \{s^n\} \rangle$.

EXAMPLE 6.18 The dihedral group has a presentation

$$\langle x, y \,|\, x^n, y^2, (xy)^2 \rangle \xrightarrow{\sim} D_n,$$

sending x to a rotation of order n and y to a reflection. The proof that there is such a presentation is left as as exercise.

EXERCISE 6.19 Suppose that $\langle S \mid R \rangle$ is a group defined by a presentation (i.e. the group F_S/N). Using the first isomorphism theorem, extend Exercise 3.10 to show that a group homomorphism $\langle S \mid R \rangle \to H$ for an arbitrary group H is equivalent to the data of a map of sets $\varphi : S \to H$ such that $R \subseteq \ker(\widetilde{\varphi})$. This setup will occur often as we introduce other constructions and we will often use this to define homomorphisms without comment.

DEFINITION 6.20 We say that the group G is *finitely presented* if there is a presentation $\langle S \mid R \rangle \xrightarrow{\sim} G$ such that the sets S and R are finite.

Every finite group is finitely presented, but we do not prove this[17]. Of course an infinite group may also be finitely presented; for example the free group F_S for any finite set S has a presentation $\langle S \mid \emptyset \rangle = F_S$, but F_S is infinite (as long as S is non-empty). If G is finitely presented, then it is obviously finitely generated (i.e. $G = \langle Q \rangle$ for some finite subset Q of G). What is more difficult to prove (and we won't) is that there are finitely generated groups which are *not* finitely presented.

[17] This isn't super hard, but requires some care. Note, the set R in Example 6.16 will never be finite, to conclude one needs to explain why the R is itself finitely generated.

Part 2

Commutative algebra

We will start with basic concepts and results concerning (commutative) rings and modules, and then focus on the topics of tensor products and localisation, introducing more of the language of category theory along the way.

7 Ring theory review

DEFINITION 7.1 A *ring* is a set R equipped with a pair of binary operations $+$ (called *addition*) and $\cdot$ (*multiplication*) satisfying:

(i) $(R,+)$ is an abelian group;

(ii) $\cdot$ is associative and has a (multiplicative) identity element[18], denoted 1_R, i.e. $r\cdot(s\cdot t)=(r\cdot s)\cdot t$ for all $r,s,t\in R$, and there is an element $1_R\in R$ such that $1_R\cdot r=r=r\cdot 1_R$ for all $r\in R$;

(iii) $\cdot$ is left and right distributive over $+$, i.e. $r\cdot(s+t)=r\cdot s+r\cdot t$ and $(r+s)\cdot t=r\cdot t+s\cdot t$ for all $r,s,t\in R$.

REMARK 7.2 Since $(R,+)$ is a group, it has a unique identity for $+$ (Exercise 1.6), which we denote by 0_R or 0 for short. Exactly, the same argument shows that even though R is not a group under $\cdot$, its multiplicative identity 1_R is unique. We usually denote it simply by 1.

EXAMPLE 7.3 The familiar number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ with the usual notions of addition and multiplication are all rings.

EXAMPLE 7.4 If n is a positive integer, then the set $\mathbb{Z}/n\mathbb{Z}$, with the operations of addition and multiplication mod n, i.e. $\overline{a}+\overline{b}=\overline{a+b}$ and $\overline{a}\cdot\overline{b}=\overline{a\cdot b}$, is a ring (where $\overline{a}$ denotes $a+n\mathbb{Z}$ for $a\in\mathbb{Z}$).

EXAMPLE 7.5 The set $\{0\}$, with the (only possible) operations $0+0=0$ and $0\cdot 0=0$, is a ring. Note that in this case $0_R=0=1_R$.

The rings above are all *commutative*, meaning that the operation $\cdot$ is commutative[19] (i.e. $r\cdot s=s\cdot r$ for all $r,s\in R$). We will focus on rings that are commutative, but here's. one that isn't:

EXAMPLE 7.6 Let $M_n(\mathbb{R})$ denote the set of $n\times n$ matrices over $\mathbb{R}$ with the usual notions of matrix addition and multiplication. Then $M_n(\mathbb{R})$ is a ring, non-commutative if $n>1$.

Starting with some rings, there are general constructions that produce more.

EXAMPLE 7.7 If R and R' are rings, then the set $R\times R'$, with componentwise addition and multiplication, is a ring. By "componentwise," we mean

$$(r,r')+(s,s')=(r+s,r'+s') \quad \text{and} \quad (r,r')\cdot(s,s')=(rs,r's')$$

for $r,s\in R$, $r',s'\in R'$, where $+$ is $r+s$ refers to the addition operation on R, etc. Note for example that the multiplicative identity on $R\times R'$ is $(1_R,1_{R'})$.

EXAMPLE 7.8 The set of polynomials with real coefficients, denoted $\mathbb{R}[X]$, is a ring under the usual addition and multiplication operations on polynomials.

More generally for any ring R, we similarly define $R[X]$ to be the *polynomial ring* over R in the variable (or indeterminate) X. We can replace X by any (formal) variable,

[18] Some authors do not insist on the existence of a multiplicative identity. The default though is that they do have one.

[19] Note that if $\cdot$ is commutative, then the left and right distributive axioms (the two equations in 3) for all $r,s,t\in R$) are equivalent.

and so for example define $R[X,Y] = (R[X])[Y]$ to be the polynomial ring over $R[X]$ in the variable Y, and similarly inductively define $R[X_1,\dots,X_n]$, the polynomial ring over R in n variables (for any $n \geq 1$).

Note that R need not be commutative for this to be defined, but that if R is commutative, so is $R[X]$ (and so by induction $R[X_1,\dots,X_n]$).

EXAMPLE 7.9 For any ring R and integer $n \geq 1$, we can define the $n \times n$ matrix ring over R, denoted $M_n(R)$, with the usual operations of matrix addition and multiplication. Again R need not be commutative, but in this case $M_n(R)$ will not be commutative even if R is (unless $n = 1$ or $R = \{0\}$).

For any ring R, we have $0_R \cdot r = 0_R = r \cdot 0_R$ for all $r \in R$. To see the first equality for example, note that it follows from the axioms that

$$(0_R \cdot r) + (1_R \cdot r) = (0_R + 1_R) \cdot r = 1_R \cdot r = 0_R + (1_R \cdot r),$$

so the cancellation law (for the abelian group $(R,+)$) implies that $0_R \cdot r = 0_R$. Similarly we have $(-1_R) \cdot r = -r = r \cdot (-1_R)$ for all $r \in R$ (where as usual, we use $-$ to denote the additive inverse of an element). More generally, it follows by induction that

$$n_R \cdot r = nr = r \cdot n_R \qquad \text{for all } r \in R,\ n \in \mathbb{Z},$$

where nr has its usual meaning for the abelian group $(R,+)$ (for example $3r = r+r+r$) and n_R denotes $n(1_R)$.

EXERCISE 7.10 Show that any ring for which $1_R = 0_R$ has only one element (so morally is the ring in Example 7.5).

DEFINITION 7.11 An element $r \in R$ is called a *unit* in R if it has a multiplicative inverse, i.e. $r \cdot s = 1_R = s \cdot r$ for some $s \in R$. We let $R^\times$ denote the set of units in R. It is a group under $\cdot$. This follows from the facts that 1) $1_R \in R^\times$, 2) if $r, r' \in R^\times$, then $r \cdot r' \in R^\times$, and 3) if $r \in R^\times$, then its inverse $r^{-1} \in R^\times$.

EXAMPLE 7.12 Every non-zero real number has a (real) multiplicative inverse, so $\mathbb{R}^\times = \mathbb{R} \smallsetminus \{0\}$.

EXAMPLE 7.13 The only integers with multiplicative inverses in $\mathbb{Z}$ are ± 1, so $\mathbb{Z}^\times = \{\pm 1\}$.

EXAMPLE 7.14 We have $\{0\}^\times = \{0\}$.

EXAMPLE 7.15 If n is a positive integer, then $(\mathbb{Z}/n\mathbb{Z})^\times$ is the set of residue classes $\overline{a}$ such that $\gcd(a,n) = 1$.

EXAMPLE 7.16 For and (non-zero) $f, g \in \mathbb{R}[X]$, we have $\deg(fg) = \deg(f) + \deg(g)$, so if $fg = 1_{\mathbb{R}[X]}$, then $\deg(f) = \deg(g) = 0$. It follows that $(\mathbb{R}[X])^\times = \mathbb{R}^\times$ (where we identify elements of $\mathbb{R}$ with constant polynomials).

EXAMPLE 7.17 Since a matrix in $M_n(\mathbb{R})$ is invertible if and only if it has non-zero determinant, we have $M_n(\mathbb{R})^\times = \mathrm{GL}_n(\mathbb{R})$.

EXAMPLE 7.18 If R and S are rings, then $(R \times S)^\times = R^\times \times S^\times$.

EXERCISE* 7.19 Recall that Ab denotes the category of abelian groups. Let $A \in \mathrm{Ob}(\mathsf{Ab})$ be any abelian group. Prove that $\mathrm{End}_{\mathsf{Ab}}(A) := \mathrm{Hom}_{\mathsf{Ab}}(A,A)$ a canonical ring structure (in particular, what is multiplication and addition?). So rings arise naturally from the concept of abelian groups (cf. Proposition 5.10)!

From now on, **we assume R is a commutative ring unless otherwise stated**.

DEFINITION 7.20 Let $r \in R \setminus \{0\}$, if there exists $s \in R \setminus \{0\}$ such that $rs = 0$, we say that r (and so also s) is a *zero divisor*. We say R is an *integral domain* (or just a *domain*) if $0_R \neq 1_R$ (i.e. the ring is not the zero ring of Example 7.5) and R has no zero divisors, i.e. for every $r, s \in R$ such that $r \neq 0_R$ and $s \neq 0_R$, we have $r \cdot s \neq 0_R$.

EXAMPLE 7.21 The ring $\mathbb{Z}$ is a domain.

EXAMPLE 7.22 If $n > 0$, then the ring $\mathbb{Z}/n\mathbb{Z}$ is domain if and only if n is prime.

EXERCISE 7.23 Prove that for a domain we have the following cancellation law: if $r, r' \in R$ and $s \neq 0_R$, then $rs = r's \iff r = r'$. Show that this need not hold for general rings R.

DEFINITION 7.24 We say R is a *field* if $R^\times = R \setminus \{0_R\}$, or equivalently if $R \neq \{0_R\}$ and every non-zero element of R is a unit.

EXAMPLE 7.25 The rings $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields, but $\mathbb{Z}$ is not. The ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime.

DEFINITION 7.26 We say a subset $R' \subset R$ is a *subring* of R if the following hold:

(i) $1_R \in R'$;
(ii) R' is a subgroup of R under $+$;
(iii) R' is closed under $\cdot$, i.e. if $r, s \in R'$, then $r \cdot s \in R'$.

Note that if R' is a subring of R, then it is a ring under the same operations as the ones on R (restricted to R').

EXAMPLE 7.27 The subsets

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{R}[X] \subset \mathbb{R}[X, Y]$$

are all subrings of the ones to their right (where we identify real numbers with constant polynomials to make sense of the inclusion $\mathbb{R} \subset \mathbb{R}[X]$).

Note that $\{0\}$ is *not* a subring of $\mathbb{Z}$. In fact $\{0_R\}$ is never a subring of R (unless $R = \{0_R\}$); it satisfies conditions (ii) and (iii), but not (i).

DEFINITION 7.28 Suppose that R and S are rings. We say a function $f : R \to S$ is a *(ring) homomorphism* if the following hold:

(i) $f(1_R) = 1_S$;
(ii) $f(r + r') = f(r) + f(r')$ for all $r, r' \in R$ (i.e. f is a homomorphism of abelian groups under $+$);
(iii) $f(r \cdot r') = f(r) \cdot f(r')$ for all $r, r' \in R$.

If $f : R \to S$ is a homomorphism of rings, then we say f is an *isomorphism (of rings)* if it bijective.

Note that in (ii) and (iii), the operations $+$ and $\cdot$ on the left-hand side of the equations are the ones on R, and those on the right are on S.

EXAMPLE 7.29 For any ring R, the function $\mathbb{Z} \to R$ defined by $n \mapsto n_R$ is a homomorphism, where recall that $n_R = 1_R + \cdots + 1_R$ (n times) if $n > 0$ and $n_R = -(-n)_R$ if $n < 0$. This follows from the Laws of Exponents for powers of 1_R of the group $(R, +)$ (except the "powers" are multiples since the operation is $+$). Note this map need not be injective.

EXAMPLE 7.30 If R' is a subring of R, then $\iota : R' \hookrightarrow R$ (defined by $\iota(r) = r$) is a homomorphism.

EXAMPLE 7.31 Complex conjugation on $\mathbb{C}$, i.e. $f : \mathbb{C} \to \mathbb{C}$ defined by $f(z) = \overline{z}$, is a homomorphism.

EXAMPLE 7.32 The identity function on any ring R is a homomorphism, and if $f : R \to S$ and $f : S \to T$ are homomorphisms, then so is $g \circ f : R \to T$. Therefore commutative rings form a category, which we denote CRings. All the other properties of a category follow from facts about composition of functions of sets.

If f is an isomorphism of rings (in the sense of Definition 7.28), then its inverse (function) is also a ring homomorphism, so the notion of isomorphism in Definition 7.28 is equivalent to the one obtained by viewing Rings as a category.

Non-commutative rings also form a category, denoted Rings.

If $f : R \to S$ is a ring homomorphism, then

$$\mathrm{im}(f) = f(R) = \{ f(r) \mid r \in R \}$$

is a subring of S. Viewing f as a homomorphism of abelian groups, we have

$$\ker(f) = \{ r \in R \mid f(r) = 0_S \},$$

which is *not* a subring of R (unless $S = \{0_S\}$ in which case $\ker(f) = R$), but a different type of subset.

DEFINITION 7.33 We say a subset $I \subset R$ is an *ideal* of R if the following hold:

(i) I is a subgroup of R under $+$;

(ii) $r \cdot a \in I$ for all $r \in R$, $a \in I$.

We use $I \trianglelefteq R$ to denote an ideal I of R.

For any homomorphism $f : R \to S$, we have that $\ker(f)$ is an ideal of R. Indeed we already know (i) by group theory, and for (ii), note that if $r \in R$ and $a \in \ker(f)$, then $f(a) = 0_S$, so

$$f(r \cdot a) = f(r) \cdot f(a) = f(r) \cdot 0_S = 0_S,$$

which means that $r \cdot a \in \ker(f)$.

We will often omit the symbol for multiplication in a ring R from now on, and just write rs instead of $r \cdot s$.

DEFINITION 7.34 If $a \in R$, then the subset

$$\{ ra \mid r \in R \} \subset R$$

is an ideal. Indeed, for (i) $0 = 0a \in Ra$, if $sa, s'a \in Ra$ (where $s, s' \in R$), then $sa + s'a = (s + s')a \in Ra$ and $-(sa) = (-s)a \in Ra$, and (ii) if $r \in R$, $sa \in Ra$ (where $s \in R$), then $r(sa) = (rs)a \in Ra$.

We denote this ideal by Ra or (a) and call it the *principal ideal generated* by a. An ideal I of R is called *principal* if $I = Ra$ for some $a \in R$.

EXAMPLE 7.35 If $R = \mathbb{Z}$ and $n \in \mathbb{Z}$, then the ideal generated by n is the set of integer multiples on n, i.e. the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$.

EXAMPLE 7.36 For any R, we have $R(1_R) = R$ and $R(0_R) = \{0_R\}$.

It is easy to see that if $a, b \in R$, then

$$Rb \subset Ra \iff b \in Ra \iff b = ra \text{ for some } r \in R.$$

When this holds, we say a *divides* b (in R), or that b is a *multiple* of a (in R), and we write $a|b$.

Note that $Ra = R$ if and only if $a|1_R$, i.e. $a \in R^\times$. It follows that if R is a ring whose only non-zero ideal is R, then every non-zero element of R is a unit, and hence R is a field. Conversely if R is a field, then it has precisely two ideals, namely $\{0_R\}$ and R.

Generalizing Definition 7.34, we can consider ideals generated by subsets of R:

DEFINITION 7.37 If $a_1, \ldots, a_n \in R$, then the *(left[20]) ideal of R generated by* $\{a_1, \ldots, a_n,\}$ is

$$(a_1, \ldots, a_n) = \left\{ \sum_{i=1}^{n} r_i a_i \;\middle|\; r_1, \ldots, r_n \in R \right\}.$$

More generally for any subset $A \subset R$, the *ideal of R generated by A* is

$$(A) = \left\{ \sum_{j=1}^{k} s_j b_j \;\middle|\; k \geq 0,\, s_1, \ldots, s_k \in R,\, b_1, \ldots, b_k \in A \right\}$$

(where we view the sum as 0_R if $k = 0$).

Note that the two versions of the definition agree if $A = \{a_1, \ldots, a_n\}$, i.e. $(a_1, \ldots, a_n) = (A)$. (Every sum of the form $\sum_{i=1}^{n} r_i a_i$ is clearly in (A); on the other hand if each $b_j \in A = \{a_1, \ldots, a_n\}$, then we can rewrite $\sum_{j=1}^{k} s_j b_j$ as $\sum_{i=1}^{n} r_i a_i$ where r_i is the sum of the s_j such that $b_j = a_i$.) We sometimes write simply $(a_1, \ldots, a_n)$ instead of $R(a_1, \ldots, a_n)$ or $(\{a_1, \ldots, a_n\})$, so for example the principal ideal generated by a would be denoted (a).

Note that it is clear from the definition that (A) is in fact an ideal. Furthermore it is the smallest ideal containing A in the usual sense: $A \subset (A)$, and if I is any ideal of R such that $A \subset I$,then $(A) \subset I$.

EXAMPLE 7.38 If $a_1, \ldots, a_n \in R = \mathbb{Z}$, then $(a_1, \ldots, a_n) = (d)$ where $d = \gcd(a_1, \ldots, a_n)$ (or 0 if $a_1 = \ldots = a_n = 0$).

EXAMPLE 7.39 If $R = \mathbb{R}[X, Y]$, then (X, Y) is the set of polynomials $f \in R$ with constant term 0, i.e.

$$(X, Y) = \left\{ \sum_{i,j=0}^{k} r_{ij} X^i Y^j \;\middle|\; k \geq 0,\, r_{ij} \in \mathbb{R} \text{ for } i, j = 1, \ldots, k,\, r_{0,0} = 0 \right\}.$$

DEFINITION 7.40 We say that R is a *principal ideal domain* (or a *PID*), if R is a domain and every ideal of R is principal.

EXAMPLE 7.41 The ring $\mathbb{Z}$ is a PID since every ideal is of the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$.

EXAMPLE 7.42 The ring $\mathbb{R}[X]$ is a PID (as a consequence of the division algorithm for polynomials); more generally if K is any field, then $K[X]$ is a PID.

EXAMPLE 7.43 The ring $R = \mathbb{Z}/n\mathbb{Z}$ (for any positive integer n) has the property that every ideal is principal; however it is a domain (and hence a PID) if and only if n is prime.

EXAMPLE 7.44 The ring $R = \mathbb{R}[X, Y]$ is a domain, but the ideal (X, Y) is not principal. Indeed if it were true that $(X, Y) = (f)$ for some $f \in R$, then we would have $f \mid X$ and $f \mid Y$, so f would have to be a non-zero constant polynomial, in which case $f \notin (X, Y)$. Therefore $\mathbb{R}[X, Y]$ is not a PID.

EXERCISE 7.45 Let $f : R \to S$ be a ring homomorphism. Let J be an ideal of S. Prove that the preimage $f^{-1}(J)$ is an ideal of R. Show that this is not, in general, true of the image of ideals.

[20] It is possible to define right ideals by putting the ring on the other side of the generators. We won't ever consider right ideals, so use ideal to refer to left ideals. Of course, if R is commutative, the two concepts coincide.

Recall that if I is an ideal of R, then it is a subgroup of the R under addition. Since R is abelian, I is in fact a normal subgroup, and we can form the quotient group

$$R/I = \{ r + I \mid r \in R \}.$$

Thus R/I is an abelian group under the inherited operation $+$, defined by $\overline{r} + \overline{s} = \overline{r+s}$ (writing $\overline{r}$ for the coset $r + I$ for $r \in R$). Furthermore it is easy that $\overline{r} \cdot \overline{s} = \overline{rs}$ gives a well-defined binary operation on cosets. (If $\overline{r} = \overline{r'}$ and $\overline{s} = \overline{s'}$, then $r - r', s - s' \in I$, so

$$rs - r's' = r(s - s') + s'(r - r') \in I,$$

and therefore $\overline{rs} = \overline{r's'}$.)

We claim that R/I is in fact a commutative ring under the resulting addition and multiplication operations. Indeed the required associativity and distributivity are immediate from the corresponding properties of R, and $\overline{1}_R$ is a multiplicative identity element. The ring R/I (read as "R mod I") is called the *quotient ring* of R by I.

Recall also (Example 3.8) that there is a surjective homomorphism of abelian groups $\pi : R \to R/I$ defined by $\pi(r) = \overline{r}$; since $\overline{rs} = \overline{r} \cdot \overline{s}$, it is in fact a ring homomorphism.

Now suppose $f : R \to S$ is any ring homomorphism. Recall that $I = \ker(f)$ is an ideal of R. By the First Isomorphism Theorem (for groups), we have a well-defined isomorphism

$$\begin{array}{rccc} \varphi : & R/I & \to & \mathrm{im}(f) \\ & \overline{r} & \mapsto & f(r) \end{array}$$

of groups under addition. Since $\varphi(\overline{1}_R) = f(1_R) = 1_S$, and

$$\varphi(\overline{r} \cdot \overline{s}) = \varphi(\overline{rs}) = f(rs) = f(r)f(s) = \varphi(\overline{rs})$$

for all $r, s \in R$, it follows that φ is in fact an isomorphism of rings. We thus obtain the "First Isomorphism Theorem" for rings!

EXAMPLE 7.46 Consider the homomorphism $\varphi : \mathbb{R}[X] \to \mathbb{C}$ defined by "evaluation at i," i.e. $\varphi(f) = f(i)$ for $f \in \mathbb{R}[X]$. We then have $\mathrm{im}(\varphi) = \mathbb{C}$ and $\ker(\varphi) = (X^2 + 1)$, so there is an isomorphism $\mathbb{R}[X]/(X^2 + 1) \to \mathbb{C}$ defined by $\overline{f} \mapsto f(i)$.

EXERCISE 7.47 Reprove Exercise 7.45 by showing that $f^{-1}(J)$ is the kernel of a map of rings using the first isomorphism theorem.

EXERCISE 7.48 Let $I \trianglelefteq R$. Prove that there is a canonical bijection

$$\{\text{ideals of } R \text{ containing } I\} \leftrightarrow \{\text{ideals of } R/I\}.$$

DEFINITION 7.49 We say an ideal I of R is *prime* if the following hold:

(i) $I \neq R$;

(ii) if $r, s \in R$ and $rs \in I$, then $r \in I$ or $s \in I$.

Note that the first condition is equivalent to $\overline{R}$ not being the zero ring, and the second means that if $\overline{r} \cdot \overline{s} = \overline{0}_R$, then $\overline{r} = \overline{0}_R$ or $\overline{s} = \overline{0}_R$. Thus I is a prime ideal of R if and only if R/I is a domain.

DEFINITION 7.50 We say an ideal I of R is *maximal* if the following hold:

(i) $I \neq R$;

(ii) if J is an ideal of R containing I, then either $J = I$ or $J = R$.

Again the first condition is equivalent to $\overline{1}_R \neq \overline{0}_R$ in $\overline{R}$. Since the ideals of R/I are in bijection with the ideals of R containing I (left as an exercise), the second condition is now equivalent to R/I having precisely two ideals (namely $\{\overline{0}_R\}$ and R/I). Thus I is a maximal ideal of R if and only if R/I is a field (how would you prove this "directly"?). In particular every maximal ideal is prime (since every field is a domain).

EXAMPLE 7.51 The maximal ideals of $\mathbb{Z}$ are those of the form $p\mathbb{Z}$, where p is prime. The ideal $\{0\}$ is prime, but not maximal.

EXAMPLE 7.52 For any ring R, the ideal $\{0_R\}$ is prime if and only if R is a domain, and maximal if and only if R is a field.

EXAMPLE 7.53 Recall from Example 7.46 that $\mathbb{R}[X]/(X^2+1)$ is isomorphic to the field $\mathbb{C}$; therefore (X^2+1) is a maximal ideal of $\mathbb{R}[X]$. More generally if K is a field and f is a non-zero polynomial in $K[X]$, then (f) is prime if and only if it is irreducible (i.e. it cannot be factored into polynomials of lower degree in $K[X]$), in which case (f) is in fact maximal. Note for example that X^2+1 is irreducible in $\mathbb{R}[X]$ (the only polynomials of lower degree are linear, and X^2+1 has no roots in $\mathbb{R}$). On the other hand $X^2+1=(X-i)(X+i)$ is reducible in $\mathbb{C}[X]$, so the ideal (X^2+1) in $\mathbb{C}[X]$ is not prime.

EXERCISE 7.54 Classify the maximal ideals of $\mathbb{R}[X]$ using results of this section.

EXAMPLE 7.55 Consider the homomorphism $\varphi : \mathbb{R}[X,Y] \to \mathbb{R}$ defined by $f \mapsto f(0,0)$. Its kernel is the set of polynomials in $\mathbb{R}[X,Y]$ with constant term 0, i.e. the ideal (X,Y) of Example 7.39. Since φ is surjective, $\mathbb{R}[X,Y]/(X,Y)$ is isomorphic to $\mathbb{R}$, so (X,Y) is a maximal ideal of $\mathbb{R}[X,Y]$.

Similarly consider the homomorphism $\psi : \mathbb{R}[X,Y] \to \mathbb{R}[X]$ defined by $\psi(f) = f(X,0)$, i.e.

$$\psi\left(\sum_{i,j=0}^{k} r_{ij}X^iY^j\right) = \sum_{i=0}^{k} r_{i,0}X^i.$$

Then $\ker(\psi)$ consists of the polynomials in which the coefficient of each monomial of the form X^iY^0 is 0, i.e. the set of polynomials f such that $Y \mid f$. Since ψ is surjective and $\ker(\psi) = (Y)$, it follows that $\mathbb{R}[X,Y]/(Y)$ is isomorphic to $\mathbb{R}[X]$, which is a domain but not a field. Therefore (Y) is a prime ideal of $\mathbb{R}[X,Y]$, but is not maximal. (Indeed one can see directly from the inclusions $(Y) \subsetneq (X,Y) \subsetneq \mathbb{R}[X,Y]$ that (Y) is not maximal.)

EXERCISE 7.56 Let $f : R \to S$ be a ring homomorphism. Prove that the preimage of a prime ideal is prime using the first isomorphism and the fact that quotienting by a prime ideals gives a domain. The analogous argument doesn't work for the preimage of maximal ideals (why?).

EXERCISE* 7.57 Let R be a commutative ring. Consider the polynomial ring $R[X_1,\ldots,X_n]$. There is a map of sets $i : \{1,\ldots,n\} \to R[X_1,\ldots,X_n]$. Show that the polynomial ring satisfies the exact analogue of the universal property of free groups (Proposition 6.9) with every instance of group replaced by commutative ring.

8 Modules

We continue to **assume R is a commutative ring**.

DEFINITION 8.1 An *R-module* is an abelian group $(M,+)$, together with a function

$$\begin{aligned} R \times M &\longrightarrow M \\ (r,m) &\longrightarrow r \cdot m \end{aligned}$$

such that the following hold:

(i) $1_r \cdot m = m$ for all $m \in M$;

(ii) $r \cdot (m+n) = r \cdot m + r \cdot n$ for all $r \in R$, $m, n \in M$;

(iii) $(r+s) \cdot m = r \cdot m + s \cdot m$ for all $r, s \in R$, $m \in M$;

(iv) $r \cdot (s \cdot m) = (rs) \cdot m$ for all $r, s \in R$, $m \in M$.

The operation sending (r, m) to $r \cdot m$ is referred to as *scalar multiplication (by R on M)*. Conditions (ii) and (iii) can be viewed as distributivity axioms, and (iv) as associativity. The distributivity easily implies that $0_R \cdot m = 0_M = r \cdot 0_M$ for all $m \in M$ and $r \in R$.

This should be compared with the action of a group on a set (Definition 4.1).

EXAMPLE 8.2 Let $R = \mathbb{Z}$ and let $(M, +)$ be any abelian group. Then M is a $\mathbb{Z}$-module, with $n \cdot m$ given its usual meaning for $n \in \mathbb{Z}$, $m \in M$ (e.g., $3 \cdot m = m + m + m$). In fact, any abelian group automatically has the structure of a $\mathbb{Z}$-module (so that these are equivalent notions).

EXAMPLE 8.3 Let $R = K$ be a field. Then a K-module is the same as a vector space over K.

EXAMPLE 8.4 For any R-module, we can view $M = R$ as an R-module by defining $r \cdot m = rm$ for $r, m \in R$ (the axioms being exactly the same as those that make R a ring).

EXAMPLE 8.5 If R is a subring of S, then $M = S$ is an R-module by defining $r \cdot s = rs$ for $r \in R$, $s \in S$. More generally if $f : R \to S$ is any homomorphism of rings, then we can make S an R-module by defining $r \cdot s = f(r)s$. The conditions in Definition 8.1 since

(i) $1_R \cdot s = f(1_R)s = 1_S s = s$ for all $s \in S$;

(ii) $r \cdot (s + s') = f(r)(s + s') = f(r)s + f(r)s' = r \cdot s + r \cdot s'$ for all $r \in R$, $s, s' \in S$;

(iii) $(r + r') \cdot s = f(r + r')s = (f(r) + f(r'))s = f(r)s + f(r')s = r \cdot s + r' \cdot s$ for all $r, r' \in R$, $s \in S$;

(iv) $r \cdot (r' \cdot s) = f(r)(f(r')s) = f(rr')s = (rr') \cdot s$ for all $r, r' \in R$, $s \in S$.

EXAMPLE 8.6 If M and N are R-modules, then so is $M \times N$, with scalar multiplication defined by $r \cdot (m, n) = (r \cdot m, r \cdot n)$ for all $r \in R$, $m \in M$, $n \in N$.

EXAMPLE 8.7 For any ring R and positive integer n, combining Examples 8.4 and 8.6 (inductively) makes R^n an R-module, with scalar multiplication defined by $r \cdot (m_1, \ldots, m_n) = (rm_1, \ldots, rm_n)$ for $r, m_1, \ldots, m_n \in R$.

EXAMPLE 8.8 Let $R = \mathbb{R}[X]$, $M = \mathbb{R}^n$ (where n is a positive integer) and fix a matrix $A \in M_n(\mathbb{R})$. Define a scalar multiplication by R on M by the formula $f \cdot \mathbf{x} = f(A)\mathbf{x}$ for $f \in R$ and $m \in M$ (where $f(A)$ has the obvious meaning, i.e. if $f = \sum_{i=0}^{k} r_i X^i$ for some $r_0, \ldots, r_k \in \mathbb{R}$, then $f(A) = \sum_{i=0}^{k} r_i A^i \in M_n(\mathbb{R})$, with A^0 interpreted as the identity matrix I_n). Note that (i) holds since $f(1) = I_n$, so $1 \cdot \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in M$, (ii) and (iii) follow from distributivity of matrix multiplication and (iv) by its associativity (for (iii) and (iv), using also that $(f + g)(A) = f(A) + g(A)$ and $(fg)(A) = f(A)g(A)$).

We remark that one can also consider modules for non-commutative rings, except that we must distinguish between *left* modules and *right* modules. For example, $\mathbb{R}^n$ is a left $M_n(\mathbb{R})$-module with scalar multiplication defined by $A \cdot \mathbf{x} = A\mathbf{x}$ for $A \in M_n(\mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^n$ (viewing $\mathbf{x}$ as a column vector), and similarly defining $\mathbf{x} \cdot A = \mathbf{x}A$ makes $\mathbb{R}^n$ a right $M_n(\mathbb{R})$-module (where each $\mathbf{x}$ is viewed as a row vector).

DEFINITION 8.9 If N is an R-module and M is a subset of N, then we say that M is an *R-submodule* of N if the following hold:

(i) M is a subgroup of N under addition;

(ii) $r \cdot m \in M$ for all $r \in R$ and $m \in M$.

Note that if M is an R-submodule of N, then M becomes an R-module with the addition and scalar multiplication inherited from N. We will often omit the $\cdot$ and write rm for $r \cdot m$.

EXAMPLE 8.10 If $N = R$, then an R-submodule of N is the same as an ideal of R.

EXAMPLE 8.11 For any R-module N, we have that $\{0_N\}$ and N itself are R-submodules of N.

EXAMPLE 8.12 If $R = \mathbb{Z}$, then an R-submodule of an abelian group N (see Example 8.2) is the same as a subgroup of N under addition.

EXAMPLE 8.13 If $R = K$ is a field and V is a K-module, i.e. a vector space over K (see Example 8.3), then a K-submodule of V is the same as a sub-K-vector space of V.

EXAMPLE 8.14 If N is an R-module and $m \in N$, then the *(cyclic) R-submodule of N generated by m* is the subset $R \cdot m = \{ rm \mid r \in R \}$. We say that an R-submodule M of N is *cyclic* if $M = R \cdot m$ for some $m \in N$. Note that an ideal of R is a cyclic R-submodule if and only if it is a principal ideal (see Definition 7.34).

EXAMPLE 8.15 Suppose that M_1 and M_2 are R-submodules of an R-module N. Then so are $M_1 \cap M_2$ and

$$M_1 + M_2 = \{ m_1 + m_2 \mid m_1 \in M_1,\, m_2 \in M_2 \}.$$

This is clear from the definition for $M_1 \cap M_2$; for $M_1 + M_2$, note that it is a subgroup since $0_N = 0_N + 0_N \in M_1 + M_2$, and if $m_1 + m_2$ and $m_1' + m_2' \in M_1 + M_2$ (where $m_1, m_1' \in M_1$ and $m_2, m_2' \in M_2$), then

$$(m_1 + m_2) + (m_1' + m_2') = (m_1 + m_1') + (m_2 + m_2') \in M_1 + M_2$$

and $-(m_1 + m_2) = (-m_1) + (-m_2) \in M_1 + M_2$, and if $r \in R$, then $r(m_1 + m_2) = (rm_1) + (rm_2) \in M_1 + M_2$.

EXAMPLE 8.16 Suppose that $m_1, \ldots, m_n$ are elements of an R-module N. Combining Examples 8.14 and 8.15 (inductively), it follows that

$$R \cdot m_1 + \cdots + R \cdot m_n = \left\{ \sum_{i=1}^{n} r_i m_i \;\middle|\; r_1, \ldots, r_n \in R \right\}$$

is an R-submodule of N. Furthermore it is the R-submodule of N generated by $A = \{m_1, \ldots, m_n\}$, in the sense that it is the smallest R-submodule of N containing A (i.e. if M is an R-submodule of N such that $A \subset M$, then $R \cdot m_1 + \cdots + R \cdot m_n \subset M$.

More generally if A is *any* subset of M, then the R-submodule of N generated by A, defined as

$$\langle A \rangle = \left\{ \sum_{j=1}^{k} s_j n_j \;\middle|\; k \geq 0,\, s_1, \ldots, s_k \in R,\, n_1, \ldots, n_k \in A \right\},$$

is the smallest R-submodule of N containing A. (Note that this reduces to Definition 7.37 in the case $N = R$.)

If M is any R-submodule of an R-module N, then M is a subgroup of N under addition, normal since N is abelian, so we can form the quotient group N/M. Not only is it an abelian group, but we make it an R-module with scalar multiplication defined by $r \cdot (n + M) = (rn) + M$ for $r \in R$, $n \in N$. We must first check that the operation is well-defined, i.e. if $\overline{n}_1 = \overline{n}_2$, then $\overline{rn}_1 = \overline{rn}_2$ (writing simply $\overline{n}$ for $n + M$, etc.): if

$\overline{n}_1 = \overline{n}_2$, then $n_1 - n_2 \in M$, so $rn_1 - rn_2 = r(n_1 - n_2) \in M$, and hence $\overline{rn}_1 = \overline{rn}_2$. The conditions in Definition 8.1 hold since

(i) $1_R \cdot \overline{n} = \overline{1_R n} = \overline{n}$ for all $n \in N$;

(ii) $r \cdot (\overline{n}_1 + \overline{n}_2) = r \cdot (\overline{n_1 + n_2}) = \overline{r(n_1 + n_2)} = \overline{rn_1 + rn_2}$
$= \overline{rn}_1 + \overline{rn}_2 = r \cdot \overline{n}_1 + r \cdot \overline{n}_2$ for all $r \in R$, $n_1, n_2 \in N$.

(iii) $(r+s) \cdot \overline{n} = \overline{(r+s)n} = \overline{rn + sn} = \overline{rn} + \overline{sn} = r \cdot \overline{n} + s \cdot \overline{n}$ for all $r, s \in R$, $n \in N$;

(iv) $r \cdot (s \cdot \overline{n}) = r \cdot \overline{sn} = \overline{rsn} = (rs) \cdot \overline{n}$ for all $r, s \in R$, $n \in N$.

We call N/M the *quotient R-module of N by M* (or just the *quotient module*, as usual read as "N mod M".)

DEFINITION 8.17 If M and N are R-modules, then a function $f : M \to N$ is an *R-module homomorphism* (or simply *R-linear*) if

(i) $f(m_1 + m_2) = f(m_1) + f(m_2)$ for all $m_1, m_2 \in M$, and

(ii) $f(r \cdot m) = r \cdot f(m)$ for all $r \in R$, $m \in M$.

If $f : M \to N$ is a homomorphism of R-modules, then we say f is an *isomorphism (of R-modules)* if it is bijective.

Note that the addition and scalar multiplication on the left-hand sides of the equations are on M, and on the right they are on N, and (i) means that f is a homomorphism of groups under addition.

EXAMPLE 8.18 If M and N are $\mathbb{Z}$-modules (i.e, abelian groups; see Example 8.2), then $f : M \to N$ is a $\mathbb{Z}$-module homomorphism if and only if it is a homomorphism of abelian groups.

EXAMPLE 8.19 If $R = K$ is a field and V and W are K-modules (i.e. K-vector spaces; see Example 8.3), then $f : V \to W$ is a K-module homomorphism if and only if it is K-linear (in the sense of maps between K-vector spaces).

EXAMPLE 8.20 If M is an R-submodule of an R-module N, then the inclusion $i : M \hookrightarrow N$ (defined by $i(m) = m$) is R-linear, as is the projection $\pi : N \to N/M$ (defined by $\pi(n) = \overline{n}$).

EXAMPLE 8.21 Let M be any R-module and fix $s \in R$. Then the function $f : M \to M$ defined by $f(m) = sm$ (i.e. scalar multiplication by s) is R-linear. Indeed $f(m+n) = s(m+n) = sm + sn = f(m) + f(n)$ for all $m, n \in M$, and $f(rm) = srm = rsm = rf(m)$ for all $r \in R$, $m \in M$.

EXAMPLE 8.22 The identity function on any R-module M is R-linear, and if $f : M \to N$ and $g : N \to P$ are homomorphisms of R-modules, then so is $g \circ f : M \to P$. Therefore R-modules form a category, denoted R-mod. If M and N are R-modules, then

$$\mathrm{Hom}_{R\text{-mod}}(M, N) = \{ f : M \to N \mid f \text{ is } R\text{-linear} \}$$

is usually just denoted $\mathrm{Hom}_R(M, N)$.

If $f : M \to N$ is an isomorphism of R-modules in the sense of Definition 8.17 (i.e. a bijective R-module homomorphism), then the inverse function of f is R-linear, so f is an isomorphism in the category of R-modules. We write $\mathrm{Aut}_R(M)$ for the group of R-module automorphisms of M (i.e. R-linear isomorphisms $f : M \to M$), so for example $\mathrm{Aut}_{\mathbb{R}}(\mathbb{R}^n)$ is isomorphic to $\mathrm{GL}_n(\mathbb{R})$ (with the matrix $A \in \mathrm{GL}_n(\mathbb{R})$ corresponding to the automorphism defined by $\mathbf{x} \mapsto A\mathbf{x}$).

EXERCISE 8.23 Let M be any R-module and fix $m \in M$. Show there is a unique R-module homomorphism $f : R \to M$ such that $1_R \mapsto m$. In other words, $\mathrm{Hom}_R(R, M) = \mathrm{Hom}_{\mathsf{Sets}}(\{1\}, M) = M$.

If M and N are R-modules and $f \in \mathrm{Hom}_R(M, N)$, then $\mathrm{im}(f) = f(M)$ is an R-submodule of N. Indeed we know already that it is a subgroup of N under addition, and if $r \in R$ and $n \in \mathrm{im}(f)$, then $n = f(m)$ for some $m \in M$, so $rn = rf(m) = f(rm) \in \mathrm{im}(f)$. Similarly $\ker(f)$ is an R-submodule of M since if $r \in R$ and $m \in \ker(f)$, then $f(rm) = rf(m) = r0_N = 0_N$, so $rm \in \ker(f)$.

EXAMPLE 8.24 Suppose that N is an R-module and fix $m \in N$. Then the function $f : R \to N$ defined by $f(r) = rm$ is R-linear since $f(r+s) = (r+s)m = rm + sm = f(r) + f(s)$ and $f(rs) = (rs)m = r(sm) = rf(s)$ for all $r, s \in R$. Note that $\mathrm{im}(f) = R \cdot m$ is the R-submodule of N generated by m (see Example 8.14), and

$$\ker(f) = \{\, r \in R \mid rm = 0_N \,\}$$

is an R-submodule (i.e. ideal) of R (called the *annihilator* of m in R).

EXAMPLE 8.25 Let N be an R-module and $m_1, \ldots, m_n \in N$. Generalizing the preceding example, consider $f : R^n \to M$ defined by

$$f(r_1, r_2, \ldots, r_n) = \sum_{i=1}^{n} r_i m_i.$$

Then $\mathrm{im}(f)$ is the R-submodule of N generated by $\{m_1, m_2, \ldots, m_n\}$ (see Example 8.16).

If $f : M \to N$ is a homomorphism of R-modules, then the First Isomorphism Theorem for groups gives an isomorphism

$$\begin{array}{rccc} \varphi : & M/\ker(f) & \to & \mathrm{im}(f) \\ & \overline{m} & \mapsto & f(m) \end{array}$$

of groups (as usual writing $\overline{m}$ for the coset $m + \ker(f)$). It is in fact an isomorphism of R-modules: we know already that it is a bijective isomorphism of groups, so it suffices to note that

$$\varphi(r\overline{m}) = \varphi(\overline{rm}) = f(rm) = rf(m) = r\varphi(\overline{m})$$

for all $r \in R$, $\overline{m} \in M/\ker(f)$.

Similarly if M and M' are R-submodules of an R-module N, then the Second Isomorphism Theorem for groups yields an isomorphism

$$M/(M \cap M') \xrightarrow{\sim} (M + M')/M'$$

of R-modules, and the Third Isomorphism Theorem yields a bijection between R-submodules of N containing M and R-submodules of N/M.

EXERCISE* 8.26 Continuing from Exercise 7.19, prove that to given an abelian group A an R-module structure is equivalent to giving a ring homomorphism $R \to \mathrm{End}_{\mathsf{Ab}}(A)$. It may be helpful to reinterpret Example 8.2 in this context. So modules are the "ring version" of group actions/representations. Writing a module this way is much less common however.

9 Direct products, direct sums and free modules

If $M_1, M_2, \ldots, M_n$ are R-modules, then so is the product of the underlying abelian groups

$$M_1 \times M_2 \times \cdots \times M_n,$$

with addition and scalar multiplication defined componentwise, so for example

$$r \cdot (m_1, m_2, \ldots, m_n) = (rm_1, rm_2, \ldots, rm_n)$$

for $r \in R$ and $(m_1, m_2, \ldots, m_n) \in M_1 \times M_2 \times \cdots \times M_n$. More generally

DEFINITION 9.1 For any collection of R-modules $\{M_\alpha\}_{\alpha \in A}$ indexed by a set A, we define their *direct product* to be the R-module

$$\prod_{\alpha \in A} M_\alpha = \{ (m_\alpha)_{\alpha \in A} \mid m_\alpha \in M_\alpha \text{ for all } \alpha \in A \}$$

with addition and scalar multiplication defined componentwise, i.e.

$$(m_\alpha)_{\alpha \in A} + (n_\alpha)_{\alpha \in A} = (m_\alpha + n_\alpha)_{\alpha \in A} \quad \text{and} \quad r \cdot (m_\alpha)_{\alpha \in A} = (rm_\alpha)_{\alpha \in A}.$$

Note that its zero element is $(0_\alpha)_{\alpha \in A}$ where 0_α is the zero element in M_α, and the additive inverse of $(m_\alpha)_{\alpha \in A}$ is $-(m_\alpha)_{\alpha \in A} = (-m_\alpha)_{\alpha \in A}$.

If $A = \{1, 2, \ldots, n\}$, then $\prod_{\alpha \in A} M_\alpha$ is just $M_1 \times M_2 \times \cdots \times M_n$, but note that A can be infinite in Definition 9.1.

EXAMPLE 9.2 Let $R = \mathbb{R}$, $A = \{ i \in \mathbb{Z} \mid i > 0 \}$ and $M_i = \mathbb{R}$ for all $i \in A$. Then

$$\prod_{i \in A} M_i = \{ (x_1, x_2, \ldots) \mid x_i \in \mathbb{R} \text{ for all } i \in A \}$$

is the set of infinite sequences of real numbers.

The direct product comes equipped with projection maps $\pi_\beta : \prod_{\alpha \in A} M_\alpha \to M_\beta$ given by $(m_\alpha)_\alpha \mapsto m_\beta$. These are in fact R-module homomorphisms. Note that given another R-module N and R-module homomorphism $\varphi : N \to \prod_{\alpha \in A} M_\alpha$, the composites

$$N \xrightarrow{\varphi} \prod_{\alpha \in A} M_\alpha \xrightarrow{\pi_\beta} M_\beta$$

are R-module homomorphisms. Denote these by φ_β. So from the data of a map $\varphi : N \to \prod_{\alpha \in A} M_\alpha$ we obtain a collection of R-module homomorphisms $(\varphi_\beta : N \to M_\beta)_{\beta \in A}$. We claim the converse also holds, i.e. given any collection of maps $(\psi_\beta : N \to M_\beta)_{\beta \in A}$, there is a unique R-module homomorphism $\psi : N \to \prod_{\alpha \in A} M_\alpha$ such that $\pi_\beta \circ \psi = \psi_\beta$. This is defined by

$$\begin{aligned} \psi : N &\longrightarrow \prod_{\alpha \in A} M_\alpha \\ n &\longmapsto (\psi_\alpha(n))_\alpha. \end{aligned}$$

After checking the details (exercise), we obtain the following proposition.

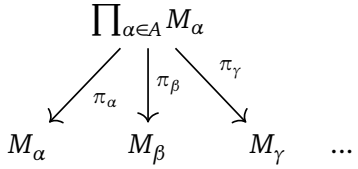
PROPOSITION 9.3 *The function*

$$\begin{array}{ccc} \mathrm{Hom}_R(N, \prod_{\alpha \in A} M_\alpha) & \to & \prod_{\alpha \in A} \mathrm{Hom}_R(N, M_\alpha) \\ \varphi & \mapsto & (\pi_\alpha \circ \varphi)_{\alpha \in A}, \end{array}$$

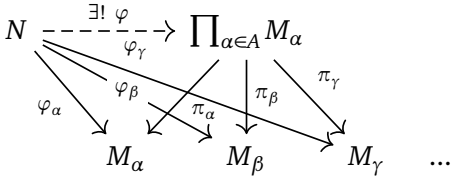
is a bijection.

Thus giving an R-module homomorphism $N \to \prod_{\alpha \in A} M_\alpha$ is equivalent to giving a collection of R-module homomorphisms $N \to M_\alpha$.

REMARK 9.4 This is actually an example of a universal property. First note that the data of a direct product and its projections can be drawn as:



(Note we do not insist that A is a countable set.) The universal property is that given a similar collection of R-module homomorphisms from a module N, those maps factor uniquely through the direct product: The universal property is then that given a diagram of solid arrows:



there is a unique R-module homomorphism φ making the diagram commute.

It is often helpful to think of the direct product $\prod_{\alpha\in A} M_\alpha$ as "the closest object to the collection $(M_\alpha)_{\alpha\in A}$.

Note again that universal properties uniquely define objects (one really should consider the maps π_α as part of the data here), but one still needs to give a construction to show they exist!

We now define the direct sum of a collection of R-modules $\{M_\alpha\}_{\alpha\in A}$, which has a similar property for R-module homomorphisms *from* (rather than *to*) the M_α

DEFINITION 9.5 For any collection of R-modules $\{M_\alpha\}_{\alpha\in A}$ indexed by a set A, we define their *direct sum*, denoted $\bigoplus_{\alpha\in A} M_\alpha$ to be the R-submodule of $\prod_{\alpha\in A} M_\alpha$ consisting of the elements $(m_\alpha)_{\alpha\in A}$ such that $m_\alpha = 0$ for all but finitely many $\alpha \in A$.

It is straightforward to check that $\bigoplus_{\alpha\in A} M_\alpha$ is indeed an R-submodule of $\prod_{\alpha\in A} M_\alpha$.

EXAMPLE 9.6 Let $M_i = R = \mathbb{R}$ for all $i \in A = \{\, i \in \mathbb{Z} \mid i > 0 \,\}$, as in Example 9.2. Then $\bigoplus_{i\in A} \mathbb{R}$ is the set of sequences $(x_1, x_2, \ldots)$ of real numbers such that $x_n = 0$ for all sufficiently large n.

For a direct sum, for each $\alpha \in A$, there is an R-module homomorphism $\iota_\alpha : M_\alpha \to \bigoplus_{\beta\in A} M_\beta$ sending $m \in M_\alpha$ to the element $(m_\beta)_{\beta\in A}$ such that

$$m_\beta = \begin{cases} m, & \text{if } \alpha = \beta; \\ 0_\beta, & \text{otherwise.} \end{cases}$$

Thus if $\psi : \bigoplus_{\alpha\in A} M_\alpha \to N$ is an R-module homomorphism, then this gives an R-module homomorphism $\psi \circ \iota_\alpha : M_\alpha \to N$ for each $\alpha \in A$. We now claim that conversely, given any collection of maps $\psi_\alpha \colon M_\alpha \to N$, there is a unique R-module homomorphism $\psi \colon \bigoplus_{\alpha\in A} M_\alpha \to N$ such that $\psi \circ \iota_\alpha = \psi_\alpha$. This is given by the formula

$$\psi\left((m_\alpha)_{\alpha\in A}\right) = \sum_{\alpha\in A} \psi_\alpha(m_\alpha). \tag{2}$$

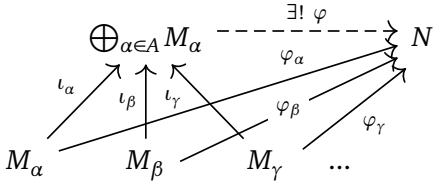
Note that by the definition of $\bigoplus_{\alpha\in A} M_\alpha$, only finitely many of the m_α can be non-zero, so the same is true of the $\psi_\alpha(m_\alpha) \in N$, and the sum makes sense. (To be precise, define it as $\sum_{\alpha\in B} \psi_\alpha(m_\alpha)$ for any finite subset B of A such that $\psi_\alpha(m_\alpha) = 0$ for all $\alpha \notin B$.) Filling in all the details, we obtain:

PROPOSITION 9.7 *The function*

$$\begin{array}{ccc} \mathrm{Hom}_R(\bigoplus_{\alpha\in A} M_\alpha, N) & \longrightarrow & \prod_{\alpha\in A}\mathrm{Hom}_R(M_\alpha, N) \\ \psi & \longmapsto & \psi\circ\iota_\alpha, \end{array}$$

is a bijection.

REMARK 9.8 This is also a universal property. Direct sums are the "closest objects from a collection $(M_\alpha)_{\alpha\in A}$". That is given a diagram of solid arrows:



there is a unique map φ making the diagram commute.

EXERCISE 9.9 The direct product also has inclusion maps $\iota_\alpha\colon M_\alpha \to \prod_{\alpha\in A} M_\alpha$. Why does the direct product not satisfy the universal property of direct products? (*Hint: Show equation* (2) *can't work, but equally we are forced to have it.*) Conversely, the direct sum has projection maps, why does the direct sum not satisfy the universal property of direct products?

DEFINITION 9.10 For any set A, we define the *free R-module* on A to be

$$F_A = \bigoplus_{\alpha\in A} R.$$

We let e_α denote the element $\iota_\alpha(1) \in F_A$.

EXAMPLE 9.11 If $A = \{1, 2, \ldots, n\}$, then $F_A = R^n$ and $e_i = (0, 0, \ldots, 0, 1, 0 \ldots, 0)$ where the 1 is in the i^{th} place.

EXAMPLE 9.12 For $R = \mathbb{R}$ and $A = \{\, i \in \mathbb{Z} \mid i > 0 \,\}$, the free $\mathbb{R}$-module $F_A = \bigoplus_{i\in A} \mathbb{R}$ is described in Example 9.6.

Note that in general $e_\alpha = (\delta_{\alpha,\beta})_{\beta\in A}$, where $\delta_{\alpha,\beta} = 1$ or 0 according to whether or not $\alpha = \beta$. Note that since

$$(r_\alpha)_{\alpha\in A} = \sum_{\alpha\in A} r_\alpha e_\alpha,$$

every element of F_A is uniquely expressed as an R-linear combination[21] of the elements e_α.

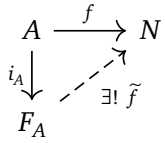
Since free modules are direct products, they have a universal property as in Remark 9.8. But, this becomes especially nice in this case. The general universal property says that to give an R-module homomorphism $F_A \to N$ is equivalent to giving an R-module homomorphism $R \to N$ for each $\alpha \in A$. But we saw in Exercise 8.23 that $\mathrm{Hom}_R(R, N) = \mathrm{Hom}_{\mathsf{Sets}}(\{1\}, N)$ (i.e. homomorphisms from the trivial module are determined by the image of 1 and any choice of image is possible). In conclusion

$$\mathrm{Hom}_R(F_A, N) \longleftrightarrow \left\{ \begin{array}{l} \text{choices of the image of 1} \\ \text{for each } \alpha \in A \end{array} \right\} \longleftrightarrow \mathrm{Hom}_{\mathsf{Sets}}(A, N).$$

We have shown:

[21] Note that since only finitely many of the r_α are non-zero, the sum can be interpreted as a finite sum, i.e. as $\sum_{\alpha\in B} r_\alpha e_\alpha$ for some finite subset B of A.

PROPOSITION 9.13 *For each map of sets $f : A \to N$, there is a unique R-module homomorphism $\varphi : F_A \to N$ such that $\varphi(e_\alpha) = f(\alpha)$ for all $\alpha \in A$. In other words, free modules satisfy the following universal property:*



where i_A denotes the map of sets $A \to F_A$ given by $\alpha \mapsto e_\alpha$.

Note that this is exactly the analogue of the universal property of free groups (Remark 6.11).

DEFINITION 9.14 We say an R-module M is *free* if M is isomorphic to F_A for some set A. If $\varphi : F_A \xrightarrow{\sim} M$ is an isomorphism, we say that the set $\{\varphi(e_\alpha) \mid \alpha \in A\}$ is an R-*basis* for M (indexed by A).

Note that if M is a free R-module and the set $\mathcal{B} = \{m_\alpha \mid \alpha \in A\}$ is an R-basis for M indexed by A, then each element of M can be expressed uniquely as a (finite) R-linear combination of elements of the R-basis $\mathcal{B}$ (since the same is true for F_A for the R-basis $\{e_\alpha \mid \alpha \in A\}$). This means that

(i) M is generated by $\mathcal{B}$ (as an R-module), i.e. $M = \langle \mathcal{B} \rangle$ (see Example 8.16),

(ii) and $\mathcal{B}$ is *linearly independent* (over R) in the sense that if $(r_\alpha)_{\alpha \in A} \in F_A$ is such that $\sum_{\alpha \in A} r_\alpha m_\alpha = 0$, then $r_\alpha = 0$ for all $\alpha \in A$.

Conversely if M is an R-module and $\mathcal{B}$ is a subset of M satisfying (i) and (ii), then the R-linear homomorphism $\varphi : F_{\mathcal{B}} \to M$ corresponding to the inclusion $\mathcal{B} \hookrightarrow M$ is an isomorphism, so M is free and $\mathcal{B}$ is an R-basis for M in the sense of Definition 9.14. (We will just say *basis* instead of *R-basis* when R is clear from the context.)

EXAMPLE 9.15 If $R = \mathbb{Z}$ and M is a free $\mathbb{Z}$-module, then M is isomorphic to $\bigoplus_{\alpha \in A} \mathbb{Z}$ for some set A, so if A is finite, say of cardinality n, then M is isomorphic to $\mathbb{Z}^n$. Note that $\mathbb{Z}^n$ has infinitely many bases (if $n > 1$); for example $\{(1, m), (0, 1)\}$ is a basis for $\mathbb{Z}^2$ (for any $m \in \mathbb{Z}$).

EXAMPLE 9.16 On the other hand $\mathbb{Q}$ is not a free $\mathbb{Z}$-module. Indeed suppose that $\mathcal{B}$ were a $\mathbb{Z}$-basis for $\mathbb{Q}$. Note that $\mathbb{Q}$ is not a cyclic $\mathbb{Z}$-module (since for example $r/2 \notin \langle r \rangle$ if $r \in \mathbb{Q}^\times$), so $\mathcal{B}$ must contain at least two distinct (necessarily non-zero) elements, say $r = a/b$, $s = c/d$, where $a, b, c, d \in \mathbb{Z}$ (with $a, b, c, d \neq 0$), but then $bcr - ads = 0$, so r and s cannot be linearly independent.

We will show that if K is a field, then every K-module (i.e. K-vector space) is free (i.e. has a basis). This fact should be very familiar for finite-dimensional vector spaces, the idea of the proof being that any maximal linearly independent subset of a vector space V in fact spans V, and hence is a basis. The general proof is based on the same idea, but one must be more careful in making sense of the word *maximal*, and the fact that such maximal subsets exist relies on an axiom from set theory called Zorn's Lemma[22].

We first recall some definitions from set theory needed for the statement of Zorn's Lemma:

[22]Usually, one instead assumes the axiom of choice and deduces Zorn's lemma as a consequence. If you are unwilling to assume Zorn's lemma/the axiom of choice, then you lose the existence of bases.

DEFINITION 9.17 A *partial ordering* on a set $\mathcal{S}$ is a binary relation $\leq$ on $\mathcal{S}$ (so $\leq$ is a comparison on pairs of elements of $\mathcal{S}$, but not every pair of elements are comparable) which is

(i) reflexive, i.e. $x \leq x$ for all $x \in \mathcal{S}$;

(ii) anti-symmetric, i.e. if $x, y \in \mathcal{S}$ are such that $x \leq y$ and $y \leq x$, then $x = y$;

(iii) transitive, i.e. if $x, y, z \in \mathcal{S}$ are such that $x \leq y$ and $y \leq z$, then $x \leq z$.

Suppose that $\mathcal{S}$ is partially ordered (with respect to $\leq$) and $\mathcal{T}$ is a subset of $\mathcal{S}$. We say that

- an element $x \in \mathcal{S}$ is an *upper bound* for $\mathcal{T}$ if $y \leq x$ for all $y \in \mathcal{T}$ (note x need not lie in $\mathcal{T}$);
- an element $x \in \mathcal{T}$ is *maximal* (in $\mathcal{T}$) if the only element $y \in \mathcal{T}$ such that $x \leq y$ is $y = x$ (note this does not mean that $y \leq x$ for all $y \in \mathcal{T}$);
- $\mathcal{T}$ is a *chain* if for every $x, y \in \mathcal{T}$, we have $x \leq y$ or $y \leq x$. Alternatively, chains are subspaces such that $\leq$ restricts to a *total ordering* (as now any two elements are comparable).

There are of course similar notions of *lower bound* and *minimal element*. Recall also that $\leq$ is a *total ordering* on $\mathcal{S}$ if (in addition) it has the property that for every $x, y \in \mathcal{S}$, we have $x \leq y$ or $y \leq x$ (i.e. $\mathcal{S}$ is itself a chain). Note also that every subset of a chain is a chain.

EXAMPLE 9.18 Consider the relation $\leq$ on $\mathcal{S} = \mathbb{R}^2$ defined by $(x, y) \leq (x', y')$ if $x \leq x'$ and $y \leq y'$. Then $\leq$ is a partial ordering.

The subset $\{(x, x) \mid x \in \mathbb{R}\}$ is a chain, with no upper bound in $\mathbb{R}^2$ and no maximal elements. Its subset $\{(x, x) \mid x < 0\}$ (also a chain) has as an upper bound in $\mathbb{R}^2$ (for example $(0, 0)$), but it has no maximal elements.

The subset $\mathcal{T} = \{(x, -x) \mid x \in \mathbb{R}\}$ is not a chain; every element of $\mathcal{T}$ is maximal, and $\mathcal{T}$ has no upper bound in $\mathbb{R}^2$.

The subset $\mathcal{C} = \{(x, y) \mid x^2 + y^2 = 1\}$ is not a chain. An element $(x, y) \in \mathcal{C}$ is maximal if and only if $x \geq 0$ and $y \geq 0$. An element $(x, y) \in \mathbb{R}^2$ is an upper bound for $\mathcal{C}$ if and only if $x \geq 1$ and $y \geq 1$.

Zorn's Lemma can be stated as follows:

LEMMA 9.19 *Suppose that $\mathcal{S}$ is a partially ordered set with the property that every chain contained in $\mathcal{S}$ has an upper bound in $\mathcal{S}$. Then $\mathcal{S}$ has a maximal element.*

(Very loosely, Zorn's lemma says that if you can deal with chains, it can deal with arbitrary subsets.)

THEOREM 9.20 *If K is a field and V is a K-module, then V is free (i.e. V has a basis).*

Proof. Let $\mathcal{S}$ denote the set of linearly independent subsets of V. Then $\mathcal{S}$ is partially ordered with respect to inclusion.

We show that $\mathcal{S}$ satisfies the hypotheses of Zorn's Lemma. Suppose that $\mathcal{T} \subset \mathcal{S}$ is a chain. (It may be helpful to think of this as a sequence of increasing subsets

$$\ldots \subseteq A_i \subseteq A_{i+1} \subseteq A_{i+2} \subseteq \ldots,$$

each of which is linearly independent, but in fact $\mathcal{T}$ need not be countable, so we need not have quite as nice a picture.) We will show that $\mathcal{T}$ has an upper bound in $\mathcal{S}$. We

want this to be given by $B = \bigcup_{A\in\mathcal{T}} A$, but we must show that $B \in \mathcal{S}$, i.e. that B is linearly independent.

Suppose then that $\sum_{i=1}^{n} r_i v_i = 0$ for some elements $r_1, \ldots, r_n \in K$ and (distinct) $v_1, \ldots, v_n \in B$. We claim that in fact $v_1, \ldots, v_n \in A$ for some $A \in \mathcal{T}$. We prove this by showing, by induction on i, that $v_1, \ldots, v_i \in A_i$ for some $A_i \in \mathcal{T}$. Clearly since $v_1 \in B = \bigcup_{A\in\mathcal{T}} A$, we have $v_1 \in A_1$ for some $A_1 \in \mathcal{T}$. Suppose then that $1 \leq i \leq n-1$ and $v_1, \ldots, v_i \in A_i$ for some $A_i \in \mathcal{T}$. Since $v_{i+1} \in B$, we have $v_{i+1} \in A'$ for some $A' \in \mathcal{T}$. By assumption $\mathcal{T}$ is a chain, so either $A' \subset A_i$ or $A_i \subset A'$. Thus letting $A_{i+1} = A_i$ or A' accordingly, we have $v_1, \ldots, v_i, v_{i+1} \in A_{i+1}$, as required.

In particular since $v_1, \ldots, v_n \in A_n$ and A_n is linearly independent (since $A_n \in \mathcal{T} \subset \mathcal{S}$), it follows that $r_1 = \ldots = r_n = 0$. This proves that B is linearly independent, i.e. $B \in \mathcal{S}$. It is clear that $A \subset B$ for all $A \in \mathcal{T}$, so B is an upper bound for $\mathcal{T}$.

It now follows from Zorn's Lemma that there is a maximal element $A \in \mathcal{S}$. By the definition of $\mathcal{S}$, A is linearly independent, so it just remains to prove that V is generated by A, i.e. that $V = \langle A \rangle$.

Suppose then that $v \in V$, and consider the subset $A' = A \cup \{v\}$ of V. If $v \in A$, then clearly $v \in \langle A \rangle$, so assume $v \notin A$. Since A is a maximal element of $\mathcal{S}$ and $A \subsetneq A'$, we must have $A' \notin \mathcal{S}$, i.e. A' is not linearly independent. This means that

$$rv + \sum_{i=1}^{n} r_i v_i = 0$$

for some $r, r_1, \ldots, r_n \in K$, not all zero, and distinct $v_1, \ldots, v_n \in A$. Since A is linearly independent, we must have $r \neq 0$. Since K is a field, we have an element $r^{-1} \in K$, and

$$v = \sum_{i=1}^{n} (-r^{-1} r_i) v_i \in \langle A \rangle,$$

as required. □

EXERCISE 9.21 In the above proof, understand why the argument wouldn't work if didn't assume that $\mathcal{T}$ was a chain (...upper bounds of non-chains need not exist...).

EXAMPLE 9.22 Let $K = \mathbb{R}$ and $A = \{i \in \mathbb{Z} \mid i > 0\}$, and consider the $\mathbb{R}$-module $V = \prod_{i\in A} \mathbb{R}$ described in Example 9.2. Note that the subset $\{e_i \mid i \in A\}$ (where $e_i = (0, \ldots, 0, 1, 0, \ldots)$ has a 1 in the i^{th} place) is a basis for the $\mathbb{R}$-submodule $F_A = \bigoplus_{i\in A} \mathbb{R}$ of V. So $\{e_i \mid i \in A\}$ is a linearly independent subset of V, but it is not a basis for V since it does not generate V (convince yourself of this!). On the other hand, Theorem 9.20 implies that V is a free $\mathbb{R}$-module, so it has some basis, say $\mathcal{B}$.

We claim that $\mathcal{B}$ is uncountable (not examinable). For this it suffices to prove that V has an uncountable linearly independent subset S[23]. For each $r \in \mathbb{R}$, let $v_r = (1, r, r^2, \ldots) \in V$, and let $S = \{v_r \mid r \in \mathbb{R}\}$. Then S is linearly independent since if $r_1, \ldots, r_n$ are distinct real numbers, then the Vandermonde determinant formula shows that the set of vectors

$$\{(1, r_i, r_i^2, \ldots, r_i^{n-1}) \mid i = 1, \ldots, n\}$$

is linearly independent in $\mathbb{R}^n$.

[23] Every element of V is in the span of a finite subset of $\mathcal{B}$, and if S is linearly independent, then each finite subset of $\mathcal{B}$ can have only finitely many elements of S in its span. If $\mathcal{B}$ were countable, then the set of finite subsets of $\mathcal{B}$ would be countable, and therefore S would have to be countable.

In particular, this shows that V is *not* isomorphic to F_A (which has a countable basis). Note that this is not immediate from the fact that F_A is a proper subspace of V. (For example, F_A is isomorphic to its proper subspace spanned by the elements $\{e_i \mid i > 1\}$.)

In the spirit of this course, it is worth pointing out that using Zorn's lemma to product maximal objects is ubiquitous across this area. Necessarily, the proofs all have the same template, showing that chains have upper bounds!

EXERCISE* 9.23 Show that, assuming Zorn's lemma, every ideal I of a ring R is contained in a maximal ideal. Take $\mathcal{S}$ to be the set of ideals containing I which are not equal to R. Once again, understand why we can only provide upper bounds for chains in general.

10 R-linear and R-bilinear maps

EXAMPLE 10.1 Consider R as a module over itself. Recall that the product $R \times R$ is also an R-module. We can define a map of sets

$$\begin{aligned} \rho : R \times R &\longrightarrow R \\ (r,s) &\longmapsto rs. \end{aligned}$$

This is not even a map of abelian groups! For example,

$$\rho((1_R, 0_R)) + \rho((0_R, 1_R)) = 1_R \cdot 0_R + 0_R \cdot 1_R = 0_R + 0_R = 0_R \neq 1_R = \rho((1_R, 1_R))$$

(unless R is the zero ring). More generally,

$$(r + r', s + s') \mapsto (r + r')(s + s') = rs + r's + rs' + r's' \neq rs + r's'.$$

The map ρ is a perfectly reasonable map to consider in nature, but doesn't fit into our rigid notion of R-linear morphisms.

DEFINITION 10.2 Suppose that M, N and P are R-modules. We say that a function $\rho : M \times N \longrightarrow P$ is *R-bilinear* if

- $\rho(m + m', n) = \rho(m, n) + \rho(m', n)$ for all $m, m' \in M$, $n \in N$;
- $\rho(m, n + n') = \rho(m, n) + \rho(m, n')$ for all $m \in M$, $n, n' \in N$;
- $\rho(\lambda m, n) = \lambda\rho(m, n) = \rho(m, \lambda n)$ for all $\lambda \in R$, $m \in M$, $n \in N$.

EXERCISE 10.3 Check this is equivalent to showing that

$$\rho(rm + r'm', sn + s'n') = rs\rho(m, n) + r's\rho(m', n) + rs'\rho(m, n') + r's'\rho(m', n')$$

for all $r, r', s, s' \in R$, $m, m' \in M$, $n, n' \in N$.

EXAMPLE 10.4 The map of Example 10.1 is R-bilinear. For example, $\rho(\lambda r, s) = (\lambda r)s = \lambda(rs) = \lambda\rho(r, s) = r(\lambda s) = \rho(r, \lambda s)$.

Note, we continue to assume that R is commutative. There is a theory for non-commutative rings, but it requires more care.

EXAMPLE 10.5 More generally let $R = M$, let $N = P$ be any R-module and consider the map $\rho : R \times N \to N$ defined by scalar multiplication, i.e. $\rho(r, n) = rn$. The fact that ρ is R-bilinear then follows from the R-module axioms, and the commutativity of R.

EXAMPLE 10.6 Let $R = \mathbb{R}$ and $M = N = P = \mathbb{C}$. We claim that the map

$$\begin{aligned} \rho : \mathbb{C} \times \mathbb{C} &\longrightarrow \mathbb{C} \\ (x, y) &\longmapsto xy \end{aligned}$$

is $\mathbb{R}$-bilinear. This follows as the map is $\mathbb{C}$-bilinear by Example 10.4 and to be $\mathbb{R}$-bilinear is weaker than to be $\mathbb{C}$-bilinear. We also have that

$$\begin{aligned} \rho' \colon \mathbb{C} \times \mathbb{C} &\longrightarrow \mathbb{C} \\ (x, y) &\longmapsto x\bar{y} \end{aligned}$$

is $\mathbb{R}$-bilinear. Now we can no longer appeal to the fact that the map is $\mathbb{C}$-bilinear. We can either check the axioms directly or note that if $\psi \colon N \times M \to P$ is a R-bilinear map and $f : M' \to M$ is an R-linear map, then $\psi \circ (\mathrm{id}, f) \colon N \times M' \to P$ is R-bilinear, where for us $\psi = \rho$ and $f(y) = \bar{y}$.

There are also many other interesting sources of R-bilinear maps. One such is outlined below.

Composition of R-module homomorphisms (not lectured)

We claim that $\mathrm{Hom}_R(M,N)$ is in fact an R-module, with addition and scalar multiplication defined as follows: if $\varphi, \varphi' \in \mathrm{Hom}_R(M,N)$ and $r \in R$, then

$$\begin{array}{rccc} \varphi + \varphi' \colon & M & \to & N \\ & m & \mapsto & \varphi(m) + \varphi'(m) \end{array} \quad \text{and} \quad \begin{array}{rccc} r \cdot \varphi \colon & M & \to & N \\ & m & \mapsto & r\varphi(m). \end{array}$$

There are many (easy) things to check to verify this claim, firstly that $\varphi + \varphi'$ and $r \cdot \varphi$ are in $\mathrm{Hom}_R(M,N)$, i.e. that they are indeed R-linear. Having done that one also needs to check that the operations thus defined satisfy the R-module axioms in Definition 8.1. We carry out a few of the checks, leaving the rest as an exercise. For example, to see that $\varphi + \varphi' \in \mathrm{Hom}_R(M,N)$, note that

$$\begin{aligned} (\varphi + \varphi')(m + m') &= \varphi(m + m') + \varphi(m + m') \\ &= \varphi(m) + \varphi(m') + \varphi'(m) + \varphi'(m') \\ &= (\varphi + \varphi')(m) + (\varphi + \varphi')(m') \end{aligned}$$

for all $m, m' \in M$, and similarly we find that $(\varphi + \varphi')(rm) = r((\varphi + \varphi')(m))$ for all $r \in R$, $m \in M$. To verify one of the R-module axioms, namely the distributive law $r \cdot (\varphi + \varphi') = (r \cdot \varphi) + (r \cdot \varphi')$ (for $r \in R$, $\varphi, \varphi' \in \mathrm{Hom}_R(M,N)$), note that

$$\begin{aligned} (r \cdot (\varphi + \varphi'))(m) &= r((\varphi + \varphi')(m) \\ &= r(\varphi(m) + \varphi'(m)) \\ &= r(\varphi(m)) + r(\varphi'(m)) \\ &= (r \cdot \varphi)(m) + (r \cdot \varphi')(m) \\ &= ((r \cdot \varphi) + (r \cdot \varphi'))(m) \end{aligned}$$

for all $m \in M$ (where all the equations follow from the definition of the operations on $\mathrm{Hom}_R(M,N)$, except the third, which uses the corresponding distributive law on N).

EXERCISE 10.7 What would have happened if we'd instead defined scalar multiplication by $r \cdot \varphi = (m \mapsto \varphi(rm))$?

EXAMPLE 10.8 Let $M = R^m$ and $N = R^n$. By Proposition 9.13, to give an R-linear map $\varphi : R^m \to R^n$ is equivalent to specifying the elements $\varphi(e_j)$ for $j = 1, \ldots, m$, where $e_1, \ldots, e_m$ are the standard basis elements of $M = R^m$. Writing $\varphi(e_j) = \sum_{i=1}^{n} a_{ij} f_i$, where $f_1, \ldots, f_n$ are the standard basis elements of $N = R^n$ and $a_{ij} \in R$ for $i = 1, \ldots, n$,

we see that to give such a φ is equivalent to giving an $n \times m$-matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

with entries $a_{ij} \in R$. Denoting this matrix by $(a_{ij})_{i,j}$ and the set of such matrices by $M_{n,m}(R)$, we thus have a bijection

$$M_{n,m}(R) \longrightarrow \mathrm{Hom}_R(R^m, R^n), \tag{3}$$

under which the matrix $A = (a_{ij})_{i,j}$ corresponds to the R-linear map $\varphi_A : R^m \to R^n$ defined by

$$\varphi_A\left(\sum_{j=1}^{m} r_j e_j\right) = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij} r_j\right) f_i,$$

i.e. writing the elements of R^m and R^n as column vectors, we have $\varphi_A(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} \in R^m$. Furthermore the bijection (3) is in fact an isomorphism of R-modules (with the R-module structure on $\mathrm{Hom}_R(M,N)$ defined above, and the obvious one on $M_{n,m}(R)$ given by addition and scalar multiplication of matrices). Note also that $M_{n,m}(R)$ is a free R-module, with basis $\{E_{ij} \mid i = 1,\ldots,n,\ j = 1,\ldots,m\}$ where E_{ij} is the matrix with (i,j) entry equal to 1 and all other entries 0, so that

$$(a_{ij})_{i,j} = \sum_{i,j} a_{ij} E_{ij}.$$

Now suppose that L, M and N are R-modules, so we have the map

$$\begin{array}{ccccc} \mathrm{Hom}_R(L,M) & \times & \mathrm{Hom}_R(M,N) & \longrightarrow & \mathrm{Hom}_R(L,N) \\ (\psi & , & \varphi) & \longmapsto & \varphi \circ \psi \end{array}$$

defined by composition. We claim that this map is R-bilinear.

EXAMPLE 10.9 For any R-modules L, M and N, the composition map

$$\mathrm{Hom}_R(L,M) \times \mathrm{Hom}_R(M,N) \longrightarrow \mathrm{Hom}_R(L,N)$$

is R-bilinear. For the first condition, we must show that if $\psi, \psi' \in \mathrm{Hom}_R(L,M)$ and $\varphi \in \mathrm{Hom}_R(M,N)$, then $\varphi \circ (\psi + \psi') = (\varphi \circ \psi) + (\varphi \circ \psi')$. This holds since

$$\begin{aligned} (\varphi \circ (\psi + \psi'))(\ell) &= \varphi((\psi + \psi')(\ell)) \\ &= \varphi(\psi(\ell) + \psi'(\ell)) \\ &= \varphi(\psi(\ell)) + \varphi(\psi'(\ell)) \\ &= (\varphi \circ \psi)(\ell) + (\varphi \circ \psi')(\ell) \\ &= ((\varphi \circ \psi) + (\varphi \circ \psi'))(\ell) \end{aligned}$$

for all $\ell \in L$, where all the equalities follow from definitions, except the third, which uses that φ is a homomorphism. We leave verification of the other two conditions in the description of bilinearity as an exercise.

Note that as a special case, we see that matrix multiplication (as in Example 10.10) is R-bilinear (as one can also check directly).

EXAMPLE 10.10 Let $L = R^\ell$, $M = R^m$ and $N = R^n$. We saw in Example 10.8 that $\mathrm{Hom}_R(M,N)$ is isomorphic to $M_{n,m}(R)$ (as an R-module), with the matrix $A \in M_{n,m}(R)$ corresponding to the R-linear map $\varphi_A : M \to N$ defined by $\varphi_A(\mathbf{x}) = A\mathbf{x}$. Similarly $\mathrm{Hom}_R(L,M)$ is isomorphic to $M_{m,\ell}(R)$, the matrix $B \in M_{m,\ell}(R)$ corresponding to φ_B :

$L \to M$ defined by $\varphi_B(\mathbf{y}) = B\mathbf{y}$. Thus the composite $\varphi_A \circ \varphi_B$ is the R-linear map $L \to N$ sending $\mathbf{x}$ to

$$\varphi_A(\varphi_B(\mathbf{x})) = A(B\mathbf{x}) = (AB)\mathbf{x} = \varphi_{AB}(\mathbf{x}),$$

where $AB \in M_{n,\ell}(R)$ is defined as usual by matrix multiplication. Thus under the R-module isomorphisms $\mathrm{Hom}_R(M,N) \cong M_{n,m}(R)$, etc., the map defined by composition of R-module homomorphisms corresponds to the map

$$\begin{array}{rccccc} \rho: & M_{m,\ell}(R) & \times & M_{n,m}(R) & \longrightarrow & M_{n,\ell}(R) \\ & (B & , & A) & \mapsto & AB \end{array}$$

defined by matrix multiplication. It is easy to directly check this is R-bilinear (as it must be by the above) and not R-linear.

11 Tensor products: the definition

Before giving the general definition of the tensor product of two R-modules, let us first consider the case where $R = K$ is a field. Suppose that V and W are vector spaces over K, say of (finite) dimensions m and n, with bases $\mathcal{B} = \{e_1, \ldots, e_m\}$ (for V) and $\mathcal{C} = \{f_1, \ldots, f_n\}$ (for W).

As a provisional first definition, let $V \otimes_K W$ denote the K-vector space of dimension mn with basis $\mathcal{B} \times \mathcal{C}$, but write $e_i \otimes f_j$ for the basis element (e_i, f_j), so that

$$V \otimes_K W = \left\{ \sum_{i,j} r_{i,j}(e_i \otimes f_j) \;\middle|\; r_{i,j} \in K \text{ for } i = 1, \ldots, m,\ j = 1, \ldots, n \right\}.$$

More generally, let us continue to assume that $R = K$ is a field, but drop the assumption that V and W are finite-dimensional. Letting $\{e_\alpha \mid \alpha \in A\}$ be a basis for V and $\{f_\beta \mid \beta \in B\}$ a basis for W, we define $V \otimes_K W$ to be the K-vector space with basis elements $e_\alpha \otimes f_\beta := (e_\alpha, f_\beta)$ indexed by $A \times B$. Thus the elements of $V \otimes_K W$ have the form

$$\sum_{\alpha \in A, \beta \in B} r_{\alpha,\beta}(e_\alpha \otimes f_\beta),$$

where the $r_{\alpha,\beta} \in K$ are such that $r_{\alpha,\beta} = 0$ for all but finitely many $(\alpha, \beta) \in A \times B$.

Note that there is a unique K-bilinear map

$$\tau : V \times W \longrightarrow V \otimes_K W$$

such that $\tau(e_\alpha, f_\beta) = e_\alpha \otimes f_\beta$ for all $\alpha \in A$, $\beta \in B$. Indeed if $v = \sum_{\alpha \in A} r_\alpha e_\alpha \in V$ and $w = \sum_{\beta \in B} s_\beta f_\beta \in W$ (with all but finitely many r_α and s_β equal to 0), we are then forced to have

$$\tau(v,w) = \tau\left(\sum_{\alpha \in A} r_\alpha e_\alpha, \sum_{\beta \in B} s_\beta f_\beta\right) = \sum_{\alpha \in A, \beta \in B} r_\alpha s_\beta\, \tau(e_\alpha, f_\beta) = \sum_{\alpha \in A, \beta \in B} r_\alpha s_\beta (e_\alpha \otimes f_\beta)$$

(with the middle map being the R-bilinearity condition). To see that this actually is K-bilinear, note that if v and w are as above and $v' = \sum_{\alpha \in A} r'_\alpha e_\alpha \in V$, then

$$\begin{array}{rcl} \tau(v + v', w) & = & \sum_{\alpha,\beta}(r_\alpha + r'_\alpha)s_\beta(e_\alpha \otimes f_\beta) \\ & = & \left(\sum_{\alpha,\beta} r_\alpha s_\beta (e_\alpha \otimes f_\beta)\right) + \left(\sum_{\alpha,\beta} r'_\alpha s_\beta (e_\alpha \otimes f_\beta)\right) \\ & = & \tau(v,w) + \tau(v',w); \end{array}$$

similarly $\tau(v, w + w') = \tau(v,w) = \tau(v,w')$ for all $v \in V$, $w, w' \in W$, and $\tau(rv, w) = r\,\tau(v,w) = \tau(v, rw)$ for all $r \in K$, $v \in V$ and $w \in W$.

For $(v,w) \in V \times W$, we will write $v \otimes w$ for the element $\tau(v,w)$ of $V \otimes_K W$. Note that this is consistent with our previous notation since $\tau(e_\alpha, f_\beta) = e_\alpha \otimes f_\beta$. The following remark is very important!

REMARK 11.1 We call elements in the image of τ (i.e. of the form $v \otimes w$) *simple tensors*. We shall show that not every element of $V \otimes_K W$ is a simple tensor, i.e. τ need not be surjective! Let $V = W = K^2$. Then consider $e_1 \otimes e_2 + e_2 \otimes e_1 \in V \otimes_K W$. Suppose that $\tau(v,w) = e_1 \otimes e_2 + e_2 \otimes e_1$ for some $v, w \in K^2$. Write

$$\begin{aligned} v &= \alpha_1 e_1 + \alpha_2 e_2, \\ w &= \beta_1 e_1 + \beta_2 e_2. \end{aligned}$$

Then

$$\begin{aligned} \tau(v,w) &= \tau(\alpha_1 e_1 + \alpha_2 e_2, \beta_1 e_1 + \beta_2 e_2) \\ &= \alpha_1\beta_1\tau(e_1,e_1) + \alpha_1\beta_2\tau(e_1,e_2) + \alpha_2\beta_1\tau(e_2,e_1) + \alpha_2\beta_2\tau(e_2,e_2) \\ &= \alpha_1\beta_1(e_1 \otimes e_1) + \alpha_1\beta_2(e_1 \otimes e_2) + \alpha_2\beta_1(e_2 \otimes e_1) + \alpha_2\beta_2(e_2 \otimes e_2). \end{aligned}$$

Since the $e_i \otimes e_j$ form a basis, we deduce that $\alpha_1\beta_1 = 0$ so that either α_1 or β_1 equals zero. In either case, one of $\alpha_1\beta_2(e_1 \otimes e_2)$ or $\alpha_2\beta_1(e_2 \otimes e_1)$ equals zero. But this is a contradiction. In fact, the image of τ isn't even a vector space (why?)!

In particular, $V \otimes_K W$ is not just $V \times W$ (in general). We can see this much more quickly since $\dim_K(V \times W) = \dim_K V + \dim_K W$ whereas $\dim_K(V \otimes W) = (\dim_K V) \cdot (\dim_K W)$.

Suppose now that U is any K-vector space and $\rho : V \times W \to U$ is any K-bilinear map. We claim that there is a unique *K-linear* map $\varphi : V \otimes_K W \to U$ such that $\rho = \varphi \circ \tau$, i.e. such that the following diagram commutes:

(4)
$$\begin{array}{ccc} V \times W & \xrightarrow{\rho} & U \\ {\scriptstyle\tau}\downarrow & \nearrow_{\exists!\ \varphi} & \\ V \otimes_K W & & \end{array}$$

Indeed to define a K-linear map $\varphi : V \otimes_K W \to U$ amounts to specifying the image of each basis element $e_\alpha \otimes f_\beta$, which need to be given by $\varphi(e_\alpha \otimes f_\beta) = \varphi(\tau(e_\alpha, f_\beta)) = \rho(e_\alpha, f_\beta)$. To see that this gives $\varphi(\tau(v,w)) = \rho(v,w)$ for *all* $v \in V$ and $w \in W$, we can write $v = \sum_{\alpha\in A} r_\alpha e_\alpha$ and $w = \sum_{\beta\in B} s_\beta f_\beta$ (with only finitely many non-zero coefficients as usual), and note that

$$\begin{aligned} \varphi(\tau(v,w)) &= \varphi\left(\sum_{\alpha\in A,\beta\in B} r_\alpha s_\beta (e_\alpha \otimes f_\beta)\right) \\ &= \sum_{\alpha\in A,\beta\in B} r_\alpha s_\beta\, \varphi(e_\alpha \otimes f_\beta) \\ &= \sum_{\alpha\in A,\beta\in B} r_\alpha s_\beta\, \rho(e_\alpha, f_\beta) = \rho(v,w), \end{aligned}$$

where the last equality follows from the K-bilinearity of ρ.

Diagram (4) of course is saying that the tensor product of vector spaces satisfies a universal property! "The map φ is the closest R-linear map to the R-bilinear map ρ!" It is now clear what we would want of the tensor product "$M \otimes_R N$" of R-modules over a general ring R. It should come equipped with an R-bilinear map $\tau : M \times N \to M \otimes_R N$ satisfying the following universal property:

For any R-module P and R-bilinear map $\rho : M \times N \to P$, there should be a unique R-linear map $\varphi : M \otimes_R N \to P$ such that the following diagram commutes:

(5)
$$\begin{array}{ccc} M \times N & \xrightarrow{\rho} & P \\ {\scriptstyle\tau}\downarrow & \nearrow_{\exists!\ \varphi} & \\ M \otimes_R N & & \end{array}$$

As with the other universal properties, this uniquely constrains the tensor product if it exists (see the following exercise). But we still have to construct it! Before it was quick to construct the object and then check it satisfied the universal property, but over time you will encounter more objects where it is much shorter to give the property characterising the object than construct it (and in fact it might not exist!). We know how the tensor product should behave even though we haven't found it!

EXERCISE 11.2 The tensor product is uniquely determined ("unique up to unique isomorphism") by the universal property. In other words, if there was another R-module T and R-bilinear map $\tau' : M \times N \to T$ such that there exists a map φ' as above for all ρ, then T is isomorphic to $M \otimes_R N$ by a unique R-module homomorphism. (Hint: look back at Remark 6.11.)

To define such a tensor product and τ, let $A = M \times N$, and let F_A be the free R-module on A, so its elements are sums of the form

$$\sum_{(m,n)\in A} r_{m,n} e_{m,n},$$

where each $r_{m,n} \in R$ and $r_{m,n} = 0$ for all but finitely many pairs $(m,n) \in A = M \times N$. We then set $M \otimes_R N = F_A/S$ where S is the R-submodule of F_A generated by the set of all elements of the form:

- $e_{m+m',n} - e_{m,n} - e_{m',n}$ for $m, m' \in M$, $n \in N$;
- $e_{m,n+n'} - e_{m,n} - e_{m,n'}$ for $m \in M$, $n, n' \in N$;
- $e_{rm,n} - r e_{m,n}$ for $r \in R$, $m \in M$, $n \in N$;
- $e_{m,rn} - r e_{m,n}$ for $r \in R$, $m \in M$, $n \in N$.

Alternatively put,

(6)
$$M \otimes_R N = \left\langle M \times N \;\middle|\; \begin{array}{l} e_{m+m',n} - e_{m,n} - e_{m',n}, \\ e_{m,n+n'} - e_{m,n} - e_{m,n'}, \\ e_{rm,n} - r e_{m,n}, \\ e_{m,rn} - r e_{m,n} \end{array} \right\rangle$$

We now define the function

$$\tau : M \times N \to M \otimes_R N$$

by $\tau(m,n) = \overline{e}_{m,n}$ (where if $x \in F_A$, then we write $\overline{x}$ for the element $x + S \in T = F_A/S$). We claim τ is in fact R-bilinear: Firstly since $e_{m+m',n} - e_{m,n} - e_{m',n} \in S$, we have

$$\tau(m+m',n) = \overline{e}_{m+m',n} = \overline{e}_{m,n} + \overline{e}_{m'n} = \tau(m,n) + \tau(m',n)$$

if $m, m' \in M$, $n \in N$, and similarly $\tau(m, n+n') = \tau(m,n) + \tau(m,n')$ if $m \in M$, $n, n' \in N$. Furthermore since $\overline{e}_{rm,n} = r\overline{e}_{m,n}$, we have

$$\tau(rm,n) = \overline{e}_{rm,n} = r\overline{e}_{m,n} = r\,\tau(m,n)$$

and similarly $\tau(m, rn) = r\,\tau(m,n)$ if $r \in R$, $m \in M$ and $n \in N$.

We now prove that (5) holds.

PROPOSITION 11.3 *Let M, N, T and τ be as above. If P is an R-module and $\rho : M \times N \to P$ is an R-bilinear map, then there is a unique R-linear map $\varphi\colon M \otimes_R N \to P$ such that $\rho = \varphi \circ \tau$.*

Proof. We must prove that there is a unique R-module homomorphism $\varphi : F_A/S \to P$ such that

$$\rho(m,n) = \varphi(\tau(m,n)) = \varphi(\overline{e}_{m,n})$$

for all $(m,n) \in M \times N = A$.

By Proposition 9.13, there is a unique R-module homomorphism $\widetilde{\varphi} : F_A \to P$ such that $\widetilde{\varphi}(e_{m,n}) = \rho(m,n)$ for all $m, n \in A$. In order to define φ, we will show that $S \subset \ker(\widetilde{\varphi})$.

Since ρ is R-bilinear, we have

$$\widetilde{\varphi}(e_{m+m',n}) = \rho(m+m',n) = \rho(m,n) + \rho(m',n) = \widetilde{\varphi}(e_{m,n}) + \widetilde{\varphi}(e_{m',n})$$

for all $m, m' \in M$, $n \in N$, and therefore $\widetilde{\varphi}(e_{m+m',n} - e_{m,n} - e_{m',n}) = 0$, i.e. $e_{m+m',n} - e_{m,n} - e_{m',n} \in \ker(\widetilde{\varphi})$. Similarly we find that

$$\widetilde{\varphi}(e_{m,n+n'}) = \widetilde{\varphi}(e_{m,n}) + \widetilde{\varphi}(e_{m,n'})$$

for all $m \in M$, $n, n' \in N$, and

$$\widetilde{\varphi}(e_{rm,n}) = r\,\widetilde{\varphi}(e_{m,n}) = \widetilde{\varphi}(e_{m,rn})$$

for all $r \in R$, $m \in M$, $n \in N$. It follows that the subset of F_A generating S is contained in $\ker(\widetilde{\varphi})$, and hence that $S \subset \ker(\widetilde{\varphi})$.

We may therefore define φ to be the composite

$$\begin{array}{ccccc} F_A/S & \longrightarrow & F_A/\ker(\widetilde{\varphi}) & \xrightarrow{\sim} & P \\ \overline{x} & \mapsto & x + \ker(\widetilde{\varphi}) & \mapsto & \widetilde{\varphi}(x), \end{array}$$

so $\varphi : F_A/S \to P$ is the unique homomorphism such that $\varphi(\overline{x}) = \widetilde{\varphi}(x)$ for all $x \in F_A$, i.e. $\widetilde{\varphi} = \varphi \circ \pi$ where $\pi : F_A \to F_A/S$ is defined by $x \mapsto \overline{x}$. In particular

$$\rho(m,n) = \widetilde{\varphi}(e_{m,n}) = \varphi(\overline{e}_{m,n})$$

for all $(m,n) \in A$, as required. Furthermore φ is the unique such R-module homomorphism: if $\varphi' : F_A/S \to P$ were another, then $\rho(m,n) = \varphi'(\overline{e}_{m,n}) = \varphi'(\pi(e_{m,n}))$ for all $(m,n) \in A$, so $\widetilde{\varphi} = \varphi' \circ \pi$ and hence $\varphi' = \varphi$. □

Note how the definition F_A/S is constructing an object and forcing all the properties (bilinearity of the map τ) we want onto it. This is a common method in modern mathematics.

DEFINITION 11.4 Suppose that M and N are R-modules. We call $M \otimes_R N$ the *tensor product of M and N (as R-modules)*. We write $m \otimes n$ for the element $\tau(m,n) \in M \otimes_R N$, where $\tau : M \times N \to M \otimes_R N$ is the R-bilinear map defined above.

We now have two very abstract descriptions of the tensor product $M \otimes_R N$ of two R-modules M and N. One in terms of a universal property and one in terms of a quotient of a very large group by a lot of relations. In the next section, we shall actually compute some concrete examples, but we first give one example we understand:

EXAMPLE 11.5 When $R = K$, the abstract construction of tensor product coincides with the original construction at the start of this section. We know this as we have seen that they both satisfy the universal property (4), so must be "the same"! More explicitly, if V, W are vector spaces with bases $\{e_\alpha \mid \alpha \in A\}$ and $\{f_\beta \mid \beta \in B\}$, then there is an isomorphism

$$\left\{ \sum_{i,j} r_{i,j}(e_i \otimes f_j) \right\} \xrightarrow{\sim} F_{M\times N}/S$$

given by

$$\sum_{i,j} r_{i,j}(e_i \otimes f_j) \longmapsto \sum_{i,j} r_{i,j} \bar{e}_{e_i,f_j}$$

(this formula is coming from the universal property).

EXERCISE* 11.6 The presentation (6) is clean in that it is general and doesn't require choices. It is, however, astronomically large, almost all of which is redundant. Suppose that M, N are generated by A, B respectively. Define an R-module of the form

$$\langle A \times B \mid R \rangle$$

for some choice of relations R such that there is a choice of τ for this module for which it satisfies the universal property of the tensor product (i.e. it is the tensor product). (*Hint: Maybe some good examples are sufficient in lieu of a complete proof. A starting point would be to compare the construction of* (6) *to the construction in the field case.*)

12 Tensor products: examples and properties

Suppose that M and N are R-modules. Recall that if $m \in M$ and $n \in N$, then $m \otimes n$ denotes the element $\bar{e}_{m,n} = \tau(m,n)$ of $M \otimes_R N = F_{M\times N}/S$. We call elements in the image of τ *simple tensors*. Recall that we have the following rules for simple tensors:

$$\begin{aligned} (m+m') \otimes n &= (m \otimes n) + (m' \otimes n), \\ m \otimes (n+n') &= (m \otimes n) + (m \otimes n'), \\ (rm) \otimes n &= r(m \otimes n) = m \otimes (rn). \end{aligned} \tag{7}$$

These come from the definition of S in the construction of $M \otimes_R N$ or more immediately from the R-bilinearity of τ. Since all elements of $M \otimes_R N$ are linear combinations of simple tensors, these rules completely describe how to manipulate elements of tensor products! Even more importantly, simple tensors are exactly those in the image of the map $\tau : M \times N \to M \otimes_R N$, so they are the ones whose image we understand completely under maps induced from the universal property (5) (they go where ρ tells them!).

Note in particular that

$$0_M \otimes n = (0_R 0_M) \otimes n = 0_R(0_M \otimes n) = 0_{M\otimes N}$$

for all $n \in N$, and similarly $m \otimes 0_N = 0_{M\otimes N}$ for all $m \in M$.

EXAMPLE 12.1 Let $R = \mathbb{Z}$. We claim that

$$\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z} = 0.$$

Since all elements are linear combinations of simple tensors, it suffices to show that all simple tensors are equivalent to zero by the relations (7). Let $a, b \in \mathbb{Z}$. Then $3a \equiv a \pmod 2$. So

$$a \otimes b = 3a \otimes b = a \otimes 3b = a \otimes 0 = 0,$$

since $3b \equiv 0 \pmod 3$.

The above example was quite constructive, using the definition of the tensor product as F_A/S. The universal property equally useful for explicit computations:

EXAMPLE 12.2 Let $R = \mathbb{Z}$. We claim that

$$(\mathbb{Z}/m\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/n\mathbb{Z})$$

is canonically isomorphic to $\mathbb{Z}/\gcd(m,n)\mathbb{Z}$, defined on simple tensors by $\varphi(a \otimes b) = ab$ (if this is well-defined, this uniquely defines φ as every element is a linear combination of simple tensors). To produce an R-module map out of a tensor product, we are always looking to use the universal property (5). In this case, we use that there is a bilinear map

$$\begin{aligned} \rho : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} &\longrightarrow \mathbb{Z}/\gcd(m,n)\mathbb{Z} \\ (a,b) &\longmapsto ab. \end{aligned}$$

For example,

$$\rho((a+a') \otimes b) = (a+a')b = ab + ab' = \rho(a \otimes b) + \rho(a' \otimes b),$$

where the middle equality uses that $\gcd(m,n)$ divides m (why?). The diagram

$$\begin{array}{ccc} \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} & \xrightarrow{\rho} & N \\ \downarrow{\scriptstyle \tau} & \nearrow_{\exists!\ \varphi} & \\ \mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z} & & \end{array}$$

We claim that φ is the inverse of ψ. Indeed,

$$\varphi(\psi(n)) = \varphi(1 \otimes n) = \varphi(\tau(1,n)) = \rho(1,n) = n.$$

then immediately gives the map φ (in particular it is a well-defined R-module homomorphism). It is also clear that φ is surjective as $\varphi(a \otimes 1) = a$. It remains to show φ is injective. To see this, note that any element of $\mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z}$ is of the form $a \otimes 1$. This is because, using (7), for any simple tensor

$$x \otimes y = x \otimes (y \cdot 1) = xy \otimes 1,$$

whilst any sum of such tensors can be contracted

$$z \otimes 1 + z' \otimes 1 = (z+z') \otimes 1.$$

So we need only show that if $a \equiv a' \pmod{\gcd(m,n)}$, then $a \otimes 1 = a' \otimes 1$. Or equivalently, if $a \equiv 0 \pmod{\gcd(m,n)}$, then $a \otimes 1 = 0$. But, under this assumption, there exist $\alpha, \beta \in \mathbb{Z}$ such that

$$\alpha n + \beta m = a.$$

So

$$\begin{aligned} a \otimes 1 &= (\alpha n + \beta m) \otimes 1 \\ &= (\alpha n) \otimes 1 + (\beta m) \otimes 1 \\ &= (\alpha n) \otimes 1 + 1 \otimes (\beta m) \\ &= 0. \end{aligned}$$

EXAMPLE 12.3 Consider the case where $M = R$. We claim that $R \otimes_R N$ is isomorphic to N (as an R-module). To see this define $\psi : N \to R \otimes_R N$ by $\psi(n) = 1 \otimes n$ (writing 1 for 1_R). Then ψ is R-linear since

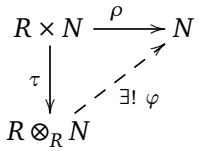
$$\psi(n+n') = 1 \otimes (n+n') = (1 \otimes n) + (1 \otimes n') = \psi(n) + \psi(n')$$

for all $n, n' \in N$, and $\psi(rn) = 1 \otimes (rn) = r(1 \otimes n) = r\psi(n)$ for all $r \in R$, $n \in N$.

To prove that ψ is an isomorphism, we shall use the universal property (Proposition 11.3) to show that ψ has an inverse. To start with, we saw in Example 10.5 that there is an R-bilinear map

$$\begin{aligned} \rho : R \times N &\longrightarrow N \\ (r,n) &\longmapsto rn. \end{aligned}$$

So by the universal property, there is an R-linear map φ such that:



We claim that φ is the inverse of ψ. Indeed,

$$\varphi(\psi(n)) = \varphi(1 \otimes n) = \varphi(\tau(1,n)) = \rho(1,n) = n.$$

Now to show that $\psi \circ \varphi = \mathrm{id}_{R \otimes_R N}$, first note that since every element of $R \otimes_R N$ is a linear combination of simple tensors and all maps are R-module homomorphisms, it suffices to show equality on elements of the form $r \otimes n$. Now

$$\psi(\varphi(r \otimes n)) = \psi(\varphi(\tau(r,n))) = \psi(\rho(r,n)) = \psi(rn) = 1 \otimes rn,$$

but $1 \otimes rn = r \otimes n$ by (7), so we win!

EXERCISE 12.4 More generally, consider the case when M is the R-module R/I, where I is an ideal of R. Show that if N is an R-module, then

$$(R/I) \otimes_R N$$

is isomorphic to N/IN, where IN is the submodule of N generated by the subset $\{\, an \mid a \in I, n \in N \,\}$. Use this to quickly prove Example 12.2.

LEMMA 12.5 *Suppose that $\theta : M \to M'$ and $\psi : N \to N'$ are homomorphisms of R-modules. Then there is a unique R-module homomorphism*

$$\varphi : M \otimes_R N \longrightarrow M' \otimes_R N'$$

such that $\varphi(m \otimes n) = \theta(m) \otimes \psi(n)$ for all $m \in M$, $n \in N$.

Proof. As usual, we will apply Proposition 11.3. We just need to show that there is an R-bilinear map

$$\rho : M \times N \longrightarrow P = M' \otimes_R N'$$

defined by $\rho(m,n) = \theta(m) \otimes \psi(n)$.

To that ρ, so defined, is R-bilinear, note that

$$\begin{aligned} \rho(m_1 + m_2, n) &= \theta(m_1 + m_2) \otimes \psi(n) \\ &= (\theta(m_1) + \theta(m_2)) \otimes \psi(n) \\ &= (\theta(m_1) \otimes \psi(n)) + (\theta(m_2) \otimes \psi(n)) \\ &= \rho(m_1,n) + \rho(m_2,n) \end{aligned}$$

for all $m_1, m_2 \in M$, $n \in N$. Similarly we find that

$$\rho(m, n_1 + n_2) = \rho(m,n_1) + \rho(m,n_2)$$

for all $m \in M$, $n_1, n_2 \in N$, and $\rho(rm,n) = r\rho(m,n) = \rho(m,rn)$ for all $r \in R$, $m \in M$, $n \in N$.

Applying Proposition 11.3, we conclude that there is a unique R-linear map $\varphi : M \otimes_R N \to P$ such that

$$\varphi(m \otimes n) = \varphi(\tau(m,n)) = \rho(m,n) = \theta(m) \otimes \psi(n)$$

for all $m \in M$, $n \in N$. □

EXAMPLE 12.6 Suppose (in the situation of Lemma 12.5) that θ and ψ are isomorphisms, and let $\zeta : M' \to M$ and $\xi : N' \to N$ be their inverses. Applying the lemma to θ and ψ gives an R-linear homomorphism $\varphi : M \otimes_R N \to M' \otimes_R N'$, and we similarly obtain a homomorphism $\chi : M' \otimes_R N' \to M \otimes_R N$ from ζ and ξ. The composite $\chi \circ \varphi : M \otimes_R N \to M \otimes_R N$ sends $m \otimes n$ to $\theta(\zeta(m)) \otimes \psi(\xi(n)) = m \otimes n$ for all $m \in M$, $n \in N$, so (by the uniqueness in Proposition 11.3) it must be the identity. Similarly we see that $\varphi \circ \chi$ is the identity, so φ and χ are isomorphisms.

Tensor product and direct sums

We are now going to prove that tensor products "commute with direct sums".

First consider the case of the direct sum of just two modules. Specifically, let M, N_1 and N_2 be R-modules, and let $N = N_1 \oplus N_2$. Applying Lemma 12.5 with $M' = M$, $N' = N_1$, $\theta = \mathrm{id}_M$ and $\psi = \pi_1 : N \to N_1$ defined by $\pi_1(n_1, n_2) = n_1$ (i.e. projection onto the first component) yields an R-module homomorphism

$$\varphi_1 : M \otimes_R N \longrightarrow M \otimes N_1$$

such that $\varphi_1(m \otimes (n_1, n_2)) = m \otimes n_1$ for all $m \in M$, $n_1 \in N_1$, $n_2 \in N_2$.

Similarly, with the same M and N as above but replacing $\psi = \pi_1$ by $\pi_2 : N \to N_2$ (defined by $\pi_2(n_1, n_2) = n_2$), we obtain an R-module homomorphism $\varphi_2 : M \otimes_R N \longrightarrow M \otimes N_2$ such that $\varphi_2(m \otimes (n_1, n_2)) = m \otimes n_2$ for all $m \in M$, $n_1 \in N_1$, $n_2 \in N_2$. We can then define

$$\varphi : M \otimes_R N \longrightarrow (M \otimes N_1) \oplus (M \otimes N_2),$$

by $\varphi(t) = (\varphi_1(t), \varphi_2(t))$, so that $\varphi(m \otimes (n_1, n_2)) = (m \otimes n_1, m \otimes n_2)$ for all $m \in M$, $n_1 \in N_1$, $n_2 \in N_2$.

Similarly, letting $\psi = \iota_1 : N_1 \to N$ be the homomorphism defined by $\iota_1(n_1) = (n_1, 0)$, we obtain a homomorphism

$$\chi_1 : M \otimes N_1 \longrightarrow M \otimes N$$

sending $m \otimes n_1$ to $m \otimes (n_1, 0)$ for $m \in M$, $n_1 \in N$. Likewise there is a homomorphism $\chi_2 : M \otimes N_2 \longrightarrow M \otimes N$ sending $m \otimes n_2$ to $m \otimes (0, n_2)$, so we obtain a homomorphism

$$\chi : (M \otimes_R N_1) \oplus (M \otimes_R N_2) \longrightarrow M \otimes N$$

such that $\chi(t_1, t_2) = \chi_1(t_1) + \chi_2(t_2)$ (recall maps out of direct products are just choices of maps on each component by Proposition 9.7).

Unsurprisingly, φ and χ are inverses of each other, and hence isomorphisms. To see that $\chi \circ \varphi$ is the identity, note that it is an R-linear homomorphism $M \otimes_R N \longrightarrow M \otimes_R N$ sending simple tensors $m \otimes (n_1, n_2)$ to

$$\begin{aligned} \chi(\varphi(m \otimes (n_1, n_2))) &= \chi(m \otimes n_1, m \otimes n_2) = \chi_1(m \otimes n_1) + \chi_2(m \otimes n_2) \\ &= (m \otimes (n_1, 0)) + (m \otimes (0, n_2)) \\ &= m \otimes ((n_1, 0) + (0, n_2)) = m \otimes (n_1, n_2) \end{aligned}$$

for all $m \in M$, $(n_1, n_2) \in N$. Since every element of $M \otimes_R N$ is a linear combination of simple tensors, $\chi \circ \varphi$ is the identity and we win!

In the other direction, to prove that $\varphi \circ \chi$ is the identity, we must show that[24]

$$\varphi(\chi(m \otimes n_1, 0)) = (m \otimes n_1, 0) \quad \text{and} \quad \varphi(\chi(0, m \otimes n_2)) = (0, m \otimes n_2)$$

for all $m \in M$, $n_1 \in N_1$, $n_2 \in N_2$. Now

$$\begin{aligned} \varphi(\chi(m \otimes n_1, 0)) &= \varphi(\chi_1(m \otimes n_1)) \\ &= \varphi(m \otimes (n_1, 0)) \\ &= (m \otimes n_1, m \otimes 0) \\ &= (m \otimes n_1, 0), \end{aligned}$$

as desired! The check for elements of the form $(0, m \otimes n_2)$ is identical and $\varphi \circ \chi = \mathrm{id}$ also.

We have now shown (rather tediously) that there is a "natural" isomorphism

$$M \otimes_R (N_1 \oplus N_2) \xrightarrow{\sim} (M \otimes_R N_1) \oplus (M \otimes_R N_2)$$

for any R-modules M, N_1, and N_2. Similarly, for any R-modules M_1, M_2 and N, there is an isomorphism

$$(M_1 \oplus M_2) \otimes_R N \xrightarrow{\sim} (M_1 \otimes_R N) \oplus (M_2 \otimes_R N_2)$$

sending $(m_1, m_2) \otimes n$ to $(m_1 \otimes n, m_2 \otimes n)$ for $m_1 \in M_1$, $m_2 \in M_2$, $n \in N$. By induction argument gives an isomorphism

$$\left(\bigoplus_{i=1}^{m} M_i\right) \otimes_R \left(\bigoplus_{j=1}^{n} N_j\right) \xrightarrow{\sim} \bigoplus_{i,j} (M_i \otimes_R N_j), \tag{8}$$

for any R-modules $M_1, M_2, \ldots, M_m, N_1, \ldots, N_n$ (where $i = 1, \ldots, m$ and $j = 1, \ldots, n$ in the product on the right). But this inductive argument can never deal with infinite direct sums!

EXERCISE* 12.7 In fact tensor products commute with arbitrary (i.e. infinite) direct sums. Show this! Since finite direct sums and finite products are isomorphic, the statement for finite products also holds. But the statement for infinite products does not. Can you give an example (hard?)?

EXAMPLE 12.8 Letting $M_1 = \cdots = M_m = N_1 = \cdots = N_n = R$, we see from (8) (and the fact that $R \otimes_R R \cong R$) that there is an isomorphism

$$R^m \otimes_R R^n \xrightarrow{\sim} R^{mn}.$$

(Recall we already saw this in the case that K is a field.)

More generally if A and B are any index sets, then Exercise 12.7 gives an isomorphism

$$F_A \otimes_R F_B \xrightarrow{\sim} F_{A \times B}.$$

It follows that if M and N are free R-modules, then so is $M \otimes_R N$ (strictly speaking, we are using Example 12.6 here).

We state a few other general properties of tensor products whose proofs we leave as exercises.

PROPOSITION 12.9 *Suppose that M, N and P are R-modules.*

[24] Note that we have just been writing 0 for 0_{N_2}, as well as for 0_{N_1}, and now also for $0_{M \otimes_R N_2}$, etc.

(i) *There is a unique isomorphism*

$$\varphi : M \otimes_R N \xrightarrow{\sim} N \otimes_R M$$

such that $\varphi(m \otimes n) = n \otimes m$ *for all* $m \in M$, $n \in N$.

(ii) *There is a unique isomorphism*

$$\psi : M \otimes_R (N \otimes_R P) \xrightarrow{\sim} (M \otimes_R N) \otimes_R P$$

such that $\psi(m \otimes (n \otimes p)) = (m \otimes n) \otimes p$ *for all* $m \in M$, $n \in N$, $p \in P$.

Note that the first part of the proposition describes a sense in which the tensor product is commutative, and the second a sense in which it is associative[25]. In particular, the associativity property allows us to drop the parentheses and write simply $M \otimes_R N \otimes_R P$ instead of $M \otimes_R (N \otimes_R P)$ (which the isomorphism ψ in the proposition identifies with $(M \otimes_R N) \otimes_R P$), or more generally $M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_n$ for the tensor product of any R-modules $M_1, M_2, \ldots, M_n$. Furthermore the commutativity property allows us to interchange the order of the factors in the tensor product. Finally we remark that Proposition 11.3 generalizes to R-multilinear maps, so that giving an R-multilinear map

$$M_1 \times M_2 \times \cdots \times M_n \longrightarrow N$$

is equivalent to giving an R-linear map $M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_n \to N$.

13 Functors

We will now introduce another useful notion from category theory, namely *functors*. Functors are "maps of categories".

EXAMPLE 13.1 Let $f : R \to S$ be a homomorphism of rings. We claim that given an S-module M, we can also view M as an R-module using f. Specifically, we can let R act on M by

$$r \cdot m = f(r)m \qquad \text{for } r \in R,\ m \in M,$$

where $f(r)$ is acting as an element of S on M. Recall that M is already assumed to be an abelian group (as part of the definition of an S-module), and it is straightforward to check that the axioms in Definition 8.1 are satisfied. For example, (iii) holds since

$$(r + r') \cdot m = f(r + r')m = (f(r) + f(r'))m = f(r)m + f(r')m = r \cdot m + r' \cdot m$$

for all $r, r' \in R$ and $m \in M$ (where the first and last equalities are by definition, the second follows from f being a homomorphism, and the third is (iii) for M as an S-module).

Recall that we already know that R-mod and S-mod are categories. We want to promote the assignment of Example 13.1 to a "functor" S-mod $\to R$-mod. But categories are more than just objects. To really be saying something, we need instructions of what to do with a morphism of S-modules also.

We claim that if $\varphi : M \to N$ is a homomorphism of S-modules, then when we consider M, N as R-modules via Example 13.1, it is also a homomorphism of R-modules. Indeed φ is a homomorphism of abelian groups, and

$$\varphi(r \cdot m) = \varphi(f(r)m) = f(r)\varphi(m) = r \cdot \varphi(m)$$

[25] In a similar vein, one can view (8) as a distributivity property.

for all $r \in R$, $m \in M$. Therefore we have a map

$$\mathrm{Hom}_S(M,N) \to \mathrm{Hom}_R(M,N).$$

This assignment of an S-module to an R-module and an S-module homomorphism to a homomorphism of the corresponding R-modules *is* an example of a functor S-mod $\to$ R-mod:

DEFINITION 13.2 Suppose that $\mathcal{C}$ and $\mathcal{D}$ are categories. A *functor* $F : \mathcal{C} \to \mathcal{D}$ is a rule that associates an object $F(A)$ of $\mathcal{D}$ to each object A of $\mathcal{C}$, and a function

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{C}}(A,B) & \longrightarrow & \mathrm{Hom}_{\mathcal{D}}(F(A),F(B)) \\ \varphi & \longmapsto & F(\varphi) \end{array}$$

to each pair of objects A, B of $\mathcal{C}$ such that the following hold:

(i) if A is an object of $\mathcal{C}$, then $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$;

(ii) if $A \xrightarrow{\varphi} B$ and $B \xrightarrow{\psi} C$ are morphisms in $\mathcal{C}$, then

$$F(\psi \circ \varphi) = F(\psi) \circ F(\varphi).$$

Thus a functor $F : \mathcal{C} \to \mathcal{D}$ sends objects and morphisms in $\mathcal{C}$ to ones of $\mathcal{D}$ in a way that's compatible with the structure of the categories, and in particular composition. Note also that the equation in part (ii) of the definition makes sense since the composite

$$F(A) \xrightarrow{F(\varphi)} F(B) \xrightarrow{F(\psi)} F(C)$$

is a morphism in $\mathrm{Hom}_{\mathcal{D}}(F(A),F(C))$, as is $F(\psi \circ \varphi)$ (since $\psi \circ \varphi \in \mathrm{Hom}_{\mathcal{C}}(A,C)$.

EXAMPLE 13.3 Suppose that $f : R \to S$ is a homomorphism of rings, then we really do obtain a functor $F : S$-mod $\to R$-mod by setting $F(M)$ to be M thought of as an R-module as in Example 13.1 and setting

$$\begin{array}{c} \mathrm{Hom}_S(M,N) \longrightarrow \mathrm{Hom}_R(M,N) \\ \varphi \mapsto F(\varphi) := \varphi \quad \begin{pmatrix} \text{considered as an } R\text{-module} \\ \text{homomorphism} \end{pmatrix}. \end{array}$$

We need only check that $F(\mathrm{id}_M) = \mathrm{id}_M = \mathrm{id}_{F(M)}$ for all S-modules M, and $F(\psi \circ \varphi) = \psi \circ \varphi = F(\psi) \circ F(\varphi)$ for all S-linear maps $\varphi : M \to N$ and $\psi : N \to P$, which they are!

(We remark also that nowhere did we use that R and S are commutative; we could similarly have defined a functor from the category of left (or right) S-modules to that of left (or right) R-modules, without assuming commutativity of the rings.)

EXAMPLE 13.4 We define a functor Gps $\to$ Sets as follows. If G is a group, then define $F(G)$ to be the underlying set G (forgetting the operation, say $*$, on G, so we should really write $F((G,*)) = G$), and if $\varphi : G \to H$ is a morphism in Gps (i.e. a group homomorphism), we define $F(\varphi)$ to be the function $\varphi \in \mathrm{Hom}_{\mathsf{Sets}}(G,H)$. Just as in Example 13.3, identities map to identities and compositions to compositions, so that (i) and (ii) in Definition 13.2 are satisfied. This is an example of a *forgetful functor* (these are functors defined by forgetting some part of the structure, these are often denoted U).

In the preceding examples, the objects of $\mathcal{C}$ and $\mathcal{D}$ were sets with additional structure, the morphisms were structure-preserving functions, and the functions

$$\mathrm{Hom}_{\mathcal{C}}(A,B) \longrightarrow \mathrm{Hom}_{\mathcal{D}}(F(A),F(B))$$

were the obvious inclusions. We now give examples where this isn't the case:

EXAMPLE 13.5 Let $\mathcal{C}$ be the category with two objects and three morphisms defined in Example 5.5 by the following diagram:

$$\mathrm{id}_A \circlearrowright A \xrightarrow{\ f\ } B \circlearrowright \mathrm{id}_B\ .$$

Let $\mathcal{D}$ be the category with one object D and one morphism:

$$\mathrm{id}_D \circlearrowright D\ .$$

There is a functor $F : \mathcal{C} \to \mathcal{D}$ defined by setting $F(A), F(B) = D$ and $F(\mathrm{id}_A), F(A), F(B) = \mathrm{id}_D$. We certainly have that $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ for every $X \in \mathrm{Ob}(\mathcal{C})$. The only compositions to check are that

$$\begin{aligned} F(\mathrm{id}_A) \circ F(\mathrm{id}_A) &= F(\mathrm{id}_A), \\ F(\mathrm{id}_B) \circ F(\mathrm{id}_B) &= F(\mathrm{id}_B) \\ F(f) \circ F(\mathrm{id}_A) &= F(f) \\ F(\mathrm{id}_B) \circ F(f) &= F(f) \end{aligned}$$

These all do hold as every one is id_D and $\mathrm{id}_D \circ \mathrm{id}_D = \mathrm{id}_D$.

EXAMPLE 13.6 Let K be a field, let $\mathcal{C}$ be the category K-vec of vector spaces over K, and let $\mathcal{D}$ denote the category defined as in Example 5.7, so its objects are vector spaces over K, but if V and W are objects, then $\mathrm{Hom}_{\mathcal{D}}(V, W)$ is the set of equivalence classes of K-linear maps $\varphi : V \to W$, where $\varphi \sim \varphi'$ if $\varphi' = r\varphi$ for some $r \in K^\times$. In other words,

$$Hom_{\mathcal{D}}(V, W) = K^\times \backslash \mathrm{Hom}_{\mathcal{C}}(V, W)$$

is the set of orbits in $\mathrm{Hom}_{\mathcal{C}}(V, W)$ under the obvious action of $K^\times$.

We can therefore define a functor $F : \mathcal{C} \to \mathcal{D}$ by $F(V) = V$ for objects V of $\mathcal{C}$, and $F(\varphi) = [\varphi] = K^\times \cdot \varphi$ for morphisms $\varphi \in \mathrm{Hom}_{\mathcal{C}}(V, W)$ (which is again easily seen to satisfy (i) and (ii) in Definition 13.2).

EXERCISE* 13.7 Show that the image of a functor $f : \mathcal{C} \to \mathcal{D}$ does not have to be a sub-category. Consequently, the notion of the image of a functor is of almost no use. (Here, we mean the collection of objects and morphisms in the image under f of objects and morphisms of $\mathcal{C}$. By does not have to be a sub-category, we mean with the composition rules for $\mathcal{D}$, this collection does not satisfy the conditions of being a category.) *Hint: try and construct a small example using categories defined by diagrams like in Example 5.5.*

EXAMPLE 13.8 Let R be a (commutative) ring, and let M be a (fixed) R-module. We will define a functor $F : R\text{-mod} \to R\text{-mod}$ by "tensoring with M." More precisely, if N is an R-module, then we let $F(N) = M \otimes_R N$, and if $\psi : N \to N'$ is a homomorphism of R-modules, then

$$F(\psi) : F(N) = M \otimes_R N \longrightarrow M \otimes_R N' = F(N')$$

is the R-linear map obtained by applying Lemma 12.5 to the pair of morphisms $\theta = \mathrm{id}_M : M \to M$ and $\psi : N \to N'$. We will denote this R-module homomorphism $\mathrm{id}_M \otimes \psi$ (and more generally, we will start systematically writing $\theta \otimes \psi$ for the R-linear map $M \otimes_R N \to M' \otimes_R N'$ obtained by applying the lemma to $\theta : M \to M'$ and $\psi : N \to N'$). Since $F(\mathrm{id}_N) = \mathrm{id}_M \otimes \mathrm{id}_N$ is the unique R-module homomorphism $M \otimes_R N \to M \otimes_R N$ sending $m \otimes n$ to $m \otimes n$ for all $m \in M$, $n \in N$, it must be $\mathrm{id}_{M \otimes_R N}$. To conclude that F is a functor, note that if $\psi : N \to N'$ and $\psi' : N' \to N''$ are R-linear maps, then

$$F(\psi' \circ \psi) = \mathrm{id}_M \otimes (\psi' \circ \psi) : M \otimes_R N \longrightarrow M \otimes_R N''$$

is the unique R-linear map sending $m \otimes n$ to

$$m \otimes \psi'(\psi(n)) = (\mathrm{id}_M \otimes \psi')(m \otimes \psi(n)) = (\mathrm{id}_M \otimes \psi')((\mathrm{id} \otimes \psi)(m \otimes n))$$

for all $m \in M$, $n \in N$, so it coincides with

$$(\mathrm{id}_M \otimes \psi') \circ (\mathrm{id}_M \otimes \psi) = F(\psi') \circ F(\psi).$$

(More generally, the same reasoning shows that

$$(\theta' \otimes \psi') \circ (\theta \otimes \psi) = (\theta' \circ \theta) \otimes (\psi' \circ \psi)$$

for any R-linear maps $\theta : M \to M'$, $\theta : M' \to M''$, $\psi : N \to N'$ and $\psi' : N' \to N''$.)

Suppose now that $f : R \to S$ is a ring homomorphism. Recall from Example 13.3 that viewing each S-module as an R-module gives rise to a functor $F : S\text{-mod} \to R\text{-mod}$. We now define a functor in the other direction, $G : R\text{-mod} \to S\text{-mod}$, based on the construction in Example 13.8. More precisely, suppose that N is an R-module. Since we can view S as an R-module via f, we can define the R-module $S \otimes_R N$. We claim that this is actually also an S-module "by acting on the first coordinate". More specifically, we claim that letting $s \in S$ act on a simple tensor by

$$s \cdot (s' \otimes n) := (ss') \otimes n$$

and "extending linearly" gives a well-defined S-module structure. We first check that for each s, this assignment is well-defined, that is there is a well-defined map $s \cdot (-) : S \otimes_R N \to S \otimes_R N$ that on simple tensors is given by the above formula[26]. Because of the identification $S \otimes_R N = F_{S \times N}/(\text{relations})$, we need only check that all the relations (7) are respected (so that it descends to a well-defined map by the first isomorphism theorem) by the above formula. For example, we know for any $r \in R$ that $rs' \otimes n = s' \otimes rn$, so we need

$$s \cdot (rs' \otimes n) = s \cdot (s' \otimes rn).$$

But

$$\begin{aligned} s \cdot (rs' \otimes n) &:= (sr's \otimes n) \\ &= ss' \otimes rn \\ &=: s \cdot (s' \otimes rn), \end{aligned}$$

as desired. Here the middle equality is as S is commutative and by the rules for $M \otimes_R N$ (note we could only do this move for elements $r \in R$, but that's all we needed to do as the relations are as R-modules). This actually shows a stronger result, that the map $s \cdot (-) : S \otimes_R N \to S \otimes_R N$ is an R-module homomorphism!

We now want to check this really gives $S \otimes_R N$ the structure of an S-module. We have four axioms to check. Normally we would need to check them on arbitrary elements of $S \otimes_R N$, i.e. elements of the form $\sum_i s_i \otimes n_i$, but actually because we know $s \cdot (-)$ is an R-module homomorphism, it suffices to only check the axioms of Definition 8.1 for simple tensors (why?):

(i) $1_S \cdot (s' \otimes n) = (1_S s') \otimes n = s' \otimes n$,

(ii) $s \cdot (s_1 \otimes n_1 + s_2 \otimes n_2) = ss_1 \otimes n_1 + ss_2 \otimes n_2 = s \cdot (s_1 \otimes n_1) + s \cdot (s_2 \otimes n_2)$,

(iii) $(s_1 + s_2) \cdot (s' \otimes n) = ((s_1 + s_2)s) \otimes n = (s_1 s + s_2 s) \otimes n = s_1 s \otimes n + s_2 s \otimes n = s_1 \cdot (s \otimes n) + s_2 \cdot (s \otimes n)$,

(iv) $s_1 \cdot (s_2 \cdot n) = s_1((s_2 s) \otimes n) = (s_1 s_2 s) \otimes n = (s_1 s_2) \cdot (s \otimes n)$.

[26] It may be useful to revisit Remark 5.16 here!

Here we mostly used the axioms for the ring S. We now define $G(N)$ to be the S-module $S \otimes_R N$.

To complete the definition of the functor G, we must specify its effect on morphisms. If $\psi : N \to N'$ is a homomorphism of R-modules, then consider the homomorphism

$$\mathrm{id}_S \otimes \psi : S \otimes_R N \longrightarrow S \otimes_R N',$$

a priori R-linear, but we claim that it is in fact S-linear, i.e. that $(\mathrm{id}_S \otimes \psi)(s \cdot m) = s \cdot ((\mathrm{id}_S \otimes \psi)(m))$ for all $s \in S$, $m \in S \otimes_R N$. Note that this is the same as saying that $(\mathrm{id}_S \otimes \psi) \circ \lambda_s = \lambda'_s \circ (\mathrm{id}_S \otimes \psi)$ for all $s \in S$, where $\lambda_s = \mu_s \otimes \mathrm{id}_N$ and $\lambda'_s = \mu_s \otimes \mathrm{id}_{N'}$. This holds since

$$(\mathrm{id}_S \otimes \psi) \circ (\mu_s \otimes \mathrm{id}_N) = \mu_s \otimes \psi = (\mu_s \otimes \mathrm{id}_{N'}) \circ (\mathrm{id}_S \otimes \psi)$$

(see the discussion at the end of Example 13.8). We can therefore define $G(\psi) = \mathrm{id}_S \otimes \psi$.

Finally the conditions in Definition 13.2 are satisfied since

$$G(\mathrm{id}_N) = \mathrm{id}_S \otimes \mathrm{id}_N = \mathrm{id}_{S \otimes_R N} = \mathrm{id}_{G(N)}$$

for all R-modules N, and

$$G(\psi' \circ \psi) = \mathrm{id}_S \otimes (\psi' \circ \psi) = (\mathrm{id}_S \otimes \psi') \circ (\mathrm{id}_S \otimes \psi) = G(\psi') \circ G(\psi)$$

for all R-module homomorphisms $\psi : N \to N'$, $\psi' : N' \to N''$.

EXERCISE 13.9 There is a slick way to define the action of S on $S \otimes_R N$ by starting with the action maps $s \cdot (-) \colon S \to S$ coming from the action of S on itself (Example 8.4). Since these are R-linear, we can use Lemma 12.5 to obtain "action by s" maps $(s \cdot (-)) \otimes \mathrm{id}_N : S \otimes_R N \to S \otimes_R N$. Check the details!

EXAMPLE 13.10 Recall from Example 12.3 (and part (i) of Proposition 12.9) that for any R-module M, we have an R-linear isomorphism $M \cong M \otimes_R R$ under which $m \in M$ corresponds to $m \otimes 1$. In particular there is an R-linear isomorphism $\varphi : S \otimes_R R \xrightarrow{\sim} S$ under which $t \otimes r \mapsto tr$. The isomorphism is in fact S-linear, since

$$s \cdot \varphi(r \otimes t) = srt = \varphi(sr \otimes t)$$

for all $s, t \in S$ and $r \in R$. Therefore $G(R)$ is isomorphic to S as an S-module.

EXAMPLE 13.11 Recall for the discussion before (8) that if M, N_1 and N_2 are R-modules, we have an R-linear isomorphism

$$M \otimes_R (N_1 \times N_2) \xrightarrow{\sim} (M \otimes_R N_1) \times (M \otimes_R N_2).$$

In particular $S \otimes_R (N_1 \times N_2)$ is isomorphic to $(S \otimes_R N_1) \times (S \otimes_R N_2)$. Again the isomorphism is easily seen to be S-linear (since it is defined by $x \mapsto (G(\pi_1)x, G(\pi_2)x)$ where π_i is the projection $N_1 \times N_2 \to N_i$ for $i = 1, 2$), so we have $G(N_1 \times N_2) \cong G(N_1) \times G(N_2)$ as S-modules. In particular it follows that

$$G(R^n) \cong G(R)^n \cong S^n$$

as S-modules (using Example 13.10 for the final isomorphism).

Without tensor products we have no obvious way to obtain an S-module from an R-module. This is one of many applications of tensor products within commutative algebra.

EXERCISE 13.12 The functor G we have just constructed is related to the functor F of Example 13.3. We leave it as an exercise to show that there is a natural bijection[27]

$$\mathrm{Hom}_S(G(N), M) \longleftrightarrow \mathrm{Hom}_R(N, F(M))$$

[27] Functors satisfying this type of relation are called *adjoint* functors.

for all R-modules N and S-modules M.

Finally we describe one more general construction involving tensor products. Suppose now that $f : R \to S$ and $g : R \to T$ are two ring homomorphisms (from the same ring R). As usual we can view S and T as R-modules via f and g, and so we can define the R-module $S \otimes_R T$. Being an R-module, it is an abelian group under addition, but we can also define a multiplication operation on $S \otimes_R T$. This multiplication, being a binary operation, can be viewed as a function

$$(S \otimes_R T) \times (S \otimes_R T) \longrightarrow S \otimes_R T.$$

We will define the function as an R-bilinear map, obtained as a composite

$$(S \otimes_R T) \times (S \otimes_R T) \xrightarrow{\tau} (S \otimes_R T) \otimes_R (S \otimes_R T) \xrightarrow{\varphi} S \otimes_R T,$$

where as usual τ is the (R-bilinear) map defined by $(m,n) \mapsto m \otimes n$ for $m, n \in S \otimes_R T$, and φ will be an R-linear map. Recall that to give an R-linear map $S \otimes_R T \otimes_R S \otimes_R T \to S \otimes_R T$ is equivalent[28] to giving an R-quadrilinear map $\rho : S \times T \times S \times T \to S \otimes_R T$, which we define by $\rho(s,t,s',t') = (ss') \otimes (tt')$ for $s, s' \in S$, $t, t' \in T$. It is straighforward to check that ρ is indeed R-quadrilinear; for example

$$\rho(r \cdot s, t, s', t') = (r \cdot ss') \otimes (tt') = r \cdot ((ss') \otimes (tt')) = r \cdot \rho(s,t,s',t')$$

for all $r \in R$, $s, s' \in S$, $t, t' \in T$, and

$$\begin{aligned} \rho(s_1 + s_2, t, s', t') &= ((s_1 + s_2)s') \otimes (tt') = (s_1 s' + s_2 s') \otimes (tt') \\ &= ((s_1 s') \otimes (tt')) + ((s_2 s') \otimes (tt')) = \rho(s_1, t, s', t') + \rho(s_2, t, s', t') \end{aligned}$$

if $s_1, s_2, s' \in S$, $t, t' \in T$. Note that the resulting map φ sends an element of $(S \otimes_R T) \otimes_R (S \otimes_R T)$ of the form

$$\tau(s \otimes t, s' \otimes t') = s \otimes t \otimes s' \otimes t'$$

to $(ss') \otimes (tt')$, so the resulting multiplication operation on $S \otimes_R T$ can be described more concretely by the formula $(s \otimes t)(s' \otimes t') = (ss') \otimes (tt')$. Recall that elements of $S \otimes_R T$ are not necessarily of the form $s \otimes t$, but can always be written as R-linear combinations (or even sums) of such elements, for which it follows that multiplication is defined by

$$\left(\sum_{i=1}^{m} r_i \cdot (s_i \otimes t_i) \right) \left(\sum_{j=1}^{n} r'_j \cdot (s'_j \otimes t'_j) \right) = \sum_{i,j} r_i r'_j \cdot ((s_i s'_j) \otimes (t_i t'_j)).$$

We leave it as an exercise to check that with multiplication so defined, $1 \otimes 1$ is an identity element and the associativity, commutativity and distributivity properties hold, making $S \otimes_R T$ a commutative ring.

EXAMPLE 13.13 As an easy first example, let $T = R$ and $g = \mathrm{id}_R$. We have already seen that $S \otimes_R R$ is isomorphic to S as an S-module. Since $s \otimes 1 \leftrightarrow s$ under the isomorphism, and $(s \otimes 1)(t \otimes 1) = (st) \otimes 1$ (and $1 \otimes 1$ is the identity in $S \otimes_R R$), it follows that $S \otimes_R R$ is in fact isomorphic to S as a ring.

EXAMPLE 13.14 Let $R = \mathbb{R}$, $S = T = \mathbb{C}$, and $f = g$ the usual inclusions $\mathbb{R} \hookrightarrow \mathbb{C}$. We leave it as an exercise to show that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ is isomorphic to the ring $\mathbb{C} \times \mathbb{C}$.

EXAMPLE 13.15 Let T denote the polynomial ring $R[X]$, and let $g : R \to R[X]$ be the usual inclusion (viewing elements of R as constant polynomials). We leave it as another exercise to show that $S \otimes_R R[X]$ is isomorphic to the polynomial ring $S[X]$ over S, and hence that $R[X_1] \otimes_R R[X_2]$ is isomorphic to $R[X_1, X_2]$.

[28] Note that the removal of the parentheses in the first tensor product is justified by Proposition 12.9(ii)

14 Localisation of rings

Starting with a domain R, one can construct a (smallest possible) field containing R, called the *field of fractions* of R. This construction should be familiar, at least in the case of $R = \mathbb{Z}$ (whose field of fractions is $\mathbb{Q}$), and it can be viewed as a special case of a more general construction, called *localisation*, which is useful in various contexts, especially algebraic geometry.

Before discussing the general notion of localisation, we recall the definition of the field of fractions of a domain R. Consider the set of all symbols

$$\left\{ \frac{r}{s} \;\middle|\; r,s \in R,\, s \neq 0_R \right\}$$

where $r \in R$ and $s \in R \setminus \{0_R\}$ (here we consider this all as one "symbol", the line cannot be removed). We define an equivalence relation on this set of symbols by

$$\frac{r}{s} \sim \frac{r'}{s'} \iff rs' = r's \in R.$$

It is obviously reflexive and symmetric. To see that it is transitive, note that if

$$\frac{r}{s} \sim \frac{r'}{s'} \quad \text{and} \quad \frac{r'}{s'} \sim \frac{r''}{s''}$$

(with $r, r', r'' \in R$ and $s, s', s'' \in R \setminus \{0\}$), then $rs' = r's$ and $r's'' = r''s'$. Now this implies that

$$(rs')s'' = (r's)s'' = (r's'')s = (r''s')s, \tag{9}$$

so since R is a domain and $s' \neq 0$, the equation $(rs'')s' = (r''s)s'$ implies that $rs'' = r''s$ (Exercise 7.23).

Let $\mathrm{Frac}(R)$ denote the set of such equivalence classes, i.e.

$$\mathrm{Frac}(R) = \left\{ \frac{r}{s} \;\middle|\; r,s \in R,\, s \neq 0_R \right\} / \sim .$$

We then make $\mathrm{Frac}(R)$ a ring by defining

$$(r/s) + (t/u) = (ru + st)/su \quad \text{and} \quad (r/s) \cdot (t/u) = rt/su$$

for $r, s, t, u \in R$, $s, u \neq 0_R$. Here the identity is $1_R/1_R$ and the zero is $0_R/1_R$. One needs to check that the operations are well-defined and that the ring axioms are satisfied, but we will do it anyway later in more generality in the discussion of localisation. It is easy to see that $\mathrm{Frac}(R)$ is a field since r/s has inverse s/r.

EXERCISE 14.1 Let R be a domain. Show that there is a ring homomorphism

$$\begin{aligned} R &\longrightarrow \mathrm{Frac}(R) \\ r &\longmapsto \frac{r}{1_R}. \end{aligned}$$

Show that this is injective.

Note that the rules for the symbols r/s (in particular the equivalence relation) exactly mimic the rules for fractions. In particular, we have that $\mathrm{Frac}(\mathbb{Z})$ is canonically isomorphic to $\mathbb{Q}$.

For the more general notion of localisation, we need the following definition. As usual, we restrict our attention to commutative rings.

DEFINITION 14.2 Suppose that R is a ring. We say that $S \subset R$ is a *multiplicatively closed subset* of R if (i) $1_R \in S$ and (ii) S is closed under multiplication (i.e. if $s, s' \in S$, then $ss' \in S$).

EXAMPLE 14.3 For any ring R, the subsets $\{1_R\}$ and $\{0_R, 1_R\}$ are multiplicative.

EXAMPLE 14.4 If R is a domain, then the subset $S = R \setminus \{0_R\}$ is multiplicative.

EXAMPLE 14.5 Let $R = \mathbb{Z}$ and let p be a prime number. Then the subset $\{ p^n \mid n \in \mathbb{Z}, n \geq 0 \}$ is multiplicative, as is $\{ m \in \mathbb{Z} \mid p \nmid m \}$.

EXAMPLE 14.6 Suppose that P is a prime ideal of a ring R (see Definition 7.49), and let $S = R \setminus P$. Then S is a multiplicative subset of R: (i) we have $1_R \notin P$ (since $P \neq R$), so $1_R \in S$, (ii) if $s, s' \in S$, then $s, s' \notin P$, so $ss' \notin P$, i.e. $ss' \in S$.

Note that Example 14.4 is a special case of this example (since if R is a domain, then $\{0_R\}$ is a prime ideal), as is the subset $\{ m \in \mathbb{Z} \mid p \nmid m \}$ in Example 14.5 (the prime ideal being $p\mathbb{Z}$.

Suppose now that S is any multiplicative subset of a ring R and consider the set of symbols

$$\left\{ \frac{r}{s} \;\middle|\; r \in R, s \in S \right\}.$$

Define an equivalence relation on this set by

$$\frac{r}{s} \sim \frac{r'}{s'} \iff \text{if } rs't = r'st \text{ for some } t \in S.$$

To see that this is an equivalence relation:

(i) We have $r/s \sim r/s$ since $rs \cdot 1_R = rs \cdot 1_R$ and $1_R \in S$.

(ii) The condition is obviously symmetric.

(iii) Suppose that $r/s \sim r'/s'$ and $r'/s' \sim r''/s''$ (where $r, r', r'' \in R$ and $s, s', s'' \in S$). This means $rs't = r'st$ and $r's''t' = r''s't'$ for some $t, t' \in S$, which implies that

$$rs''(s'tt') = (rs't)s''t' = (r'st)s''t' = (r's''t')st = (r''s't')st = r''s'(s'tt').$$

Since $s', t, t' \in S$ and S is multiplicative, we have $s'tt' \in S$, and therefore $r/s \sim r''/s''$.

Note that if R is a domain, then $rs't = r'st$ if and only if $rs' = r's$ (Exercise 7.23), so the equivalence condition is actually the same as the one in the definition of $\mathrm{Frac}(R)$ (where $S = R \setminus \{0_R\}$). When R is not a domain, we really do need the extra $t \in S$ in order for $\sim$ to be transitive (see (9)). This is new for phenomena for "fractions" and falls outside of any intuition coming from $\mathbb{Z}, \mathbb{Q}$.

We write R_S for the set of these equivalence classes:

$$R_S = \left\{ \frac{r}{s} \;\middle|\; r \in R, s \in S \right\} / \sim .$$

We claim that R_S is actually a ring. We use the same formulae as in the definition of $\mathrm{Frac}(R)$, i.e.

$$(r/s) + (t/u) = (ru + st)/su \quad \text{and} \quad (r/s) \cdot (t/u) = rt/su$$

for $r, t \in R$ and $s, u \in S$. Let us check that addition is well-defined, i.e. that if $r/s = r'/s'$ and $t/u = t'/u'$ (where $r, r', t, t' \in R$ and $s, s', u, u' \in S$), then $(ru + st)/su = (r'u' +$

$s't')/s'u'$. The assumption that $r/s \sim r'/s'$ and $t/u \sim t'/u'$ means that $rs'v = r'sv$ and $tu'w = t'uw$ for some $v, w \in S$. This implies that

$$\begin{aligned}(ru+st)s'u'vw &= (rs'v)uu'w + (tu'w)ss'v \\ &= (r'sv)uu'w + (t'uw)ss'v \quad = \quad (r'u'+s't')suvw,\end{aligned}$$

and it follows that $(ru+st)/(su) \sim (r'u'+s't')/(s'u')$ (since $vw \in S$). Therefore addition is well-defined. Similarly (but more easily), we have that

$$rts'u'vw = (rs'v)(tu'w) = (r'sv)(t'uw) = r't'suvw,$$

so $rt/su = r't'/s'u'$, so multiplication is also well-defined.

Now we check that the ring axioms are satisfied for these operations on R_S. First note that the addition operation is commutative since

$$(r/s) + (t/u) = (ru+st)/su = (ts+ur)/us = (t/u) + (r/s),$$

and it is associative since

$$\begin{aligned}((r/s) + (t/u)) + (v/w) &= ((ru+st)/su) + (v/w) \\ &= ((ru+st)w + suv)/suw = (r(uw) + s(tw+uv))/suw \\ &= (r/s) + (tw+uv)/uw = (r/s) + ((t/u) + (v/w))\end{aligned}$$

for all $r, t, v \in R$, $s, u, w \in S$. Recall that $1_R \in S$ (as part of the definition of a multiplicative subset), so $0_R/1_R \in R_S$, and it is an identity element for addition since

$$(0_R/1_R) + (r/s) = (0_R \cdot s + 1_R \cdot r)/(1_R \cdot s) = r/s$$

for all $r/s \in R_S$. Furthermore the inverse of (r/s) under addition is given by $(-r)/s \in R_S$ since

$$(r/s) + ((-r)/s) = (rs + (-r)s)/s^2 = 0_R/s^2 = 0_R/1_R$$

(where the last equality holds since $0_R \cdot 1_R = 0_R = 0_R \cdot s^2$ (and $s^2 \in S$).

We have now shown that R_S is an abelian group under addition. For the axioms involving multiplication, note that $1_R/1_R$ is an identity element for the operation, and associativity follows easily from the corresponding property for R:

$$((r/s) \cdot (t/u)) \cdot (v/w) = (rt)v/(su)w = r(tv)/s(uw) = (r/s) \cdot ((t/u) \cdot (v/w))$$

for all $r/s, t/u, v/w \in R_S$. Finally the distributive axiom holds since

$$\begin{aligned}(r/s) \cdot ((t/u) + (v/w)) &= (r/s) \cdot (tw+uv)/uw \\ &= r(tw+uv)/suw = (rtw+ruv)/suw = (rtsw+rvsu)/susw \\ &= (rt)/(su) + (rv)/(sw) = ((r/s) \cdot (t/u)) + ((r/s) \cdot (v/w))\end{aligned}$$

for all $r/s, t/u, v/w \in R_S$.

DEFINITION 14.7 If S is a multiplicative subset of a (commutative) ring R, then the ring R_S defined above is called the *localisation of R with respect to S*.

EXAMPLE 14.8 The localisation of R with respect to $S = \{1_R\}$ is isomorphic to R via $r/1_R \leftrightarrow r$. (We have $r/1_R = r'/1_R$ if and only if $(r \cdot 1_R)s = (1_R \cdot r')s$ for some $s \in S = \{1_R\}$, i.e. $s = 1_R$, so $r/1_R = r'/1_R$ if and only $r = r'$. Note also that the resulting bijection between R and R_S is compatible with the addition and multiplication operations.

On the other hand the localisation of R with respect to $\{0_R, 1_R\}$ is the zero ring (i.e. it has only one element). To see this note that $r/1_R = r'/0_R$ for all $r, r' \in R$ since $(r \cdot 0_R)s = (1_R \cdot r')s$ for some $s \in \{0_R, 1_R\}$, namely $s = 0_R$. (More generally, it is usually damaging to localise at a multiplicatively closed set S containing zero divisors, see e.g Example 14.12.)

EXAMPLE 14.9 Suppose that R is a domain and $S = R \setminus \{0_R\}$ (as in Example 14.4). Since the equivalence relation and the operations are the same as in the definition of $\mathrm{Frac}(R)$, we recover the definition of the field of fractions of R as a localisation.

EXAMPLE 14.10 Consider the localisation of $\mathbb{Z}$ with respect to the multiplicative subsets in Example 14.5. We find that $\mathbb{Z}_{\{p^n \mid n \geq 0\}}$ is the subring

$$\{ r/p^n \mid r, n \in \mathbb{Z}, n \geq 0 \}$$

of $\mathbb{Q}$ consisting of those rational numbers whose denominator is a power of p. Similarly if $S = \{ m \in \mathbb{Z} \mid p \nmid m \}$, then $\mathbb{Z}_S$ is the subring of $\mathbb{Q}$ consisting of those rational numbers whose denominator is not divisible by p.

Recall from Example 14.6 that if P is a prime ideal of R, then its complement $S = R \setminus P$ is a multiplicative subset of R. The localisation $R_{R\setminus P}$ is called the *localisation* of R at P and denoted simply (but confusingly[29]!) by R_P. Note P is not itself a multiplicatively closed set and so this is technically not overloading the notation R_S!

For any multiplicative subset of a ring R, we have a ring homomorphism $i : R \to R_S$ defined by $i_S(r) = r/1_R$. (Note that i_S sends 1_R to $1_R/1_R = 1_{R_S}$, and it is compatible with addition and multiplication since

$$(r/1_R) + (r'/1_R) = (r + r')/1_R \quad \text{and} \quad (r/1_R)(r'/1_R) = (rr')/1_R.)$$

EXAMPLE 14.11 If R is a domain and $S = R \setminus \{0\}$, then $i_S : R \to R_S$ is the injective homomorphism identifying R as a subring of its field of fractions.

EXAMPLE 14.12 The homomorphism $i_S : R \to R_S$ is not necessarily injective. For example let $R = \mathbb{Z}/6\mathbb{Z}$ and $S = \{\overline{1}, \overline{3}, \overline{5}\}$ (which is easily seen to be a multiplicative subset). We claim that $\overline{2}) \in \ker(i_S)$. Indeed

$$i_S(\overline{2}) = \overline{2}/\overline{1} = \overline{0}/\overline{1}$$

since $\overline{2} \cdot \overline{1} \cdot s = \overline{0} \cdot \overline{1} \cdot s$ for some $s \in S$, namely $s = \overline{3}$. It follows that the entire ideal $(\overline{2})$ is contained in $\ker(i_S)$; on the other hand $\overline{1}/\overline{1} \neq \overline{0}/\overline{1}$ (since $\overline{1} \cdot \overline{1} \cdot s \neq \overline{0} \cdot \overline{1} \cdot s$ for all $s \in S$), so $\ker(i) = \{\overline{2}, \overline{4}, \overline{6}\}$. Furthermore in this case i_S is surjective: if $r \in R$ and $s \in S$, then $r/s = r/\overline{1} = i_S(r)$ since $r(s-\overline{1})\overline{3} = \overline{0}$. Therefore $R_S = \{\overline{0}/\overline{1}, \overline{1}/\overline{1}\}$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

We now note a general property of the homomorphism $i_S : R \to R_S$: if $s \in S$, then $i_S(s) \in R_S^\times$, i.e. $s/1_R$ has a multiplicative inverse in R_S, namely $1_R/s$. There is a sense in which R_S is the "smallest" ring admitting a homomorphism from R with this property. This is formulated in the following proposition, whose proof we leave as an exercise:

PROPOSITION 14.13 *Let S be a multiplicative subset of a ring R, and $i_S : R \to R_S$ the ring homomorphism defined by $i_S(r) = r/1_R$. Then these satisfy the following universal property: any ring homomorphism $f : R \to T$ such that $f(S) \subseteq T^\times$ extends uniquely to a ring homomorphism $\widetilde{f}$ such that the following diagram commutes:*

$$\begin{array}{ccc} R & \xrightarrow{f} & T \\ {\scriptstyle i_S}\downarrow & \nearrow_{\exists!\ \widetilde{f}} & \\ R_S & & \end{array} .$$

[29] From an algebra perspective it would be sensible to think of localising at the multiplicatively closed set $R \setminus P$ as the localisation "away from" P. The word local and the reason to think of this as happening "at" P is coming from geometry and unfortunately we are unable to explain the correct intuition here.

(Recall that i_S need not be injective, so to use the word "extends" in this context might be new!)

Considering the case of a domain R and $S = R \setminus \{0\}$ gives the following:

COROLLARY 14.14 *Suppose that R is a domain, T is a field, and $f : R \to T$ is an injective ring homomorphism. Then there is a unique field homomorphism $g : \mathrm{Frac}(R) \to T$ whose restriction to R is f.*

Note that the injectivity is needed in order to ensure the hypothesis $f(S) \subset T^\times$ is satisfied. In this case $S = R \setminus \{0_R\}$ and $T^\times = T \setminus \{0_T\}$, so we are requiring that $f(r) \neq 0_T$ if $r \neq 0_R$.

15 Localisation of modules and ideals

There is also a notion of localisation for modules. We continue to assume that S is a multiplication subset of R, and now let M be an R-module. Consider the set of symbols

$$\left\{ \frac{m}{s} \;\middle|\; m \in M, s \in S \right\},$$

i.e. "elements of M with formal denominators in S". We can then define an equivalence relation by

$$m/s \sim m'/s' \quad \text{if } ts'm = tsm' \text{ for some } t \in S.$$

The proof that this is an equivalence relation is the same as for the one for R_S (which can be viewed as the particular case where $M = R$).

We then write m/s for the equivalence class of (m,s), and define the *localisation of M with respect to S* to be the set

$$M_S = \{\, m/s \mid m \in M, s \in S \,\}$$

of equivalence classes. We then define an addition operation on M_S by the formula

$$(m/s) + (m'/s') = (s'm + sm')/ss',$$

and a scalar multiplication by $(r/s)(m/s') = (rm)/(ss')$ for $r/s \in R_S$. The proof that these operations are well-defined and make M_S an R_S-module are very similar to the proof that R_S is a ring, and we leave it as an exercise.

We may also similarly define $j_S : M \to M_S$ by $j(m) = m/1_R$, and verify that it is a homomorphism of R-modules, where R_S-module M_S is viewed as an R-module via the homomorphism $i_S : R \to R_S$, so $r \cdot (m/s) = (r/1_R)(m/s) = (rm)/s$.

REMARK 15.1 We briefly mention that the localisation M_S can also be interpreted as a tensor product. Recall that if $R \to T$ is a ring homomorphism, M is an R-module and N is a T-module, then there is a bijection[30]

$$\mathrm{Hom}_T(T \otimes_R M, N) \longleftrightarrow \mathrm{Hom}_R(M, N).$$

So letting $T = R_S$ and $N = M_S$, there is a unique R_S-linear homomorphism $R_S \otimes_R M \to M_S$ corresponding to the R-linear homomorphism $j_S : M \to M_S$ (sending $(r/s) \otimes m$ to $(rm)/s$ for $r \in R$, $s \in S$ and $m \in M$). We leave it as an exercise to prove that this in fact gives an isomorphism

$$R_S \otimes_R M \xrightarrow{\sim} M_S.$$

[30] The bijection is defined by sending $\varphi : T \otimes_R M \to N$ to its composite with the R-linear homomorphism $M \to T \otimes_R M$ defined by $m \mapsto 1_T \otimes m$; the proof was left as an exercise.

Recall that if $\varphi : M \to N$ is a homomorphism of R-modules, and $f : R \to T$ is a ring homomorphism, then we obtain a homomorphism $\mathrm{id}_T \otimes \varphi : T \otimes_R M \to T \otimes_R N$ of T-modules. Applying this with $T = R_S$ and $f = i : R \to R_S$ gives a homomorphism of R_S-modules

$$R_S \otimes_R M \longrightarrow R_S \otimes_R N,$$

so using the isomorphisms $R_S \otimes_R M \xrightarrow{\sim} M_S$ and $R_S \otimes_R N \xrightarrow{\sim} N_S$, we get a homomorphism $M_S \to N_S$, which we denote φ_S. It is straightforward to check that this homomorphism is described by the formula $\varphi_S(m/s) = \varphi(m)/s$. for $m \in M$, $s \in S$. (One can also check more directly from the definition of the localisations that this map is well-defined and R_S-linear.) Since $(\mathrm{id}_M)_S = \mathrm{id}_{M_S}$ for any R-module M, and $(\psi \circ \varphi)_S = \psi_S \circ \varphi_S$ for any homomorphisms $\varphi : M \to N$ and $\psi : N \to P$ of R-modules, we get a functor from the category of R-modules to the category of R_S-modules, defined by $F(M) = M_S$ and $F(\varphi) = \varphi_S$. We leave it as an exercise to show that if φ is injective, then so is φ_S, and if φ is surjective, then so is φ_S.

Given an ideal I of R, we have two ways we can define a corresponding ideal of R_S:

(i) Since we have a map $i_S : R \to R_S$, we can consider the ideal $(i_S(I))$ generated by the image of I.

(ii) The ideal I is itself an R-module. So we can consider its localisation

$$I_S = \{\, a/s \mid a \in I, s \in S \,\}.$$

Since the inclusion $I \hookrightarrow R$ is an injective homomorphism of R-modules, we also obtain an injective homomorphism $I_S \to R_S$ of R_S-modules. (This is a special case of the above remark, but really if you look at what it means to be equivalent for I_S and R_S, they're really the same[31].) Since the image of an R_S-module homomorphism is an R_S-submodule, and R_S-submodules of R_S are ideals, we find that the image (also denoted I_S) is an ideal.

We claim these constructions are equal, i.e. $(i_S(I)) = I_S$. First note that given $x/s \in I_S$ for $x \in I$ and $s \in S$, then within R_S we have that

$$\frac{x}{s} = \frac{1_R}{s} \cdot \frac{x}{1_R} = \frac{1_R}{s} \cdot i_S(x). \tag{10}$$

So $I_S \subseteq (i_S(I))$ (the ideal of R_S generated by $i_S(I)$). But conversely $i_S(x) = x/1_R$ is certainly in I_S, so we must have equality.

On the other hand, the preimage of any ideal of R_S under $i_S : R \to R_S$ is an ideal of R (Exercise 7.45).

EXAMPLE 15.2 As an example, consider the ideal $J := (p/1) \trianglelefteq \mathbb{Z}_{(p)}$ (so $S = \mathbb{Z} \setminus (p)$). We claim that $i_S^{-1}(J) = (p)$. One inclusion is easy as certainly $i_S((p)) \subseteq (p/1)$. We want to show the reverse inclusion. General elements of J are of the form $(r/s)\cdot(p/1) = (rp)/s$ for some $r \in \mathbb{Z}$ and $s \in \mathbb{Z} \setminus (p)$. We must show that whenever $a \in \mathbb{Z}$ is such that $i_S(a) = a/1$ is equivalent to such an element, then $a \in (p)$. But

$$\begin{aligned} \frac{a}{1} \sim \frac{rp}{s} &\iff \exists\ t \in \mathbb{Z} \setminus (p) \text{ s.t. } rpt = ast \\ &\implies ast \in (p), \end{aligned}$$

[31] Note however that given an inclusion $\varphi : M \hookrightarrow N$, elements might become equivalent to things not in the image of φ. This is familiar though, for $\varphi : \mathbb{Z} \hookrightarrow \mathbb{Q}$ we know that $2/1 \sim 6/3$.

so as $st \in S = \mathbb{Z} \setminus (p)$ and (p) is a prime ideal, we find that $a \in (p)$ as desired[32].

We therefore have maps in each direction between the sets of ideals of the rings R and R_S:

$$\{\text{ideals of } R\} \underset{i_S^{-1}(J) \hookleftarrow J}{\overset{I \mapsto I_S}{\rightleftarrows}} \{\text{ideals of } R_S\}. \tag{11}$$

We claim that one of the two composites is the identity: every ideal of R_S is the localisation of its preimage.

LEMMA 15.3 *If J is an ideal of R_S, then $(i_S^{-1}(J))_S = J$.*

Proof. Note first that since $i_S(i_S^{-1}(J)) \subset J$, we certainly have that the corresponding containment $(i_S(i_S^{-1}(J))) \subset (J)$ of the ideals they generate within R_S. But the former is $(i_S^{-1}(J))_S$ (see (10)) and since J is already an ideal $J = (J)$.

To prove the other containment, suppose that $a/s \in J$ (where $a \in R$ and $s \in S$). Since J is an ideal of R_S and $s/1_R \in R_S$, we have

$$i_S(a) = a/1_R = (s/1_R)(a/s) \in J.$$

Therefore $a \in i_S^{-1}(J)$, and $a/s \in (i_S^{-1}(J))_S$, as required. □

It follows that the map in the right-hand direction in (11) is surjective (i.e. there are no "new" ideals of localisations they all come from R), and the map in the left-hand direction is injective:

COROLLARY 15.4 *Every ideal of R_S is of the form I_S, where I is an ideal of R. Furthermore if J and J' are ideals of R_S such that $i_S^{-1}(J) = i_S^{-1}(J')$, then $J = J'$.*

It is easy to see that the maps in (11) are not necessarily bijective (Example 15.2 is unusually well-behaved). Suppose for example that I is any ideal of R such that the intersection $I \cap S$ is non-empty. If $s \in I \cap S$, then

$$1_{R_S} = 1_R/1_R = (1_R/s)(s/1_R) \in I_S,$$

so $I_S = R_S$. But this can regularly happen for ideals that aren't R, for example $(2), (3) \trianglelefteq \mathbb{Z}$ both have image in $\mathrm{Frac}(\mathbb{Z}) = \mathbb{Q}$ the whole ring (indeed we know the map $\{\text{ideals of } \mathbb{Z}\} \to \{\text{ideals of } \mathbb{Q}\}$ must be highly non-injective. (We leave it as an exercise to show the converse is true: if I is an ideal of R such that $I_S = R_S$, then $I \cap S \neq \emptyset$.) So if S contains some $a \notin R^\times$, then letting I be the principal ideal (a) of R, we have $I \neq R$, but $I_S = R_S$. (As an extreme example, consider what happens if $S = \{0_R, 1_R\}$, as in Example 14.8.)

Rings with less ideals are considered "easier" (with fields being the easiest!). The set up of (11) shows that localisation reduces the number of ideals and therefore actually makes rings simpler (but usually larger)!

On the other hand, we can actually show that certain ideals do survive localisation and give a subset of $\{I \trianglelefteq R\}$ on which $I \mapsto I_S$ becomes bijective.

LEMMA 15.5 *Suppose P is a prime ideal of R such that $P \cap S = \emptyset$. Then*

(i) P_S is a prime ideal of R_S;

[32] Note also that if R is a domain and $\{0\} \notin S$, then the map $i_S : R \to R_S$ is injective so that i_S^{-1} corresponds to "intersection" with R.

(ii) $i_S^{-1}(P_S) = P$.

Proof. To prove (i), we must show that if $rt/su = (r/s)(t/u) \in P_S$, then $r/s \in P_S$ or $t/u \in P_S$ (where $r,t \in R$ and $s,u \in S$). If $rt/su \in P_S$, then $rt/su = v/w$ for some $v \in P$, $w \in S$, i.e.

$$rtwx = suvx \quad \text{for some } x \in S.$$

Since $v \in P$, it follows that $rtwx = suvx \in P$. On the other hand, since $w \in S$ and $x \in S$ we have $wx \in S$ (since S is multiplicative) and therefore $wx \notin P$ (since we are assuming that $P \cap S = \emptyset$). Since P is a prime ideal of R, it follows that $rt \in P$, and so at least one of r or t is an element of P. We therefore conclude that $r/s \in P_S$ or $t/u \in P_S$, as required.

To prove (ii), note first that since $i(P) \subset P_S$, we have $P \subset i^{-1}(P_S)$.

For the other containment, suppose $r \in i^{-1}(P_S)$, i.e. $r \in R$ is such that $r/1_R \in P_S$. This means that $r/1_R = s/t$ for some $s \in P$, $t \in S$, i.e.

$$rtu = su \quad \text{for some } u \in S.$$

Since $s \in P$, we have $rtu = su \in P$, but since $tu \in S$ and $P \cap S = \emptyset$, we have $tu \notin P$, and therefore $r \in P$, as required. □

Note that we made crucial use of the hypothesis that $P \cap S = \emptyset$ in proving both parts of the lemma. If this hypothesis does not hold, then $P_S = R_S$ is certainly not a prime ideal of R_S, nor is $i^{-1}(P_S) = R$ equal to P.

Recall that the preimage of a prime ideal under a ring homomorphism is a prime ideal (Exercise 7.56). In particular if Q is a prime ideal of R_S, then $i_S^{-1}(Q)$ is a prime ideal of R. Furthermore $i_S^{-1}(Q) \cap S = \emptyset$, for otherwise Lemma 15.3 would imply that $Q = (i_S^{-1}(Q))_S = R_S$, contradicting that Q is a prime ideal of R_S. Combining this with Lemma 15.5(i), we see that the maps in (11) restrict to maps

$$(12) \qquad \left\{ \begin{array}{l} \text{prime ideals } P \text{ of } R \\ \text{such that } P \cap S = \emptyset \end{array} \right\} \underset{i_S^{-1}(Q) \hookleftarrow Q}{\overset{P \mapsto P_S}{\rightleftarrows}} \{ \text{prime ideals } Q \text{ of } R_S \}.$$

Furthermore Lemmas 15.3 and 15.5(ii) say that these two maps are inverses of each other. Note also that they are inclusion-preserving (i.e. if $P \subset P'$, then $P_S \subset P'_S$, and if $Q \subset Q'$ then $i_S^{-1}(Q) \subset i_S^{-1}(Q')$). We have therefore proved the following:

COROLLARY 15.6 *The maps in (12) define an inclusion-preserving bijection between the set of prime ideals of R_S and the set of prime ideals of R disjoint from S.*

EXAMPLE 15.7 Consider the localisation of $R = \mathbb{Z}$ with respect to $S = \{ p^n \mid n \geq 1 \}$, where p is a prime number (see Example 14.10). Recall from Example 7.51 that the prime ideals of $\mathbb{Z}$ are precisely those of the form $q\mathbb{Z}$, where q is a prime number, and $\{0\}$. The only one of these that contains a power of p is $p\mathbb{Z}$. Therefore the prime ideals of

$$\mathbb{Z}_S = \{ r/p^n \mid r, n \in \mathbb{Z}, n \geq 0 \}$$

are $\{0\}_S = \{0\}$ and $(q\mathbb{Z})_S = q\mathbb{Z}_S$ for prime numbers $q \neq p$. (Note that $p\mathbb{Z}_S$ is the whole ring $\mathbb{Z}_S$.)

EXAMPLE 15.8 Now consider the localisation of $R = \mathbb{Z}$ with respect to $S = \mathbb{Z} \setminus p\mathbb{Z}$ (where p is a prime number), so

$$\mathbb{Z}_S = \{ m/n \mid m, n \in \mathbb{Z}, p \nmid n \}$$

(again as described in Example 14.10). Note that S contains every prime number other than p, so the only prime ideals of $\mathbb{Z}$ disjoint from S are $\{0\}$ and $p\mathbb{Z}$. Therefore the only prime ideals of $\mathbb{Z}_S$ are $\{0\}$ and $p\mathbb{Z}_S$.

EXERCISE 15.9 Show that the preimage of the zero ideal $i_S^{-1}((0))$ is strictly larger than the zero ideal (0) if and only if S contains zero divisors[33]. Use this to see that Corollary 15.6 is false without the prime hypothesis.

EXERCISE* 15.10 Extend Exercise 15.9 to completely classify when an ideal is the restriction of an ideal from R_S. (*Hint: Reduce to the above case by considering R/I. This will require thinking about how localisation behaves under quotients.*)

Suppose now that Q is a prime ideal of R, so that $S = R \setminus Q$ is a multiplicative subset of R (as in Example 15.8). Note that if P is any prime ideal of R, then $P \cap S = \emptyset$ if and only if $P \subset Q$. So in this case Corollary 15.6 gives an inclusion-preserving bijection

$$\{\text{prime ideals of } R \text{ contained in Q}\} \longleftrightarrow \{\text{prime ideals of } R_S\}. \tag{13}$$

In particular every prime ideal of R_S is contained in Q_S. Since every maximal ideal of a ring is prime (see Definition 7.50), it follows that Q_S is the unique maximal ideal of R_S.

The bijection of (13) should be compared with the bijection

$$\{\text{ideals of } R \text{ containing } Q\} \leftrightarrow \{\text{ideals of } R/Q\}$$

of Exercise 7.48.

DEFINITION 15.11 A ring is called a *local ring* if it has a unique maximal ideal.

Thus if S is the complement of a prime ideal in a ring R, then R_S is a local ring. So for example the ring $\mathbb{Z}_S$ in Example 15.8 is a local ring. Note also that every field is a local ring (its unique maximal ideal being $\{0\}$).

Before proving an interesting property of local rings, we need the following very general fact about rings and ideals:

LEMMA 15.12 *If I is a proper[34] ideal of a ring R, then there is a maximal ideal P of R such that $I \subset P$.*

Proof. We will appeal to Zorn's Lemma (9.19).

Let $\mathcal{S}$ denote the set of proper ideals $J \subsetneq R$ such that $I \subset J$. We view $\mathcal{S}$ as being partially ordered by inclusion.

We will show that $\mathcal{S}$ satisfies the hypotheses of Zorn's Lemma. Suppose then that $\mathcal{T} \subset \mathcal{S}$ is a non-empty[35] chain. We claim that $A = \bigcup_{J \in \mathcal{T}} J$ is an upper bound for $\mathcal{T}$.

We must prove that A is an ideal of R. Note first that $0 \in J \subset A$ (for any $J \in \mathcal{T}$), and that if $x \in A$, then $x \in J$ for some $J \in \mathcal{T}$, and hence $-x \in J \subset A$. Furthermore if $x, x' \in A$ then $x \in J$ and $x' \in J'$ for some $J, J' \in \mathcal{T}$. Since $\mathcal{T}$ is a chain, we have $J \subset J'$ or $J' \subset J$, and so $x \in J'$ or $x' \in J$, and hence $x + x' \in J \cup J' \subset A$. Finally note that if $r \in R$ and $x \in A$, then $x \in J$ for some $J \in \mathcal{T}$, so also $rx \in J \subset A$. Therefore A is an ideal of R.

[33] Recall $s \in R$ is a zero divisor if there exists a non-zero $r \in R$ such that $rs = 0$.

[34] The meaning here of *proper* is that $I \neq R$

[35] We can assume that $\mathcal{T}$ is non-empty since any element of $\mathcal{S}$ is an upper bound for the empty set (and $\mathcal{S}$ is non-empty since $I \in \mathcal{S}$).

Furthermore we have $I \subset J \subset A$ for any $J \in \mathcal{T}$, so $A \in \mathcal{S}$ and A is an upper bound for $\mathcal{T}$.

We can now invoke Zorn's Lemma to conclude that $\mathcal{S}$ has a maximal element P. Note that $I \subset P$, and P is a maximal ideal of R since if $P \subset J$, then either $J = R$ or $J \in \mathcal{S}$, in which case the maximality of P (as an element of $\mathcal{S}$) implies that $J = P$. □

The above Zorn's lemma argument is required to know to switch between Definition 15.11 and the following subtly but importantly stronger equivalent definition. The latter fits much closer to the intuition of a local ring!

LEMMA 15.13 *A ring R is local if and only if there exists a maximal ideal P such that every proper ideal I is contained in Q.*

Proof. Suppose that we have a proper ideal I. By Lemma 15.12, there exists a maximal ideal containing I. But since R is local, it has only one maximal ideal and $Q = P$. So P does contain I. The backwards direction is obvious. □

We can now deduce the following:

PROPOSITION 15.14 *Suppose that R is a local ring and P is its maximal ideal. Then $R^\times = R \setminus P$.*

Proof. Note that $r \in R^\times$ if and only if $(r) = R$ ("r generates the unit ideal"). If $r \in P$, then $(r) \subseteq P$. So this immediately shows that if $r \in R^\times$, then r cannot be in P.

Now let $r \notin P$. Then (r) is an ideal and so by Lemma 15.13 either contained in P or is not proper. The first is a contradiction, so $(r) = R$ and $r \in R^\times$. □

EXERCISE 15.15 Prove the converse, that is if I is a proper ideal of R such that $R \setminus \{I\} = R^\times$, then I is maximal and R is a local ring.

EXAMPLE 15.16 Consider the ring

$$\mathbb{Z}_S = \{ m/n \mid m, n \in \mathbb{Z}, p \nmid n \}$$

of Example 15.8. It is a local ring with maximal ideal

$$p\mathbb{Z}_S = \{ m/n \mid m, n \in \mathbb{Z}, p \mid m, p \nmid n \}.$$

The complement of $p\mathbb{Z}_S$ in $\mathbb{Z}_S$ is

$$\mathbb{Z}_S^\times = \{ m/n \mid m, n \in \mathbb{Z}, p \nmid m, p \nmid n \}.$$

Part 3

Homological algebra

Homological algebra involves the study of certain properties of sequences of algebraic objects. It was initially developed for applications to topology, but the ideas and methods have turned out to be useful much more broadly.

16 Exact sequences and chain complexes

We start with the following basic definition:

DEFINITION 16.1 A sequence

$$M \xrightarrow{f} N \xrightarrow{g} P$$

of homomorphisms of R-modules is called *exact* if $\mathrm{im}(f) = \ker(g)$.

EXAMPLE 16.2 Consider the sequence of homomorphisms of $\mathbb{Z}$-modules (i.e. abelian groups):

$$\mathbb{Z}/6\mathbb{Z} \xrightarrow{f} \mathbb{Z}/4\mathbb{Z} \xrightarrow{g} \mathbb{Z}/2\mathbb{Z},$$

where f is defined[36] by $f(\overline{a}) = 2\overline{a}$ and g by $g(\overline{b}) = \overline{b}$ (using $\bar{\cdot}$ to denote the residue class mod n for the appropriate n). The sequence is exact since $\mathrm{im}(f) = \{\overline{0}, \overline{2}\} = \ker(g)$.

The notion also applies to (longer) sequences of homomorphisms:

DEFINITION 16.3 Suppose that $n \geq 2$. We say that a sequence of homomorphisms of R-modules

$$M_n \xrightarrow{f_n} M_{n-1} \xrightarrow{f_{n-1}} \cdots \xrightarrow{f_2} M_1 \xrightarrow{f_1} M_0$$

is *exact* if $\ker(f_i) = \mathrm{im}(f_{i+1})$ for $i = 1, \ldots, n-1$. (Note that $\ker(f_i)$ and $\mathrm{im}(f_{i+1})$ are both submodules of M_i.)

EXAMPLE 16.4 Let m be a positive integer and consider the sequence of homomorphisms of $\mathbb{Z}$-modules[37]:

$$0 \xrightarrow{f_4} \mathbb{Z} \xrightarrow{f_3} \mathbb{Z} \xrightarrow{f_2} \mathbb{Z}/m\mathbb{Z} \xrightarrow{f_1} 0,$$

where f_3 is defined by $f_3(a) = ma$, f_2 by $f_2(b) = \overline{b}$ (and of course $f_4(0) = 0$ and $f_1(\overline{c}) = 0$ for all $\overline{c} \in \mathbb{Z}/m\mathbb{Z}$).

The sequence is exact since

- $\ker(f_1) = \mathbb{Z}/m\mathbb{Z} = \mathrm{im}(f_2)$;
- $\ker(f_2) = m\mathbb{Z} = \mathrm{im}(f_3)$;
- $\ker(f_3) = 0 = \mathrm{im}(f_4)$.

Note that for a sequence of the form

$$0 \xrightarrow{e} M \xrightarrow{f} N \xrightarrow{g} P \xrightarrow{h} 0, \tag{14}$$

we necessarily have $\mathrm{im}(e) = 0$ and $\ker(h) = P$, so it is exact if and only if the following hold:

- $\ker(f) = 0$, i.e. f is injective;
- $\ker(g) = \mathrm{im}(f)$;
- $\mathrm{im}(g) = P$, i.e. g is surjective.

[36] Note that f is well-defined: by the first isomorphism theorem, we need only check that $6\mathbb{Z}$ maps to zero in $\mathbb{Z}/4\mathbb{Z}$ under multiplication by 2 (which it does).

[37] Here 0 denotes the $\mathbb{Z}$-zero module, that is the $\mathbb{Z}$-module with underlying set $\{0\}$ consisting just of one element.

EXERCISE 16.5 What does it mean for a sequence $0 \to M \to N \to 0$ with just two non-zero terms to be exact?

Before continuing, we make a brief remark about notation: For any R-modules M, N, we will write 0 for the zero homomorphism sending all elements to zero. In particular, e, h above are special cases of the homomorphism. Furthermore we will usually not bother with notation for homomorphisms which are *forced* to be 0 because their domain or target is, so (14) would be written simply as $0 \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P \longrightarrow 0$.

We can also consider infinite sequences of R-modules and homomorphisms, i.e. an R-module M_i for each integer i, along with an R-module homomorphism $f_i : M_i \to M_{i-1}$ for each $i \in \mathbb{Z}$.

DEFINITION 16.6 A sequence of R-module homomorphisms

$$\cdots \xrightarrow{f_{i+2}} M_{i+1} \xrightarrow{f_{i+1}} M_i \xrightarrow{f_i} M_{i-1} \xrightarrow{f_{i-1}} M_{i-2} \xrightarrow{f_{i-2}} \cdots$$

is *exact* if $\ker(f_i) = \operatorname{im}(f_{i+1})$ for all $i \in \mathbb{Z}$.

An exact sequence of the form (14) is often called a *short exact sequence*, and an exact sequence as in Definition 16.6 is often called a *long exact sequence*.

EXAMPLE 16.7 Consider the sequence of $\mathbb{R}$-modules (i.e. $\mathbb{R}$-vector spaces)

$$\cdots \xrightarrow{f_{i+2}} \mathbb{R}^2 \xrightarrow{f_{i+1}} \mathbb{R}^2 \xrightarrow{f_i} \mathbb{R}^2 \xrightarrow{f_{i-1}} \mathbb{R}^2 \xrightarrow{f_{i-2}} \cdots,$$

where for each $i \in \mathbb{Z}$, $M_i = \mathbb{R}^2$ and $f_i : \mathbb{R}^2 \to \mathbb{R}^2$ is the $\mathbb{R}$-linear homomorphism defined by left-multiplication (on column vectors) by the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, i.e. $f_i(x_1\mathbf{e}_1 + x_2\mathbf{e}_2) = x_2\mathbf{e}_1$ for $x_1, x_2 \in \mathbb{R}$, where $\mathbf{e}_1 = {}^t(1,0)$ and $\mathbf{e}_2 = {}^t(0,1)$ are the standard basis vectors. Then $\operatorname{im}(f_{i+1}) = \mathbb{R}\mathbf{e}_1 = \ker(f_i)$ for all $i \in \mathbb{Z}$, so the sequence is exact.

EXAMPLE 16.8 Consider the sequence of $\mathbb{Z}$-modules

$$\cdots \xrightarrow{\mathrm{id}_{\mathbb{Z}}} \mathbb{Z} \xrightarrow{0} \mathbb{Z} \xrightarrow{\mathrm{id}_{\mathbb{Z}}} \mathbb{Z} \xrightarrow{0} \mathbb{Z} \xrightarrow{\mathrm{id}_{\mathbb{Z}}} \cdots$$

where $M_i = \mathbb{Z}$ for each $i \in \mathbb{Z}$, and $f_i : \mathbb{Z} \to \mathbb{Z}$ is either zero or the identity, according to whether i is even or odd. If i is even, then $f_i = 0$, so $\ker(f_i) = \mathbb{Z}$, and $i+1$ is odd, so $f_{i+1} = \mathrm{id}_{\mathbb{Z}}$ and $\operatorname{im}(f_{i+1}) = \mathbb{Z}$. On the other hand if i is odd, then $f_i = \mathrm{id}_{\mathbb{Z}}$, so $\ker(f_i) = 0$, and $i+1$ is even, so $f_{i+1} = 0$ and $\operatorname{im}(f_{i+1}) = 0$. Therefore in either case we have $\operatorname{im}(f_{i+1}) = \ker(f_i)$, so the sequence is exact.

Note that if a sequence

$$M_n \xrightarrow{f_n} M_{n-1} \xrightarrow{f_{n-1}} \cdots \xrightarrow{f_2} M_1 \xrightarrow{f_1} M_0$$

is exact, then we have $f_i \circ f_{i+1} = 0$ for $i = 1, \ldots, n-1$. Indeed for all $m \in M_{i+1}$, we have $f_{i+1}(m) \in \operatorname{im}(f_{i+1}) = \ker(f_i)$, so $f_i(f_{i+1}(m)) = 0$. Note that we only used the containment $\operatorname{im}(f_{i+1}) \subset \ker(f_i)$, and in fact the converse holds as well[38]: if $f_i \circ f_{i+1} = 0$, then $f_{i+1}(m) \in \ker(f_i)$ for all $m \in M_{i+1}$, so $\operatorname{im}(f_{i+1}) \subset \ker(f_i)$.

We will also be interested in sequences of homomorphisms that satisfy this weaker property for all $i \in \mathbb{Z}$:

[38] In particular, (14) is exact if and only if it is a chain complex, f is injective, g is surjective and $\ker(g) \subseteq \operatorname{im}(f)$

DEFINITION 16.9 A sequence of R-module homomorphisms

$$\cdots \xrightarrow{d_{i+2}} M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \xrightarrow{d_{i-1}} M_{i-2} \xrightarrow{d_{i-2}} \cdots$$

is called a *chain complex* if $d_i \circ d_{i+1} = 0$ (or equivalently $\mathrm{im}(d_{i+1}) \subset \ker(d_i)$) for all $i \in \mathbb{Z}$.

EXAMPLE 16.10 Letting M_0 be any R-module and $M_i = 0$ for all integers $i \neq 0$ gives a chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow M_0 \longrightarrow 0 \longrightarrow \cdots .$$

EXAMPLE 16.11 More generally for any R-modules M_i, we have a chain complex

$$\cdots \xrightarrow{0} M_{i+1} \xrightarrow{0} M_i \xrightarrow{0} M_{i-1} \xrightarrow{0} M_{i-2} \xrightarrow{0} \cdots$$

with $d_i = 0$ for all $i \in \mathbb{Z}$.

EXAMPLE 16.12 Every long exact sequence (as in Definition 16.6) is a chain complex.

EXAMPLE 16.13 Consider the sequence of $\mathbb{Z}$-modules

$$\cdots \xrightarrow{d_{i+2}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i+1}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_i} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i-1}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i-2}} \cdots$$

where $d_i : \mathbb{Z}/8\mathbb{Z} \to \mathbb{Z}/8\mathbb{Z}$ is defined by $d_i(\overline{a}) = 4\overline{a}$ for all $i \in \mathbb{Z}$. This is a chain complex since $d_i(d_{i+1}(\overline{a})) = 16\overline{a} = \overline{0}$ for all $i \in \mathbb{Z}$.

EXAMPLE 16.14 Consider the sequence of $\mathbb{Z}$-modules

$$\cdots \xrightarrow{d_3} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_2} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_1} \mathbb{Z}/4\mathbb{Z} \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \cdots$$

with $M_i = \mathbb{Z}/4\mathbb{Z}$ if $i \geq 0$, $M_i = 0$ if $i < 0$ and $d_i : \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z}$ defined by $d_i(\overline{a}) = 2\overline{a}$ if $i > 0$ (and $d_i = 0$ if $i \leq 0$). This is a chain complex since $d_i(d_{i+1}(\overline{a})) = d_i(2\overline{a}) = 4\overline{a} = \overline{0}$ if $i > 0$ (and clearly $d_i \circ d_{i+1} = 0$ for $i \leq 0$).

Finally we introduce some more notation and terminology. We will sometimes simply write $(M_\bullet, d_\bullet)$ for the chain complex

$$\cdots \xrightarrow{d_{i+2}} M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \xrightarrow{d_{i-1}} M_{i-2} \xrightarrow{d_{i-2}} \cdots$$

(thinking of the subscript $\bullet$ as running through the integers).

We say that a chain complex $(M_\bullet, d_\bullet)$ is *bounded below* if there is an $n \in \mathbb{Z}$ such that $M_i = 0$ for all $i < n$. So the chain complex in Example 16.14 is bounded below (taking $n = 0$). There is similar notion of a chain complex being *bounded above* $M_i = 0$ for all i greater than some n. Furthermore we say a chain complex is *finite* if it is both[39] bounded above and bounded below, so for example the chain complex in Example 16.10 is finite.

If a chain complex is bounded below, with $M_i = 0$ for $i < n$, we will often not bother to write the terms M_i for $i < n$, so the chain complex in Example 16.14 might just be denoted

$$\cdots \xrightarrow{d_3} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_2} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_1} \mathbb{Z}/4\mathbb{Z}$$

(and similarly for complexes that are bounded above or finite).

[39] These are often referred to as *bounded*, but we'll use *finite* to avoid the potential ambiguity created by the possibility of interpreting *bounded* as meaning bounded above *or* below.

17 Homology

Suppose that $(M_\bullet, d_\bullet)$, i.e.

$$\cdots \xrightarrow{d_{i+2}} M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \xrightarrow{d_{i-1}} M_{i-2} \xrightarrow{d_{i-2}} \cdots,$$

is a chain complex of R-modules. Recall this means that $\operatorname{im}(d_{i+1}) \subset \ker(d_i)$ for all i, and it is *exact* if $\operatorname{im}(d_{i+1}) = \ker(d_i)$. We will now define the *homology* of the complex, which can be viewed as measuring the failure of the chain complex to be exact.

DEFINITION 17.1 Suppose that $(M_\bullet, d_\bullet)$ is a chain complex of R-modules. We let $Z_i = \ker(d_i)$ and $B_i = \operatorname{im}(d_{i+1})$. (Note that B_i is an R-submodule of Z_i because $(M_\bullet, d_\bullet)$ is a chain complex.) We define the *homology in degree i* (or *i^{th} homology* of $(M_\bullet, d_\bullet)$ to be the R-module Z_i/B_i. We call elements of M_i *chains (in degree i)*, those of Z_i *cycles*, B_i *boundaries*, and H_i *homology classes*.

EXAMPLE 17.2 Consider the chain complex

$$\cdots \xrightarrow{d_{i+2}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i+1}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_i} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i-1}} \mathbb{Z}/8\mathbb{Z} \xrightarrow{d_{i-2}} \cdots$$

of $\mathbb{Z}$-modules in Example 16.13 (with $d_i(\overline{a}) = 4\overline{a}$ for all i). We then have

- $Z_i = \ker(d_i) = 2\mathbb{Z}/8\mathbb{Z}$,
- $B_i = \operatorname{im}(d_{i+1}) = 4\mathbb{Z}/8\mathbb{Z}$,
- $H_i = Z_i/B_i \cong 2\mathbb{Z}/4\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$.

EXAMPLE 17.3 If $(M_\bullet, d_\bullet)$ is exact, then $Z_i = \ker(d_i) = \operatorname{im}(d_{i+1}) = B_i$ for all i, so $H_i = Z_i/B_i = 0$. Note that for any chain complex $(M_\bullet, d_\bullet)$, we have that $H_i = 0$ if and only if $\ker(d_i) = \operatorname{im}(d_{i+1})$ if and only

$$M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1}$$

is exact. More generally H_i can be viewed as quantifying the difference (as an R-module) between $\ker(d_i)$ and $\operatorname{im}(d_{i+1})$, i.e. the extent to which $M_{i+1} \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1}$ fails to be exact.

EXAMPLE 17.4 Consider the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow M_0 \longrightarrow 0 \longrightarrow \cdots.$$

in Example 16.10. Then $Z_0 = M_0$, $B_0 = 0$, and $B_i = Z_i = 0$ for all $i \neq 0$. Therefore $H_0 = M_0$ and $H_i = 0$ for all $i \neq 0$.

More generally for the chain complex in Example 16.11 (where all $d_i = 0$), we have $Z_i = M_i$ and $B_i = 0$ for all i, so $H_i = M_i$ for all i.

EXAMPLE 17.5 For the chain complex

$$\cdots \xrightarrow{d_3} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_2} \mathbb{Z}/4\mathbb{Z} \xrightarrow{d_1} \mathbb{Z}/4\mathbb{Z} \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \cdots$$

of Example 16.14, we have

- $Z_i = B_i = 2\mathbb{Z}/4\mathbb{Z}$ if $i > 0$, so $H_i = 0$;
- $Z_0 = \mathbb{Z}/4\mathbb{Z}$ and $B_0 = 2\mathbb{Z}/4\mathbb{Z}$, so $H_0 \cong \mathbb{Z}/2\mathbb{Z}$;
- $Z_i = B_i = 0$ if $i < 0$, so $H_i = 0$.

Given a complex $(P_\bullet, d_\bullet)$, write $H_i((P_\bullet, d_\bullet)$ or simply $H_i(P_\bullet)$ for the i^{th} homology of $(P_\bullet, d_\bullet)$.

The notion of a "morphism" of chain complexes is provided by a chain map:

DEFINITION 17.6 Suppose that $(A_\bullet, e_\bullet)$ and $(A'_\bullet, e'_\bullet)$ are chain complexes (of R-modules). A *chain map* $\varphi_\bullet : (A_\bullet, e_\bullet) \to (A'_\bullet, e'_\bullet)$ is a sequence of R-module homomorphisms $\varphi_i : A_i \to A'_i$ such that the squares in the diagram

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & A_{i+1} & \xrightarrow{e_{i+1}} & A_i & \xrightarrow{e_i} & A_{i-1} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle \varphi_{i+1}} & & \downarrow{\scriptstyle \varphi_i} & & \downarrow{\scriptstyle \varphi_{i-1}} & & \\ \cdots & \longrightarrow & A'_{i+1} & \xrightarrow{e'_{i+1}} & A'_i & \xrightarrow{e'_i} & A'_{i-1} & \longrightarrow & \cdots \end{array} \tag{15}$$

commute, i.e. $\varphi_{i-1} \circ d_i = d'_i \circ \varphi_i$ for all $n \in \mathbb{Z}$.

EXERCISE 17.7 Show that taking chain maps as morphisms makes chain complexes of R-modules into a category, denoted Ch(R-mod).

LEMMA 17.8 *Given a chain map $\varphi_\bullet : (A_\bullet, e_\bullet) \to (A'_\bullet, e'_\bullet)$, for every $i \in \mathbb{Z}$ there is an induced map on homology $\varphi_{i,*}\colon H_i((A_\bullet, e_\bullet)) \to H_i((A'_\bullet, e'_\bullet))$ which given $\bar{a} \in H_i((A_\bullet, e_\bullet)) = \ker(e_i)/\operatorname{im}(e_{i+1})$ is defined by $\overline{\varphi_{i,*}(a)} \in \ker(e'_i)/\operatorname{im}(e'_{i+1})$.*

Proof. First note that the commutativity of the right hand square of (15) ensures that $\varphi_i(\ker(e_i)) \subseteq \ker(e'_i)$. So we have an R-module map $\ker(e_i) \to \ker(e'_i)/\operatorname{im}(e'_{i+1})$ defined by the above formula. Now the commutativity of the left hand square ensures that $\varphi_i(\operatorname{im}(e_{i+1})) \subseteq \operatorname{im}(e'_{i+1})$. So by the first isomorphism theorem, the map descends to a map $\ker(e_i)/\operatorname{im}(e_{i+1}) \to \ker(e'_i)/\operatorname{im}(e'_{i+1})$ as desired. □

EXAMPLE 17.9 Let $(A_\bullet, e_\bullet)$, $(A'_\bullet, e'_\bullet)$ be the chain complexes define by the rows of the following diagram and $\varphi_\bullet$ the chain map:

$$\begin{array}{lccccccccc} (A, \bullet, e_\bullet) & \cdots & \longrightarrow & 0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\pi} & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 & \longrightarrow & \cdots \\ & & & \downarrow & & \downarrow{\scriptstyle 2} & & \downarrow & & \downarrow & & \\ (A'_\bullet, e'_\bullet) & \cdots & \longrightarrow & 0 & \longrightarrow & \mathbb{Z} & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \end{array}$$

Here $\pi\colon \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is the usual projection and the complexes are numbered so that the only non-trivial φ_i is when $i = 0$. Then

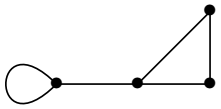
$$H_i(A_\bullet) = \begin{cases} 0 & i \neq 0 \\ 2\mathbb{Z} & i = 0 \end{cases}, \qquad H_i(A'_\bullet) = \begin{cases} 0 & i \neq 0 \\ \mathbb{Z} & i = 0 \end{cases}.$$

The map $H_0(A_\bullet) \to H_0(A'_\bullet)$ is then $2\mathbb{Z} \to \mathbb{Z}, 2n \mapsto 4n$.

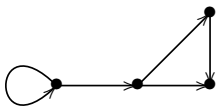
Homology of graphs

For reasons of time, we can't really explain why homology is so helpful in/arose from topology. We can however give a basic version, the "homology of graphs". We do not assume any familiarity with topological spaces!

Roughly speaking, a graph is a diagram like this:



consisting of vertices and edges, where each edge has vertices (possibly the same) as its endpoints. We in fact will be considering "directed graphs", where the edges have a direction. For example:



DEFINITION 17.10 A *directed graph* $\vec{\Gamma}$ is a disjoint pair of sets V and E and a function

$$\begin{array}{rccc} \vec{t}: & E & \to & V \times V \\ & e & \mapsto & (t_0(e), t_1(e)). \end{array}$$

The elements of V are called *vertices* of $\vec{\Gamma}$, the elements of E are called *edges* of $\vec{\Gamma}$, $t_0(e)$ is called the *initial endpoint* of e and $t_1(e)$ is called the *final endpoint* of e.

We will now show how to associate a chain complex to a directed graph $\vec{\Gamma} = (V, E, \vec{t})$. In turn we shall show how homology can be used to give algebraic invariants of directed graphs, that otherwise would be difficult to define.

Let $R = \mathbb{Z}$, and define the $\mathbb{Z}$-modules M_i as follows:

- $M_0 = F_V = \bigoplus_{v \in V} \mathbb{Z}v$ is the free $\mathbb{Z}$-module[40] generated by the vertices V;
- $M_1 = F_E = \bigoplus_{e \in E} \mathbb{Z}e$ is the free $\mathbb{Z}$-module generated by the edges E;
- $M_i = 0$ if $i \neq 0, 1$.

We then define $d_1 : M_1 \to M_0$ to be the unique homomorphism such that $d_1(e) = t_1(e) - t_0(e)$ for all $e \in E$. Recall that a homomorphism from a free module is determined by its effect on the basis elements; arbitrary elements of F_E have the form $\sum_{e \in E} n_e e$ (with all but finitely many $n_e = 0$), on which we have

$$d_1\left(\sum_{e \in E} n_e e\right) = \sum_{e \in E} n_e (t_1(e) - t_0(e)) \in M_0.$$

We have now defined a finite chain complex

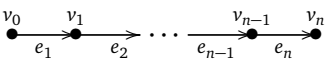
$$\cdots \longrightarrow 0 \longrightarrow M_1 \longrightarrow M_0 \longrightarrow 0 \longrightarrow \cdots.$$

We then have:

- $Z_0 = M_0$ and $B_0 = \operatorname{im}(d_1)$, so $H_0 = M_0 / \operatorname{im}(d_1)$;
- $Z_1 = \ker(d_1)$ and $B_1 = 0$, so $H_1 = \ker(d_1)$;
- $H_i = 0$ for $i \neq 0, 1$ (since $Z_i = B_i = 0$.

Let us now consider some explicit examples.

EXAMPLE 17.11 For $n \geq 0$, define $\vec{P}_n$ to be the directed graph with $V = \{v_0, v_1, \ldots, v_n\}$, $E = \{e_1, \ldots, e_n\}$ and $t(e_i) = (v_{i-1}, v_i)$:



We thus have:

[40] Our previous notation would have used e_v to denote the basis element of F_V associated to v, but since there is no risk of confusion, we will just write v instead of e_v, especially since the previous notation for bases would clash with our use of e for edges.

- $M_0 = \mathbb{Z}v_0 \oplus \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_n \cong \mathbb{Z}^{n+1}$;
- $M_1 = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2 \oplus \cdots \oplus \mathbb{Z}e_n \cong \mathbb{Z}^n$;
- $d_1 : M_1 \to M_0$ is defined by

$$d_1(e_1) = v_1 - v_0,\ d_1(e_2) = v_2 - v_1,\ \ldots,\ d_1(e_n) = v_n - v_{n-1}.$$

The resulting homomorphism[41] $\mathbb{Z}^n \to \mathbb{Z}^{n+1}$ therefore corresponds to the matrix

$$A = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 \\ 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

in the sense that $d_1(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} = {}^t(x_1, x_2, \ldots, x_n) \in \mathbb{Z}^n$ (see Example 10.8). Note that if $A\mathbf{x} = \mathbf{0}$, then

$$A\mathbf{x} = {}^t(-x_1,\ x_1 - x_2,\ \cdots,\ x_{n-1} - x_n,\ x_n) = \mathbf{0},$$

which easily implies that $\mathbf{x} = \mathbf{0}$. Therefore d_1 is injective, and hence $H_1 = Z_1 = 0$.

To determine $B_0 = \mathrm{im}(d_1)$, note that if $A\mathbf{x} = \mathbf{y} = {}^t(y_0, y_1, \ldots, y_n)$ for some $\mathbf{x} \in \mathbb{Z}^n$, then

$$\sum_{i=0}^{n} y_i = -x_1 + (x_1 - x_2) + \cdots + (x_{n-1} - x_n) + x_n = 0.$$

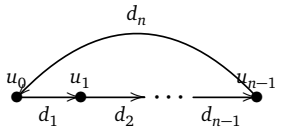
Conversely if $\mathbf{y} \in \mathbb{Z}^{n+1}$ is such that $\displaystyle\sum_{i=0}^{n} y_i = 0$, then $\displaystyle y_0 = -\sum_{i=1}^{n} y_i$, and letting

$$\mathbf{x} = (y_1 + y_2 + \cdots + y_n,\ y_2 + \cdots + y_n,\ \ldots,\ y_{n-1} + y_n,\ y_n)$$

gives $A\mathbf{x} = \mathbf{y}$. Therefore $B_0 = \ker(\deg)$, where $\deg : \mathbb{Z}^{n+1} \to \mathbb{Z}$ is defined by $\deg({}^t(y_0, y_1, \ldots, y_n)) = \displaystyle\sum_{i=0}^{n} y_i$. Since deg is surjective, it follows that

$$H_0 \cong \mathbb{Z}^{n+1}/B_0 = \mathbb{Z}^{n+1}/\ker(\deg) \cong \mathrm{im}(\deg) = \mathbb{Z}.$$

EXAMPLE 17.12 Consider now the directed graph $\vec{C}_n$ (a "loop") for $n \geq 1$ defined by $V = \{v_0, v_1, \ldots, v_{n-1}\}$, $E = \{e_1, \ldots, e_n\}$, $t(e_i) = (v_{i-1}, v_i)$ for $i = 1, \ldots, n-1$ and $t(e_n) = (v_{n-1}, v_0)$:



We have $M_0 \cong \mathbb{Z}^n$, $M_1 \cong \mathbb{Z}^n$ and $d_1 : M_1 \to M_0$ is given by the matrix

$$A = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 & 1 \\ 1 & -1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}.$$

[41] Identifying $\sum_{i=1}^{n} y_i e_i \in M_0$ with the column vector $\mathbf{y} = {}^t(y_1, y_2, \ldots, y_n) \in \mathbb{Z}^n$, and similarly $\sum_{i=0}^{n} x_i v_i \in M_0$ with the column vector $\mathbf{x} = {}^t(x_0, x_1, \ldots, x_n) \in \mathbb{Z}^{n+1}$.

Note that d_1 is no longer injective since

$$\begin{aligned} &d_1(e_n + e_{n-1} + \cdots + e_2 + e_1) \\ &\quad = (v_0 - v_{n-1}) + (v_{n-1} - v_{n-2}) + \cdots + (v_2 - v_1) + (v_1 - v_0) = 0. \end{aligned}$$

We leave it as exercise to show that in fact $H_1 = Z_1 = \mathbb{Z}\mathbf{x}$ where $\mathbf{x} = \sum_{i=1}^{n} e_n$, so that $H_1 \cong \mathbb{Z}$, and that $H_0 \cong \mathbb{Z}$ as in the preceding example.

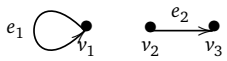
EXAMPLE 17.13 Consider the "bouquet" graph with one vertex v and n loops $e_1, \ldots, e_n$, so

- $M_0 = \mathbb{Z}v \cong \mathbb{Z}$,
- $M_1 = \bigoplus_{i=1}^{n} \mathbb{Z}e_i \cong \mathbb{Z}^n$,
- and $d_1(e_i) = v - v = 0$ for $i = 1, \ldots, n$, so $d_1 = 0$.

Since $d_1 = 0$, we have $H_1 = M_1 \cong \mathbb{Z}^n$ and $H_0 = M_0 \cong \mathbb{Z}$.

The examples above reflect a relation between H_1 and circuits in $\vec{\Gamma}$. Note also that H_0 was isomorphic to $\mathbb{Z}$; we leave it as an Exercise to show this is always the case if the graph is connected. More generally H_0 can be viewed as counting connected components, as in the following example:

EXAMPLE 17.14 Consider the directed graph with $V = \{v_1, v_2, v_3\}$, $E = \{e_1, e_2\}$, and $\vec{t} : E \to V \times V$ defined by $\vec{t}(e_1) = (v_1, v_1)$ and $\vec{t}(e_2) = (v_2, v_3)$, so the graph is represented by:



We then have $Z_1 = \mathbb{Z}e_1$ and $B_0 = \mathbb{Z}(v_2 - v_3)$, so that $H_1 \cong \mathbb{Z}$ and $H_0 = M_0/B_0 \cong \mathbb{Z}^2$, where one can realize the latter isomorphism, for example, by noting that the surjective homomorphism $\alpha : M_0 \to \mathbb{Z}^2$ defined by

$$\alpha(x_1 v_1 + x_2 v_2 + x_3 v_3) = (x_1, x_2 + x_3)$$

has kernel B_0.

EXERCISE* 17.15 Show that an "inclusion" of directed graphs yields a chain map between their associated chain complexes. Find examples of such inclusions where the induced map on homology is <u>not</u> injective.

18 Snake Lemma

In this section, we consider the following set up. Suppose we have a commutative diagram of R-modules:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \xrightarrow{f_1} & B_1 & \xrightarrow{g_1} & C_1 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d_A} & & \downarrow{\scriptstyle d_B} & & \downarrow{\scriptstyle d_C} & & \\ 0 & \longrightarrow & A_0 & \xrightarrow{f_0} & B_0 & \xrightarrow{g_0} & C_0 & \longrightarrow & 0 \end{array} \tag{16}$$

where here the rows are short exact sequences (note we could consider this as a chain map between two complexes given by short exact sequences). Note that by commutativity of the left hand square, $f_1(\ker(d_A)) \subseteq \ker(d_B)$ and similarly for the right hand square. We obtain a sequence

$$0 \longrightarrow \ker(d_A) \xrightarrow{f_{1,*}} \ker(d_B) \xrightarrow{g_{1,*}} \ker(d_C).$$

(Where we write $f_{1,*}$, $g_{1,*}$ for the restrictions of f_1, g_1 to these kernels.) We claim this is exact. Certainly $g_{1,*} \circ f_{1,*} = 0$ since this was true for the full f_1, g_1. We also have that the map $\ker(d_A) \to \ker(d_B)$ is injective since f_1 is. It remains to show that $\operatorname{im}(f_{1,*}) = \ker(g_{1,*})$. But we know that the original sequence was exact, so that $\operatorname{im}(f_1) = \ker(g_1)$. This means that given $b \in \ker(g_{1,*})$, we can at least find an element $a \in A_1$ such that $f_1(a) = b$. But we know that $b \in \ker(d_B)$, so $d_B(b) = 0$. Since f_0 is injective, and by commutativity $f_0(d_A(a)) = d_B(b) = 0$, this forces that $a \in \ker(d_A)$ also, and we win!

DEFINITION 18.1 Given a homomorphism of R-modules $f : M \to N$, we call $B/\operatorname{im}(f)$ the *cokernel* of f and denote it by $\operatorname{coker}(f)$.

Given the diagram (16), we also obtain a sequence

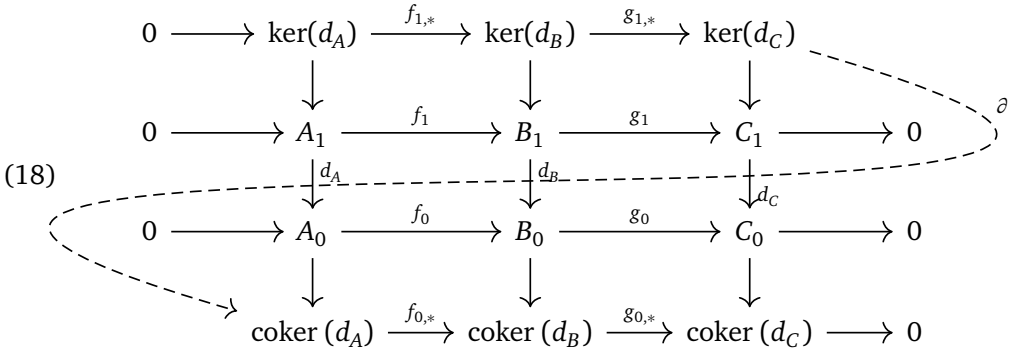
$$\operatorname{coker}(d_A) \xrightarrow{f_{0,*}} \operatorname{coker}(d_B) \xrightarrow{g_{0,*}} \operatorname{coker}(d_C) \to 0. \tag{17}$$

For example, commutativity of the left hand square gives that $f_0(\operatorname{im}(d_A)) \subseteq \operatorname{im}(d_B)$, and so by the first isomorphism theorem, we obtain a map $\operatorname{coker}(d_A) = A_0/\operatorname{im}(d_A) \to B_0/\operatorname{im}(d_B) = \operatorname{coker}(d_B)$, which we are denoting by $f_{0,*}$. We then have the analogue of the exactness of the kernel sequence:

EXERCISE 18.2 Show that in the setup of (16), we have that the cokernel sequence (17) is exact.

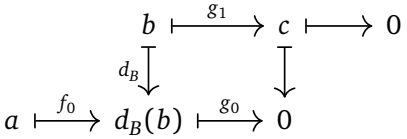
The key part of the snake lemma says that kernel sequence and cokernel sequence can be joined together into a single exact sequence by means of an R-module homomorphism called the "boundary map" ∂:

(18)



The construction of the image of an element $c \in \ker(d_C)$ under the boundary map ∂ is as follows. Consider c as lying in C_1. Then we can lift under g_1 it to an element $b \in B_1$ since g_1 is surjective. The image of b under d_B lies in the kernel of g_0 by commutativity, so lies in the image of f_0 by exactness of the lower sequence. The preimage a considered as an element of $\operatorname{coker}(d_A)$ is our element $\partial(c)$:

(19)



We must check ∂ is well-defined. The only choice was the preimage b of c. Suppose that b' is a different choice and a' is the preimage of $d_B(b')$. Then $(b - b') \overset{g_1}{\mapsto} (c - c) = 0$. So by exactness of the top sequence, $b - b'$ has a preimage in A_1. Call this $\tilde{a}$. Then $d_A(\tilde{a})$ is equal to $a - a'$ (since f_0 is injective, it suffices to check this equality after applying f_0, but then by commutativity of the left hand square $f_0(d_A(\tilde{a})) = d_B(f_1(\tilde{a})) = d_B(b - b') =$

$d_B(b)-d_B(b')=f_0(a)-f_0(a')$). So the difference between a and a' lies in $\mathrm{im}(d_A)$, that is to say they define equal elements of $\mathrm{coker}\,(d_A)$.

It is then easy to check that ∂ is an R-module homomorphism. For example, suppose that $c,c'\in\ker(d_C)$ with corresponding choices $b,b'\in B_1$ such that $\partial(c)=a,\partial(c')=a'$ with $f_0(a)=d_B(b)$, $f_0(a')d_B(b')$. Then $c+c'\in\ker(d_C)$ and we can choose our lift of $c+c'$ to be $b+b'$. This then makes $\partial(c+c')=a+a'=\partial(a)+\partial(a')$.

LEMMA 18.3 *Given a diagram of short exact sequences as in* (16)*, the following is an exact sequence of R-modules:*
(20)

$$0\to\ker(d_A)\xrightarrow{f_{1,*}}\ker(d_B)\xrightarrow{g_{1,*}}\ker(d_C)\xrightarrow{\partial}\mathrm{coker}\,(d_A)\xrightarrow{f_{0,*}}\mathrm{coker}\,(d_B)\xrightarrow{g_{0,*}}\mathrm{coker}\,(d_C)\to 0.$$

Proof. It remains to show exactness in the middle two positions. For $\ker(d_C)$ note first that if $c\in\mathrm{im}(g_{1,*}\colon\ker(d_B)\to\ker(d_C))$, then we can choose our lift b in (19) to lie in $\ker(d_B)$. But then $d_B(b)=0$ as does a. This means that $\mathrm{im}(g_{1,*})\subseteq\ker(\partial)$. Conversely, suppose that $c\in\ker(\partial)$. Then $a\in\mathrm{im}(d_A)$ and so has a preimage under d_A, call this $\tilde{a}$. We don't know that $f_1(\tilde{a})=b$, but we do know that $g_1(f_1(\tilde{a}))=0$ (the sequence is a complex), so $(b-f_1(\tilde{a}))\xmapsto{g_1}(c-0)=c$. But $d_B(b-f_1(\tilde{a}))=d_B(b)-f_0(d_A(\tilde{a}))=d_B(b)-f_0(a)=0$ (by assumption a is the preimage of $d_B(b)$ under f_0). So c is in the image of an element of $\ker(d_B)$ under $g_{1,*}$.

The proof of exactness at $\mathrm{coker}\,(d_A)$ is a similar fun diagram chase! □

We remark that in diagram (16), if we drop the exactness at A_1 (i.e. omit the initial zero), then we still obtain a snake lemma sequence which is exact except at the first $\ker(d_A)$ (check this!). Similarly, we can drop exactness at C_0 at the expense of losing exactness at $\mathrm{coker}\,(d_C)$.

EXAMPLE 18.4 In the case of the map of short exact sequences:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & 0 & \longrightarrow & \mathbb{Z} & = & \mathbb{Z} & \longrightarrow & 0\\
 & & \downarrow & & \downarrow{\scriptstyle 2} & & \downarrow & & \\
0 & \longrightarrow & \mathbb{Z} & = & \mathbb{Z} & \longrightarrow & 0 & \longrightarrow & 0
\end{array}$$

We have that the snake sequence (20) is given by

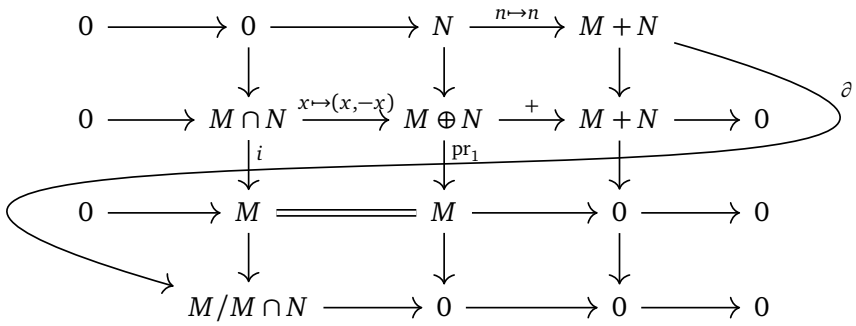
$$0\to 0\to 0\to\mathbb{Z}\xrightarrow{2}\mathbb{Z}\to\mathbb{Z}/2\mathbb{Z}\to 0\to 0.$$

EXAMPLE 18.5 Suppose that N,M are both R-submodules of an R-module T. We will prove the second isomorphism theorem, i.e. that $(M+N)/N$ is canonically isomorphic to $M/N\cap M$, using the snake lemma. Here $M+N$ denotes the submodule $\langle M,N\rangle$ of T generated M,N. Consider the diagram:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & M\cap N & \xrightarrow{x\mapsto(x,-x)} & M\oplus N & \xrightarrow{+} & M+N & \longrightarrow & 0\\
 & & \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle \mathrm{pr}_1} & & \downarrow & & \\
0 & \longrightarrow & M & = & M & \longrightarrow & 0 & \longrightarrow & 0
\end{array}$$

Here the map $+$ denotes the map induced by the universal property applied to the obvious inclusions $M\hookrightarrow M+N$, $N\hookrightarrow M+N$ (this involves summing, see Proposition 9.7). This is exact as the kernel of $+$ consists exactly of pairs (m,n) such that $m+n=0$ (check!). The lower sequence is obviously exact and the squares commute, so we can

apply the snake lemma. When we compute the kernels and cokernels, we obtain:



That $0 \to N \to (M+N) \xrightarrow{\partial} M/M\cap N \to 0$ is exact is precisely the claim that $(M+N)/N \cong M/M\cap N$.

Long exact sequences from short exact sequences of chain complexes

In the remainder of this section, we prove one of the most important results in homological algebra (an upgraded/iterated snake lemma).

DEFINITION 18.6 Let $(A_\bullet, d_{A,\bullet}), (B_\bullet, d_{B,\bullet}), (C_\bullet, d_{C,\bullet})$ be chain complexes and let 0 denote the chain complex with all zeroes. We say a sequence of chain maps $0 \to A_\bullet \xrightarrow{f_\bullet} B_\bullet \xrightarrow{g_\bullet} C_\bullet \to 0$ is a *short exact sequence of chain complexes* if the "horizontal rows"

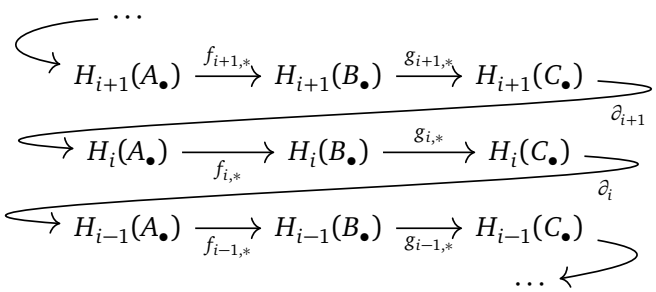
$$0 \to A_i \xrightarrow{f_i} B_i \xrightarrow{g_i} C_i \to 0$$

are exact for all $i \in \mathbb{Z}$.

By Lemma 17.8, we know that given a short exact sequence of complexes, we obtain R-module homomorphisms

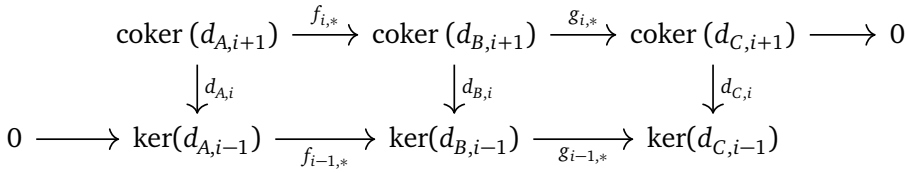
$$H_i(A_\bullet) \xrightarrow{f_{i,*}} H_i(B_\bullet) \xrightarrow{g_{i,*}} H_i(C_\bullet) \qquad \forall i \in \mathbb{Z}.$$

We claim there are R-module homomorphism boundary maps $\partial_i : H_i(C_\bullet) \to H_{i-1}(A_\bullet)$ such that



is a long exact sequence.

We construct each boundary map ∂_i by applying the snake lemma to the diagram



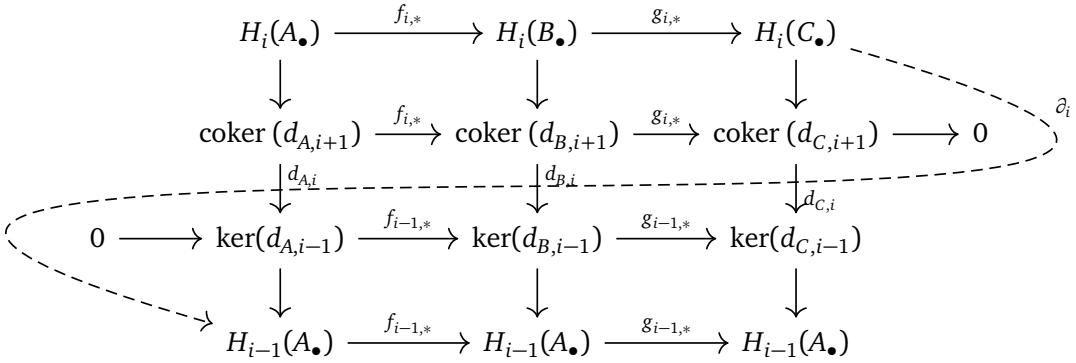
Note we showed the exactness of the top and bottom rows en route to showing the snake lemma (recall that the snake lemma can be weakened to omit exactness at the first and last positions!). The vertical maps

$$A_i/\operatorname{im}(d_{A,i+1}) \xrightarrow{d_{A,i}} \ker(d_{A,i-1}) \subseteq A_{i-1}$$

exist as the image of $d_{A,i}: A_i \to A_{i+1}$ is contained in $\ker(d_{A,i-1})$ since $(A_\bullet, d_{A,\bullet})$ is a complex and this factors through $A_i/\operatorname{im}(d_{A,i+1})$ since $d_{A,i}(d_{A,i+1}(A_{i+1}) = 0$ again because $(A_\bullet, d_{A,\bullet})$ is a complex.

We now want to calculate the kernels and cokernels of the vertical maps. Consider the vertical cokernel first, i.e. $\ker(d_{A,i-1})/\operatorname{im}(\operatorname{coker}(d_{A,i+1}))$. This "numerator" consists of the cycles Z_{i-1} for $(A_\bullet, d_{A,\bullet})$. The "denominator" is actually then the boundaries B_{i-1}. In other words, the image of $\operatorname{coker}(d_{A,i+1})$ is the same as the image of the whole of A_{i+1}. This is clear as $\operatorname{coker}(d_{A,i+1}) = A_{i+1}/\operatorname{im}(d_{A,i+1})$ (with $\operatorname{im}(d_{A,i+1})$ mapping to zero in A_{i-1}). In other words, the cokernel of the left-hand vertical map is $Z_i/B_i = H_i(A_\bullet)$ (and similarly for the other vertical maps). For the vertical kernels simply note that $\ker(\operatorname{coker}(d_{A,i+1}) \xrightarrow{d_{A,i}} \ker(d_{A,i-1})) = \ker((A_i/B_i) \xrightarrow{d_{A,i}} A_{i-1})$ and this is just $Z_i/B_i = H_i(A_\bullet)$.

We can then apply the snake lemma to get the boundary ∂_i



LEMMA 18.7 *Given a short exact sequence of chain complexes, $0 \to A_\bullet \xrightarrow{f_\bullet} B_\bullet \xrightarrow{g_\bullet} C_\bullet \to 0$, we obtain a long exact sequence of their cohomology groups*

$$\cdots \to H_{i+1}(C_\bullet) \xrightarrow{\partial_{i+1}} H_i(A_\bullet) \xrightarrow{f_{i,*}} H_i(B_\bullet) \xrightarrow{g_{i,*}} H_i(C_\bullet) \xrightarrow{\partial_i} H_{i-1}(A_\bullet) \to \cdots.$$

Proof. We should remark that the iterated snake lemmas really do "paste" together as the maps $H_i(A_\bullet) \xrightarrow{f_{i,*}} H_i(B_\bullet) \xrightarrow{g_{i,*}} H_i(C_\bullet)$ are the same for the two applications of the snake lemma that contain them. We then get exactness in every position by the conclusion of (at least) one snake lemmas. □

REMARK 18.8 For applications, it is extremely important to know that both the snake lemma and Lemma 18.7 are "functorial[42]". All we shall say here is that chain maps between chain complexes induce maps on homology. So if we have a short exact sequence of chain complexes mapping by chain maps to another short exact sequence of complexes, then we obtain induced maps between their homology long exact sequences (and in fact a chain map between them). This will be heavily exploited in §23.

[42] By which we mean, if you have maps between the input objects of these lemmas, you receive maps between their outputs. These maps should satisfy the conditions of Definition 13.2 so as to define a functor.

19 Exactness of functors

Recall that if M is an R-module and $\psi : A \to B$ is a homomorphism of R-modules, then there is an R-module homomorphism

$$\mathrm{id}_M \otimes \psi : M \otimes_R A \longrightarrow M \otimes_R B$$

characterized by the property that $(\mathrm{id}_M \otimes \psi)(m \otimes n) = m \otimes \psi(a)$ for all $m \in M$, $a \in A$ (see Example 13.8).

We claim that if ψ is surjective, then so is $\mathrm{id}_M \otimes \psi$. Indeed $M \otimes_R B$ is generated[43] as an R-module by the set of elements of the form $m \otimes b$ (for $m \in M$, $b \in B$), i.e. if $x \in M \otimes_R B$, then $x = \sum_{i=1}^{k} r_i(m_i \otimes b_i)$ for some $k \geq 1$, $m_1, \ldots, m_k \in M$ and $b_1, \ldots, b_k \in B$. Since ψ is surjective, we have $b_i = \psi(a_i)$ for some $a_i \in A$, and hence

$$x = \sum_{i=1}^{k} r_i(m_i \otimes \psi(a_i)) = (\mathrm{id}_M \otimes \psi)\left(\sum_{i=1}^{k} r_i(m_i \otimes a_i)\right)$$

is in the image of $\mathrm{id}_M \otimes \psi$.

On the other hand if ψ is injective, then it need not be the case that $\mathrm{id}_M \otimes \psi$ be injective:

EXAMPLE 19.1 Let $R = \mathbb{Z}$, $M = \mathbb{Z}/2\mathbb{Z}$, $A = B = \mathbb{Z}$ and define $\psi : \mathbb{Z} \to \mathbb{Z}$ by $\psi(a) = 2a$. Then clearly ψ is injective, but consider the homomorphism

$$\mathrm{id}_M \otimes \psi : (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z} \longrightarrow (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}.$$

Recall from Example 12.3 (and Proposition 12.9(i)) that $(\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ (with $\overline{m} \otimes a$ corresponding to $a \cdot \overline{m} = \overline{ma}$, where as usual $\overline{m}$ denotes $m + 2\mathbb{Z}$). Under these isomorphisms, $\mathrm{id}_M \otimes \psi$ corresponds to the homomorphism

$$\begin{array}{ccccccc} \mathbb{Z}/2\mathbb{Z} & \cong & (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z} & \xrightarrow{\mathrm{id}_M \otimes \psi} & (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z} & \cong & \mathbb{Z}/2\mathbb{Z} \\ \overline{m} & \leftrightarrow & \overline{m} \otimes 1 & \mapsto & \overline{m} \otimes 2 & \leftrightarrow & \overline{2m} = \overline{0}, \end{array}$$

which, being 0, is clearly not injective.

We can also consider the preceding example in the context of the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\psi} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0 \tag{21}$$

of Example 16.4 (with $m = 2$, and π defined by $\pi(a) = \overline{a}$). Applying the functor $F : \mathbb{Z}\text{-Mod} \to \mathbb{Z}\text{-Mod}$ defined by tensoring with $M = \mathbb{Z}/2\mathbb{Z}$, i.e. $F(A) = M \otimes_{\mathbb{Z}} A$ for all $\mathbb{Z}$-modules A and $F(\varphi) = \mathrm{id}_M \otimes \varphi$ for all $\mathbb{Z}$-module homomorphisms φ (see Example 13.8), we obtain from (21) another sequence of homomorphisms

$$F(0) \longrightarrow F(\mathbb{Z}) \xrightarrow{F(\psi)} F(\mathbb{Z}) \xrightarrow{F(\pi)} F(\mathbb{Z}/2\mathbb{Z}) \longrightarrow F(0) \tag{22}$$

of $\mathbb{Z}$-modules. The new sequence however is *not* exact. Indeed we already recalled how $F(\mathbb{Z}) = (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}$ can be identified with $\mathbb{Z}/2\mathbb{Z}$; note also that $F(0) = 0$ and $F(\mathbb{Z}/2\mathbb{Z}) = (\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$ can be identified with $\mathbb{Z}/2\mathbb{Z}$ (see Example 12.4). We already saw that $F(\psi) = 0$, and similarly one sees that $F(\pi)$ corresponds to the identity

[43] Recall that an arbitrary element of $M \otimes_R B$ is not necessarily of the form $m \otimes b$, but is a linear combination of such elements.

(as a homomorphism $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$), so that (22) can be viewed as the sequence of homomorphisms

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{0} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\mathrm{id}} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

This is clearly not exact since the first homomorphism $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ fails to be injective. Thus the functor F does not preserve exact sequences.

We now give an example of a different functor that occurs in nature but is not exact.

Let us again fix a ring R and an R-module M, but now define a functor $F : R\text{-Mod} \to R\text{-Mod}$ by $F(A) = \mathrm{Hom}_R(M,A)$ for all R-modules A. Recall that we already explained in §10 how to view $\mathrm{Hom}_R(M,A)$ as an R-module, but in order to complete the definition of the functor F, we also need to describe its effect on morphisms. More precisely, if $\varphi : A \to B$ is any homomorphism of R-modules, then we need to define the R-module homomorphism

$$F(\varphi) : F(A) = \mathrm{Hom}_R(M,A) \longrightarrow \mathrm{Hom}_R(M,B) = F(B)$$

(and verify that it satisfies properties (i) and (ii) in Definition 13.2).

Note that if $\varphi \in \mathrm{Hom}_R(A,B)$ and $f \in F(A) = \mathrm{Hom}_R(M,A)$, then the composite $\varphi \circ f$ is an element of $\mathrm{Hom}_R(M,B) = F(B)$, so we may define

$$\begin{array}{rccc} F(\varphi): & F(A) & \longrightarrow & F(B) \\ & f & \longmapsto & \varphi \circ f; \end{array}$$

i.e. $F(\varphi)$ is the function defined by "post-composition with φ". To see that $F(\varphi) \in \mathrm{Hom}_R(F(A),F(B))$, we need to check that $F(\varphi)$ is R-linear. To ease notation, let us write φ_* for $F(\varphi)$, i.e. $\varphi_*(f) = \varphi \circ f$ for $f \in F(A) = \mathrm{Hom}_R(M,A)$, and we need to check that

- $\varphi_*(f+g) = \varphi_*(f) + \varphi_*(g)$ for all $f, g \in F(A)$,
- and $\varphi_*(rf) = r(\varphi_*(f))$ for all $r \in R$, $f \in F(A)$.

Since $f+g$ is defined by $(f+g)(m) = f(m) + g(m)$ (for $m \in M$), so

$$\begin{aligned} (\varphi_*(f+g))(m) &= (\varphi \circ (f+g))(m) = \varphi((f+g)(m)) \\ &= \varphi(f(m) + g(m)) = \varphi(f(m)) + \varphi(g(m)) \end{aligned}$$

(since φ is R-linear), and this is the same as

$$(\varphi_*(f) + \varphi_*(g))(m) = (\varphi_*(f))(m) + (\varphi_*(g))(m) = (\varphi \circ f)(m) + (\varphi \circ g)(m),$$

showing that $\varphi_*(f+g) = \varphi_*(f) + \varphi_*(g)$. Similarly

$$(\varphi_*(rf))(m) = (\varphi \circ (rf))(m) = \varphi((rf)(m)) = \varphi(r(f(m))) = r(\varphi(f(m)))$$

is the same as $(r(\varphi_*(f)))(m) = r((\varphi_*(f))(m))$, so $\varphi_*(rf) = r(\varphi_*(f))$.

We have now defined the functions

$$\begin{array}{ccc} \mathrm{Hom}_R(A,B) & \longrightarrow & \mathrm{Hom}_R(F(A),F(B)) \\ \varphi & \longmapsto & F(\varphi) = \varphi_*. \end{array}$$

To conclude that F is a functor, note that $\varphi = \mathrm{id}_A \in \mathrm{Hom}_R(A,A)$, then $\varphi_*(f) = \mathrm{id}_A \circ f = f$ for all $f \in F(A)$, so $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$. Furthermore if $\varphi \in \mathrm{Hom}_R(A,B)$ and $\psi \in \mathrm{Hom}_R(B,C)$, then the composite

$$F(A) \xrightarrow{\varphi_*} F(B) \xrightarrow{\psi_*} F(C)$$

is defined by

$$\psi_*(\varphi_*(f)) = \psi_*(\varphi \circ f) = \psi \circ (\varphi \circ f) = (\psi \circ \varphi) \circ f = (\psi \circ \varphi)_*(f),$$

so $F(\psi)\circ F(\varphi)=\psi_*\circ\varphi_*=(\psi\circ\varphi)_*=F(\psi\circ\varphi)$ as required. Therefore F is a functor, commonly denoted $\mathrm{Hom}_R(M,-)$.

Again the functor F might not preserve exact sequences. We can use the same example as before. Letting $R=\mathbb{Z}$, $M=\mathbb{Z}/2\mathbb{Z}$ and applying F to the short exact sequence (21) now gives another sequences of homomorphisms as in (22), but now with $F(A)=\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},A)$. Note that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z})=0$ (as there are no non-zero homomorphisms $\mathbb{Z}/2\mathbb{Z}\to\mathbb{Z}$), and $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z})$ can be identified with $\mathbb{Z}/2\mathbb{Z}$, so (22) becomes

$$0\longrightarrow 0\longrightarrow 0\longrightarrow \mathbb{Z}/2\mathbb{Z}\longrightarrow 0,$$

which is clearly not exact.

We now introduce terminology in order to make this notion more precise. We restrict our attention to functors on categories of modules over commutative rings. Suppose that R and S are commutative rings, and $F:R\text{-Mod}\to S\text{-Mod}$ is a functor. For example, we could have $R=S$ and F as one of the functors discussed above, or given a ring homomorphism $f:R\to S$, we have the functor F defined by $F(A)=S\otimes_R A$ and $F(\varphi)=\mathrm{id}_S\otimes\varphi$ (see §13, where the functor is denoted G).

First recall that if A and B are R-modules, then $\mathrm{Hom}_R(A,B)$ has the structure of an R-module, hence an abelian group. Similarly since $\mathrm{Hom}_S(F(A),F(B))$ is an S-module, it is an abelian group.

DEFINITION 19.2 We say that a functor $F:R\text{-Mod}\to S\text{-Mod}$ is *additive* if the function

$$\begin{array}{ccc}\mathrm{Hom}_R(A,B) & \longrightarrow & \mathrm{Hom}_S(F(A),F(B))\\ \varphi & \longmapsto & F(\varphi)\end{array}$$

is a homomorphism of abelian groups for all R-modules A and B, i.e. $F(f+g)=F(f)+F(g)$ for all $f,g\in\mathrm{Hom}_R(A,B)$.

The examples discussed above, with $F(A)=M\otimes_R A$, $\mathrm{Hom}_R(M,A)$ (for a fixed R-module M) or $S\otimes_R A$ (for a fixed ring homomorphism $R\to S$), are all easily seen to be additive.

Recall that the identity element of the abelian group $\mathrm{Hom}_R(A,B)$ is the homomorphism $A\xrightarrow{0}B$, and that of the group $\mathrm{Hom}_S(F(A),F(B))$ is $F(A)\xrightarrow{0}F(B)$, so if F is additive, then[44] $F(0)=0$.

REMARK 19.3 Additive functors are defined in terms of respecting addition of morphisms. In fact, this strongly constrains how an additive functor behaves on objects also. For example, we claim that if $A=\{0\}$ and F is additive, then $F(A)=\{0\}$. Indeed if $F(A)=\{0\}$, then $\mathrm{id}_A=0$ (i.e. the identity function $A\xrightarrow{\mathrm{id}_A}A$ on A is the same as the constant function $A\xrightarrow{0}A$ sending every element of A to 0). Therefore

$$\mathrm{id}_{F(A)}=F(\mathrm{id}_A)=F(0)=0,$$

i.e. the identity function $F(A)\xrightarrow{\mathrm{id}_{F(A)}}F(A)$ is the same as the constant function $F(A)\xrightarrow{0}F(A)$, from which it follows that $F(A)=\{0\}$. Therefore if F is additive, then we can also view $F(0)=0$ as an equality of S-modules, where as usual, we simply write 0 for any R-module (or S-module) of the form $\{0\}$.

[44] Note that $A\xrightarrow{0}B$ and $F(A)\xrightarrow{0}F(B)$ are typically functions between different pairs of modules, but we are using 0 to denote any homomorphism with constant value 0 (i.e. the additive identity element of the target module).

DEFINITION 19.4 Suppose that $F : R\text{-Mod} \to S\text{-Mod}$ is an additive functor. We say that F is *left-exact* if for every exact sequence

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

of R-modules, the resulting sequence

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

of S-modules is also exact. Similarly we say that F is *right-exact* if for every exact sequence

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

of R-modules, the resulting sequence

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

of S-modules is also exact. Finally we say that F is *exact* if it is both left-exact and right-exact.

EXAMPLE 19.5 Let $f : S \to R$ be a ring homomorphism, and consider the functor $F : R\text{-Mod} \to S\text{-Mod}$ of Example 13.3 (with the roles of R and S reversed), i.e. if A is an R-module, then $F(A) = A$ as an S-module (with the same addition operation, and with scalar multiplication defined by $s \cdot a = f(s)a$ for $s \in S$, $a \in A$), and $F(\varphi) = \varphi$ for any R-module (hence S-module) homomorphism $A \xrightarrow{\varphi} B$.

For any sequence of R-module homomorphisms

$$A_n \xrightarrow{\varphi_n} A_{n-1} \xrightarrow{\varphi_{n-1}} \cdots \xrightarrow{\varphi_2} A_1 \xrightarrow{\varphi_1} A_0,$$

the conditions that determine exactness, i.e. $\ker(\varphi_i) = \operatorname{im}(\varphi_{i+1})$ for $i = 1, \ldots, n-1$, do not depend on whether the objects (and homomorphisms) are viewed as being (of) R-modules or S-modules. It follows that F is both left-exact and right-exact, and hence exact.

EXAMPLE 19.6 Let M be an R-module, and consider the functor $F : R\text{-Mod} \to R\text{-Mod}$ defined by $F(A) = \operatorname{Hom}_R(M, A)$ and $F(\varphi) = \varphi_*$ (as defined above).

We claim that F is left-exact. To prove this, suppose that

$$0 \longrightarrow A \xrightarrow{\varphi} B \xrightarrow{\psi} C$$

is an exact sequence of R-modules, i.e. $\ker(\varphi) = 0$ (or equivalently φ is injective), and $\operatorname{im}(\varphi) = \ker(\psi)$. We must prove that

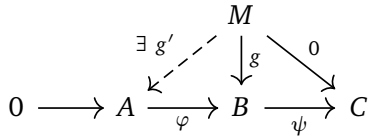
$$0 \longrightarrow \operatorname{Hom}_R(M, A) \xrightarrow{\varphi_*} \operatorname{Hom}_R(M, B) \xrightarrow{\psi_*} \operatorname{Hom}_R(M, C)$$

is also exact, i.e. that φ_* is injective and $\operatorname{im}(\varphi_*) = \ker(\psi_*)$.

To prove that φ_* is injective, we must show that if $f \in \operatorname{Hom}_R(M, A)$ is such that $\varphi_*(f) = 0$, then $f = 0$. Recall that $\varphi_*(f) = \varphi \circ f$, so if $\varphi_*(f) = 0$, then $\varphi(f(a)) = 0$ for all $a \in A$. Since φ is assumed to be injective, this implies that $f(a) = 0$ for all $a \in A$, i.e. $f = 0$.

Next note that since $\psi \circ \varphi = 0$, it follows for free that $\psi_* \circ \varphi_* = (\psi \circ \varphi)_* = 0$, and hence $\operatorname{im}(\varphi_*) \subset \ker(\psi_*)$ (alternatively put, additive functors always take chain complexes to chain complexes).

It remains to prove that if $g \in \ker(\psi_*)$ (i.e. if $\psi \circ g = 0$), then $g \in \mathrm{im}(\varphi_*)$ (i.e. there is some $g' \colon M \to A$ such that $g = \varphi \circ g'$). We can represent this as a saying that whenever g is such that the solid arrows commute:



there is an R-module homomorphism g' making the whole diagram commute. But this is actually tautological! Since $\psi \circ g = 0$, we must have that $\mathrm{im}\, g \subseteq \ker(\psi)$. Now using that the lower sequence is exact, we know that $\ker(\psi) = \mathrm{im}(\varphi)$. So the image of g is actually contained in $\mathrm{im}(\varphi)$. Since φ is injective, it is an R-module isomorphism onto its image (for example, by the first isomorphism theorem (why?)). So we can just let $g' = \varphi^{-1} \circ g$ (where $\varphi^{-1} \colon \mathrm{im}(\varphi) \xrightarrow{\sim} A$) so that $\varphi \circ g' = \varphi \circ \varphi^{-1} \circ g = g$ as desired. We have now shown that the functor F is left-exact, but note that it is not necessarily right-exact. Indeed we have (as part of (21)) the exact sequence

$$\mathbb{Z} \xrightarrow{\psi} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

(with $R = \mathbb{Z}$), but we have seen that for $M = \mathbb{Z}/2\mathbb{Z}$, the resulting sequence

$$F(\mathbb{Z}) \xrightarrow{F(\psi)} F(\mathbb{Z}) \xrightarrow{F(\pi)} F(\mathbb{Z}/2\mathbb{Z}) \longrightarrow 0$$

is not exact.

EXERCISE 19.7 Let $f : M \to N$ be injective. Show that if F is a left-exact functor, then $F(f) \colon F(M) \to F(N)$ is injective also "left-exact functors preserve injective maps". Similarly, show that right-exact functors preserve surjective maps[45].

We now give some alternative characterizations of exactness.

LEMMA 19.8 *Suppose that $F : R\text{-Mod} \to S\text{-Mod}$ is an additive functor. Then the following are equivalent:*

(i) F is exact;

(ii) if $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is an exact sequence of R-modules, then $0 \longrightarrow F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C) \longrightarrow 0$ is an exact sequence of S-modules;

(iii) if $A \xrightarrow{f} B \xrightarrow{g} C$ is an exact sequence of R-modules, then $F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C)$ is an exact sequence of S-modules.

Proof. (i) $\Rightarrow$ (ii): Suppose that F is exact (i.e. both left-exact and right-exact) and let $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ be an exact sequence of R-modules. Since F is left-exact and $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$ is exact, it follows that $0 \longrightarrow F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C)$ is an exact sequence (of S-modules). Similarly since F is right-exact and $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is exact, it follows that $F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C) \longrightarrow 0$ is exact[46]. Therefore $0 \longrightarrow F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C) \longrightarrow 0$ is exact.

[45] In fact, there is a stronger statement that an additive functor is left-exact if and only if it "preserves kernels" and right-exact if and only if it "preserves cokernels".

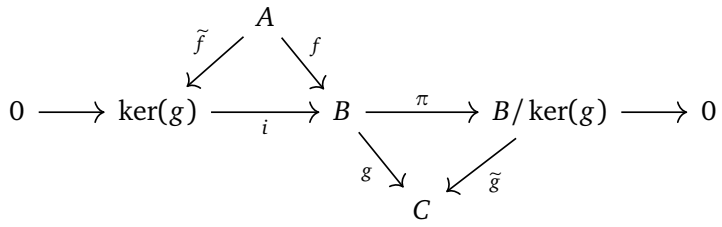
[46] Note that we already knew that $F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C)$ is exact; the new input is that we also now know $F(g)$ is surjective.

(ii) $\Rightarrow$ (iii): Suppose that (ii) holds, and let $A \xrightarrow{f} B \xrightarrow{g} C$ be an exact sequence of R-modules. Recall our aim is to prove that $\mathrm{im}(f) = \ker(g)$.

We need a short exact sequence to which we can apply (ii). We have the short exact sequence

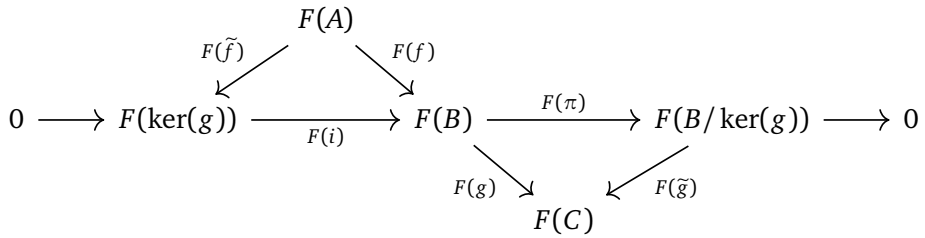
$$0 \longrightarrow \ker(g) \xrightarrow{i} B \xrightarrow{\pi} B/\ker(g) \longrightarrow 0,$$

where i is the inclusion ($i(d) = d$) and π the obvious projection ($\pi(b) = b + D$). We obtain the following diagram:



which has a short exact sequence in the middle and our original exact sequence going diagonally. Here, the homomorphism $\widetilde{f}$ exists as we are assuming the image of $\mathrm{im}(f) = \ker(g)$ by exactness and the homomorphism $\widetilde{g}$ exists by the first isomorphism theorem.

Now apply the functor F to the entire diagram. We obtain a commutative diagram (functors always take commutative diagrams to commutative diagrams (why?)):



By (ii) we know that the middle sequence is exact, so that $\mathrm{im}\, F(i) = \ker(F(\pi))$. But we also know that F takes injective maps to injective maps and surjective maps to surjective maps (by the exact same argument as in Exercise 19.7). Since $F(\widetilde{f})$ is therefore surjective, this means that $\mathrm{im}(F(i)) = \mathrm{im}(F(f))$ and since $F(\widetilde{g})$ is injective, this means that $\ker(F(g)) = \ker(F(\pi))$. We obtain

$$\mathrm{im}(F(f)) = \mathrm{im}(F(i)) = \ker(F(\pi)) = \ker(F(g)),$$

as desired.

(iii) $\Rightarrow$ (i): We must show that if (iii) holds, then F is left-exact and right-exact. Suppose then that $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$ is an exact sequence of R-modules. Then $0 \longrightarrow A \xrightarrow{f} B$ is exact, and therefore (by (iii)) so is $0 \longrightarrow F(A) \xrightarrow{F(f)} F(B)$. Similarly since $A \xrightarrow{f} B \xrightarrow{g} C$ is exact, so is $F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C)$. Putting these together, it follows that $0 \longrightarrow F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C)$ is exact. We have now shown that F is left-exact.

The proof that F is right-exact is entirely similar. □

20 Projective modules

Recall from Example 19.6 that if M is an R-module, then the functor $F : R\text{-Mod} \to S\text{-Mod}$ defined by $F(A) = \mathrm{Hom}_R(M,A)$ is left-exact, but it may fail to be right-exact. We will start denoting this functor F by $\mathrm{Hom}_R(M,-)$.

There are actually some choices of M for which $\mathrm{Hom}_R(M,-)$ is also right-exact (and hence exact):

DEFINITION 20.1 We say that R-module M is *projective* if the functor $\mathrm{Hom}_R(M,-)$ is exact.

EXAMPLE 20.2 The R-module R itself is projective. Indeed if A is any R-module, then there is an isomorphism of R-modules

$$\varepsilon : \mathrm{Hom}_R(R,A) \xrightarrow{\sim} A$$

defined by $\varepsilon(f) = f(1)$ (i.e. evaluate the homomorphism at 1). It is R-linear since

$$\varepsilon(f+g) = (f+g)(1) = f(1) + g(1) = \varepsilon(f) + \varepsilon(g)$$

and $\varepsilon(rf) = (rf)(1) = r(f(1)) = r\varepsilon(1)$ for all $f, g \in \mathrm{Hom}_R(R,A)$ and $r \in R$, and it is bijective since for any $a \in A$, there is a unique R-linear homomorphism $f : R \to A$ such that $f(1) = a$, namely $f(r) = ra$. Furthermore if $\varphi : A \to B$ is a homomorphism of R-modules, then the resulting diagram

$$\begin{array}{ccc} \mathrm{Hom}_R(R,A) & \xrightarrow{\varphi_*} & \mathrm{Hom}_R(R,B) \\ \downarrow{\scriptstyle \varepsilon_A}\wr & & \downarrow{\scriptstyle \varepsilon_B}\wr \\ A & \xrightarrow{\varphi} & B \end{array}$$

commutes, where we have added the relevant subscripts to the isomorphisms ε. This shows that $\mathrm{Hom}_R(R,A)$ is effectively the "identity functor" and so certainly exact. Explicitly, if $A \xrightarrow{\varphi} B \xrightarrow{\psi} C$ is an exact sequence of R-modules, the resulting commutative diagram

$$\begin{array}{ccccc} \mathrm{Hom}_R(R,A) & \xrightarrow{\varphi_*} & \mathrm{Hom}_R(R,B) & \xrightarrow{\psi_*} & \mathrm{Hom}_R(R,C) \\ \downarrow{\scriptstyle \varepsilon_A}\wr & & \downarrow{\scriptstyle \varepsilon_B}\wr & & \downarrow{\scriptstyle \varepsilon_C}\wr \\ A & \xrightarrow{\varphi} & B & \xrightarrow{\psi} & C, \end{array}$$

shows that $\mathrm{im}(\varphi_*) = \varepsilon_B^{-1}(\mathrm{im}(\varphi)) = \varepsilon_B^{-1}(\ker(\psi)) = \ker(\psi_*)$. Therefore the functor $\mathrm{Hom}_R(R,-)$ satisfies condition (iii) in Lemma 19.8.

Similarly, the R-module R^n is projective for any $n \geq 1$. In fact, the following proposition shows that every free R-module is projective.

PROPOSITION 20.3 *Let M be an R-module. Then the following are equivalent:*

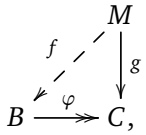
(i) M is projective;

(ii) if $\varphi : B \to C$ is a surjective homomorphism of R-modules, then the homomorphism

$$\varphi_* : \mathrm{Hom}_R(M,B) \longrightarrow \mathrm{Hom}_R(M,C)$$

is also surjective;

(iii) *if $g : M \to C$ and $\varphi : B \to C$ are homomorphisms of R-modules, and φ is surjective, then there exists a homomorphism $f : M \to B$ of R-modules, i.e. we can complete the diagram:*



so that $g = \varphi \circ f$,

(iv) *M is a direct summand of a free R-module, i.e. there exists an R-module N such that $M \oplus N$ is a free R-module.*

Proof. (i) $\implies$ (ii): Suppose that M is projective, and let $\varphi : B \to C$ be a surjective homomorphism of R-modules. Then the sequence $B \xrightarrow{\varphi} C \longrightarrow 0$ is exact, so by Lemma 19.8, so is the sequence $\mathrm{Hom}_R(M,B) \xrightarrow{\varphi_*} \mathrm{Hom}_R(M,C) \longrightarrow 0$, i.e. φ_* is surjective.

(ii) $\implies$ (i): Using Lemma 19.8, we want to show that given a short exact sequence $0 \to A \to B \to C \to 0$, the sequence after applying $\mathrm{Hom}_R(M,-)$ is exact. But we already know we have exactness except maybe at $\mathrm{Hom}_R(M,B) \to \mathrm{Hom}_R(M,C) \to \mathrm{Hom}_R(M,0)$ (Example 19.6). That this is exact is precisely (ii).

(ii) $\iff$ (iii) is just an unraveling of what it means for φ_* to be surjective, and (iii) $\iff$ (iv) is left as an exercise. □

EXAMPLE 20.4 It is immediate from Proposition 20.3 that if M is a free R-module then R is projective (take $N = 0$).

For an example of a projective module which is not free, let $R = \mathbb{Z}/6\mathbb{Z}$, and let $M = \mathbb{Z}/3\mathbb{Z}$ and $N = \mathbb{Z}/2\mathbb{Z}$ (viewed in the obvious way as R-modules). We then have an isomorphism $R \xrightarrow{\sim} M \oplus N$ of R-modules, so M and N are both projective, but they are clearly not free. (If M, for example, were free, then being finite, it would have to be isomorphic to R^n for some n, so its cardinality would have to be a power of 6.)

Before continuing the discussion of projective modules, we introduce another notion from category theory that will be useful later, namely contravariant functors.

We start with a motivating example. We have been working with the functor $\mathrm{Hom}_R(M,-)$ (for a fixed R and M), but let's consider what happens if we change this slightly and try to define a functor by $\mathrm{Hom}_R(-,M)$, i.e. a functor $F : R\text{-Mod} \to S\text{-Mod}$ such that if A is an R-module, then $F(A) = \mathrm{Hom}_R(A,M)$. We then also need to define $F(\varphi) \in \mathrm{Hom}_R(F(A),F(B))$ for each homomorphism $\varphi : A \to B$ of R-modules. Thus to each R-module homomorphism $f : A \to M$, we need to associate an R-module homomorphism $(F(\varphi))(f) : B \to M$. Recall that in defining $\mathrm{Hom}_R(M,-)$, we simply composed $A \xrightarrow{\varphi} B$ with each morphism $M \to A$ in $\mathrm{Hom}_R(M,A)$ to get a morphism $M \to B$ in $\mathrm{Hom}_R(M,B)$. However there is no natural way to use $A \xrightarrow{\varphi} B$ to define a function from $\mathrm{Hom}_R(A,M)$ to $\mathrm{Hom}_R(B,M)$. On the other hand pre-composition with φ defines a function from $F(B) = \mathrm{Hom}_R(B,M)$ to $F(A) = \mathrm{Hom}_R(A,M)$:

$$\begin{array}{rccc} F(\varphi): & \mathrm{Hom}_R(B,M) & \longrightarrow & \mathrm{Hom}_R(A,M) \\ & f & \longmapsto & f \circ \varphi. \end{array}$$

It is also straightforward to check that $F(\varphi)$ is a homomorphism of R-modules, i.e. that $(f+g)\circ\varphi = (f\circ\varphi)+(g\circ\varphi)$ and $(rf)\circ\varphi = r(f\circ\varphi)$ for all $f,g \in \mathrm{Hom}_R(A,M)$ and

$r \in R$, so $F(\varphi) \in \mathrm{Hom}_R(F(B), F(A))$. So for each morphism $A \xrightarrow{\varphi} B$ in R-Mod, we have defined a morphism $F(A) \xleftarrow{F(\varphi)} F(B)$ in R-Mod in the opposite direction.

Recall now the notion of an "opposite" category from Example 5.22. If $\mathcal{C}$ is a category, then $\mathcal{C}^{\mathrm{op}}$ is a category with the same objects as $\mathcal{C}$, but with $\mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(A,B) = \mathrm{Hom}_{\mathcal{C}}(B,A)$. In the above example, we can therefore view $F(\varphi)$ as a morphism in

$$\mathrm{Hom}_{(R\text{-Mod})^{\mathrm{op}}}(F(A), F(B)) = \mathrm{Hom}_{R\text{-Mod}}(F(B), F(A)).$$

DEFINITION 20.5 Suppose that $\mathcal{C}$ and $\mathcal{D}$ are categories. A *contravariant functor* F from $\mathcal{C}$ to $\mathcal{D}$ is a functor[47] $F : \mathcal{C} \to \mathcal{D}^{\mathrm{op}}$; equivalently F is a rule that associates an object $F(A)$ of $\mathcal{D}$ to each object A of $\mathcal{C}$, and a function

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{C}}(A,B) & \longrightarrow & \mathrm{Hom}_{\mathcal{D}}(F(B), F(A)) \\ \varphi & \mapsto & F(\varphi) \end{array}$$

to each pair of objects A, B of $\mathcal{C}$ such that the following hold:

(i) if A is an object of $\mathcal{C}$, then $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$;

(ii) if $A \xrightarrow{\varphi} B$ and $B \xrightarrow{\psi} C$ are morphisms in $\mathcal{C}$, then

$$F(\psi \circ \varphi) = F(\varphi) \circ F(\psi).$$

EXAMPLE 20.6 Returning to the construction above, let M be an R-module, and define F by $F(A) = \mathrm{Hom}_R(A, M)$ for each R-module A, and $F(\varphi) \in \mathrm{Hom}_R(F(B), F(A))$ by $f \mapsto f \circ \varphi$ for each homomorphism of R-modules $\varphi : A \to B$. We leave it as an exercise to show that this defines a contravariant functor, denoted $\mathrm{Hom}_R(-, M)$, from the category R-Mod to itself, i.e. that conditions (i) and (ii) in the definition above are satisfied.

Notions of additivity and exactness carry over to contravariant functors.

DEFINITION 20.7 Suppose that R and S are rings, and F is a contravariant functor from R-Mod to S-Mod. We say that F is *additive* if, for all R-modules A and B, the function

$$F(\varphi) : \mathrm{Hom}_R(A,B) \longrightarrow \mathrm{Hom}_S(F(B), F(A))$$

is a homomorphism of abelian groups.

We say that an additive contravariant functor F is *left-exact* if for every exact sequence $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ of R-modules, the resulting sequence $0 \longrightarrow F(C) \xrightarrow{F(g)} F(B) \xrightarrow{F(f)} F(A)$ of S-modules is exact. Similarly we say F is *right-exact* if $F(C) \xrightarrow{F(g)} F(B) \xrightarrow{F(f)} F(A) \longrightarrow 0$ is exact whenever $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C$ is an exact sequence of R-modules, and F is *exact* if it is both left-exact and right-exact.

The analogue of Lemma 19.8 holds for contravariant additive functors. More precisely, F is exact if and only if for every exact sequence of R-modules $0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ of R-modules, the resulting sequence $0 \longrightarrow F(C) \xrightarrow{F(g)} F(B) \xrightarrow{F(f)} F(A) \longrightarrow 0$ is exact, or equivalently if $F(C) \xrightarrow{F(g)} F(B) \xrightarrow{F(f)} F(A)$ is exact for every exact sequence of R-modules $A \xrightarrow{f} B \xrightarrow{g} C$. The proof is essentially the same as that of Lemma 19.8, with again most of the work going into showing that if F preserves short exact sequences, then it preserves three-term exact sequences. This is done using the same diagram as in the proof that (ii) $\Rightarrow$ (iii) in Lemma 19.8, which now gives the exactness of

$$0 \longrightarrow F(B/\ker(g)) \xrightarrow{F(\pi)} F(B) \xrightarrow{F(i)} F(\ker(g)) \longrightarrow 0,$$

[47] What we are calling *functors* are sometimes referred to as *covariant functors*.

along with the surjectivity of $F(C) \xrightarrow{F(\widetilde{g})} F(B/\ker(g))$ and injectivity of $F(\ker(g)) \xrightarrow{F(\widetilde{f})} F(A)$.

EXAMPLE 20.8 We leave it as an exercise to show that the contravariant functor $\mathrm{Hom}_R(-,M)$ (Example 20.6) is left-exact. As with the functor $\mathrm{Hom}_R(M,-)$, it is not necessarily exact (for example, take $M = R = \mathbb{Z}$ and apply $\mathrm{Hom}_R(-,M)$ to the stock exact sequence (21)). Again it can be useful to consider R-modules M for which the functor is exact; these are called *injective* modules. We leave it as an exercise to show that $\mathbb{Q}$ is an example of an injective $\mathbb{Z}$-module.

Returning now to projective modules, recall that every free R-module is projective. We claim that for any R-module M, there is a surjective homomorphism $\varphi : F \to M$ for some free, hence projective, R-module M. Indeed let $A \subset M$ be any subset of M that generates M as an R-module (let $A = M$, for example), and let $\varphi : F_A \to M$ be the R-linear homomorphism sending $e_\alpha \to \alpha$ for $\alpha \in A$ (where F_A is the free R-module on A; see Proposition 9.13), i.e.

$$\begin{array}{rccc} \varphi : & F_A & \longrightarrow & M \\ & \displaystyle\sum_{\alpha\in A} r_\alpha e_\alpha & \longmapsto & \displaystyle\sum_{\alpha\in A} r_\alpha \alpha. \end{array}$$

Recall that this makes sense since only finitely many of the r_α are non-zero in the sum, and note that φ is surjective since its image contains A, and we assumed A generates M (as an R-module), so the only R-submodule of M containing A is M itself.

DEFINITION 20.9 Let M be an R-module. A *projective resolution* of M is an exact sequence of R-modules of the form

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0,$$

where P_i is a projective R-module for all $i \geq 0$. Similarly if each P_i is a free R-module in such an exact sequence, then it is called a *free resolution*.

Note that every free resolution is also a projective resolution. We will prove that every R-module has free (hence projective) resolutions, but first we give some examples which illustrate how such resolutions are constructed.

EXAMPLE 20.10 Let $R = \mathbb{Z}$ and let $M = \mathbb{Z}/2\mathbb{Z}$. We have the surjective homomorphism $\epsilon : \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ (of $\mathbb{Z}$-modules) defined by $\epsilon(a) = \overline{a}$ (where as usual $\overline{a}$ denotes the residue class of a for $a \in \mathbb{Z}$). The kernel of ϵ is $2\mathbb{Z}$, which is also free (being isomorphic to $\mathbb{Z}$ as a $\mathbb{Z}$-module), so we can let $P_0 = \mathbb{Z}$, $P_1 = 2\mathbb{Z}$ and $P_i = 0$ for $i \geq 2$ to obtain a free resolution

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow 2\mathbb{Z} \xrightarrow{d_1} \mathbb{Z} \xrightarrow{\epsilon} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

of $M = \mathbb{Z}/2\mathbb{Z}$, where $d_1 : 2\mathbb{Z} \to \mathbb{Z}$ is the inclusion.

EXAMPLE 20.11 Again let $M = \mathbb{Z}/2\mathbb{Z}$, but now view M as a module over $R = \mathbb{Z}/4\mathbb{Z}$. We again have a surjective homomorphism of R-modules $\epsilon : R \to M$, now defined by $a + 4\mathbb{Z} \mapsto a + 2\mathbb{Z}$, but now $\ker(\epsilon) = 2\mathbb{Z}/4\mathbb{Z}$, which is not a free R-module. In fact $\ker(\epsilon)$ is isomorphic to $M = \mathbb{Z}/2\mathbb{Z}$ via the homomorphism

$$\begin{array}{ccc} \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 2\mathbb{Z}/4\mathbb{Z} \\ a + 2\mathbb{Z} & \longmapsto & 2a + 4\mathbb{Z}. \end{array}$$

Furthermore M is not even a projective R-module; note for example that ϵ is surjective, but $\epsilon_* : \mathrm{Hom}_R(M,R) \longrightarrow \mathrm{Hom}_R(M,M)$ is not. (In fact $\epsilon_* = 0$, but $\mathrm{Hom}_R(M,M)$ contains $\mathrm{id}_M \neq 0$.)

Note however that we have a surjective homomorphism $R \to 2\mathbb{Z}/4\mathbb{Z}$ defined by $\overline{a} \mapsto \overline{2a}$, so we have the sequence of R-module homomorphisms

$$\mathbb{Z}/4\mathbb{Z} \xrightarrow{d_1} \mathbb{Z}/4\mathbb{Z} \xrightarrow{\epsilon} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

where ϵ was defined above and d_1 is given by $\overline{a} \mapsto \overline{2a}$. This is exact since ϵ is surjective and $\mathrm{im}(d_1) = 2\mathbb{Z}/4\mathbb{Z} = \ker(\epsilon)$. Furthermore note that $\ker(d_1) = 2\mathbb{Z}/4\mathbb{Z}$, so we can iterate the construction above to obtain a free resolution

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0$$

with $P_i = R = \mathbb{Z}/4\mathbb{Z}$ for all $i \geq 0$, ϵ as above, and $d_i : P_i \to P_{i-1}$ defined by $d_i(\overline{a}) \mapsto \overline{2a}$ for all $i \geq 1$.

LEMMA 20.12 *Every R-module M has a free (hence projective) resolution.*

Proof. Choose a free R-module F_0 such that[48] there is a surjective R-linear homomorphism $\epsilon : F_0 \to M$. (Recall that we can obtain such an P_0 and ϵ by letting $P_0 = F_A$ for any subset A of M that generates M as an R-module.)

Now let $M_1 = \ker(\epsilon) \subset P_0$, and choose a free R-module P_1 such that there is a surjective R-linear homomorphism $\epsilon_1 : P_1 \to M_1$. Let $d_1 : P_1 \to P_0$ be the composite $i_1 \circ \epsilon_1$, where $i_1 : M_1 \to P_0$ is the inclusion, so $\mathrm{im}(d_1) = M_1 = \ker(\epsilon)$ and the sequence

$$P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0$$

is exact.

We now iterate this process. More precisely, suppose that $n \geq 1$ and we have an exact sequence of R-module homomorphisms

$$P_n \xrightarrow{d_n} P_{n-1} \xrightarrow{d_{n-1}} \cdots \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0$$

such that $P_0, P_1, \ldots, P_n$ are free R-modules. Let $M_{n+1} = \ker(d_n) \subset P_n$ and choose a free R-module P_{n+1} and a surjective R-linear homomorphism $\epsilon_{n+1} : P_{n+1} \to M_{n+1}$. Letting $d_{n+1} : P_{n+1} \to P_n$ be the composite $i_{n+1} \circ \epsilon_{n+1}$ (where $i_{n+1} : M_{n+1} \to P_n$ is the inclusion) then yields an exact sequence of R-module homomorphisms

$$P_{n+1} \xrightarrow{d_{n+1}} P_n \xrightarrow{d_n} \cdots \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0$$

such that $P_0, P_1, \ldots, P_n, P_{n+1}$ are free. □

21 Ext: definition and examples

We have seen that if M is an R-module, then the functor $\mathrm{Hom}_R(M, -)$ does not necessarily preserve exactness of sequences. Recall that homology provides a way to measure the failure of exactness. We will make this precise in the context of such "Hom" functors, leading to the construction of certain functors denoted "Ext." (The notation is short for "extensions;" we will give an indication later of what this refers to.)

Suppose that M is an R-module. Recall from Lemma 20.12 that M has a projective resolution

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0.$$

[48] The property that such an R-module exists is to say that R-mod "has enough projectives". As the proof shows, this formally gives the existence of projective resolutions.

Let $(P_\bullet, d_\bullet)$ denote the resulting (bounded below) chain complex

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \cdots .$$

Note that, unlike the resolution itself, this sequence is not exact (unless $M = 0$); its homology is $H_i = \ker(d_i)/\operatorname{im}(d_{i+1}) = 0$ if $i \neq 0$, but

$$H_0 = P_0/\operatorname{im}(d_1) = P_0/\ker(\epsilon) \cong \operatorname{im}(\epsilon) = M.$$

Suppose now that N is another R-module, and apply the contravariant functor $F = \operatorname{Hom}_R(-, N)$ of Example 20.6 to the chain complex $P_\bullet$ obtained from a projective resolution of M. Since F is *contravariant*, this yields a sequence of homomorphisms of R-modules

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow F(P_0) \xrightarrow{F(d_1)} F(P_1) \xrightarrow{F(d_2)} F(P_2) \xrightarrow{F(d_3)} \cdots .$$

Since F is additive and $d_i \circ d_{i+1} = 0$ for all $i \geq 0$, it follows that

$$F(d_{i+1}) \circ F(d_i) = F(d_i \circ d_{i+1}) = F(0) = 0$$

for all $i \geq 0$, so the resulting sequence is a (bounded above) chain complex. For consistency with our conventions in Definition 16.9, let $C_i = F(P_{-i})$ and $\delta_i = F(d_{-i+1})$ (for $i \leq 0$). Note that since $P_i \xrightarrow{d_i} P_{i-1}$ (for $i \geq 1$), this gives $C_{-i+1} \xrightarrow{\delta_{-i+1}} C_{-i}$, where δ_{-i+1} is defined by

$$\begin{array}{ccccc} C_{-i+1} = & \operatorname{Hom}_R(P_{i-1}, N) & \longrightarrow & \operatorname{Hom}_R(P_i, N) & = C_i \\ & f & \longmapsto & f \circ d_i, & \end{array}$$

and our chain complex can be rewritten[49] as

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow C_0 \xrightarrow{\delta_0} C_{-1} \xrightarrow{\delta_{-1}} C_{-2} \xrightarrow{\delta_{-2}} \cdots ,$$

or simply $(C_\bullet, \delta_\bullet)$.

We will now use the homology of this chain complex to define Ext modules; more precisely, we let $\operatorname{Ext}^i_R(M, N) = H_{-i}((C_\bullet, \delta_\bullet))$. However the construction of the chain complex depended on a *choice* of projective resolution of M. Before we show that $\operatorname{Ext}^i_R(M, N)$ is well-defined (i.e. independent of that choice) and study its properties, let us consider some examples.

EXAMPLE 21.1 If N is any R-module and P is projective, then

$$\operatorname{Ext}^i_R(P, N) = \begin{cases} \operatorname{Hom}_R(P, N) & i = 0 \\ 0 & \text{else} \end{cases}.$$

This is because taking $P_0 = P$ and $P_i = 0$ for $i > 0$ immediately gives a resolution

$$\cdots 0 \to 0 \to 0 \to P_0 \to P$$

of P. Then $\operatorname{Ext}^i_R(P, N)$ is the homology of

$$\cdots \to 0 \to 0 \to \operatorname{Hom}_R(P_0, N) \to \operatorname{Hom}_R(0, N) \to \operatorname{Hom}_R(0, N) \to \cdots .$$

[49] An alternative to using the negative indices would be to write C^i and δ^i for what we are calling C_{-i} and δ_{-i}. The resulting sequence

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow C^0 \xrightarrow{\delta^0} C^1 \xrightarrow{\delta^1} C^2 \xrightarrow{\delta^2} \cdots ,$$

or simply $(C^\bullet, \delta^\bullet)$, is called a *cochain complex*, and its i^{th} cohomology group, denoted $H^i = H^i(C^\bullet, \delta^\bullet)$, is defined as $\ker(\delta^i)/\operatorname{im}(\delta^{i-1})$.

EXAMPLE 21.2 Let $R = \mathbb{Z}$, $M = \mathbb{Z}/2\mathbb{Z}$ and $N = \mathbb{Z}$. Recall from Example 20.10 that we have a free (hence projective) resolution

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow 2\mathbb{Z} \xrightarrow{d_1} \mathbb{Z} \xrightarrow{\epsilon} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

We may therefore take $(P_\bullet, d_\bullet)$ to be the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow 2\mathbb{Z} \xrightarrow{d_1} \mathbb{Z} \longrightarrow 0 \longrightarrow 0 \longrightarrow \cdots,$$

with $P_1 = 2\mathbb{Z}$, $P_2 = \mathbb{Z}$ and d_1 the obvious inclusion. Applying the functor $\mathrm{Hom}_{\mathbb{Z}}(-, \mathbb{Z})$ then yields the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow C_0 \xrightarrow{\delta_0} C_{-1} \longrightarrow 0 \longrightarrow 0 \longrightarrow \cdots,$$

where $C_0 = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z})$, $C_{-1} = \mathrm{Hom}_{\mathbb{Z}}(2\mathbb{Z}, \mathbb{Z})$, and $\delta_0 : C_0 \to C_{-1}$ sends a homomorphism $f : \mathbb{Z} \to \mathbb{Z}$ to its restriction $f \circ d_1 : 2\mathbb{Z} \to \mathbb{Z}$. Recall that $C_0 = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z})$ is isomorphic to $\mathbb{Z}$ (via $f \mapsto f(1)$, so $a \in \mathbb{Z}$ corresponds to the homomorphism $\mathbb{Z} \to \mathbb{Z}$ defined by $b \mapsto ab$). Since $2\mathbb{Z}$ is isomorphic to $\mathbb{Z}$ (as a $\mathbb{Z}$-module, via $2a \leftrightarrow a$), we also have that $C_{-1} = \mathrm{Hom}_{\mathbb{Z}}(2\mathbb{Z}, \mathbb{Z})$ is isomorphic to $\mathbb{Z}$ (via $g \mapsto g(2)$ for $g \in C_{-1}$). The chain complex $(C_\bullet, \delta_\bullet)$ will therefore have homology isomorphic to that of the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbb{Z} \xrightarrow{\delta_0'} \mathbb{Z} \longrightarrow 0 \longrightarrow 0 \longrightarrow \cdots,$$

where now δ_0' is the homomorphism corresponding to δ_0 under the isomorphisms $C_0 \cong \mathbb{Z}$ and $C_{-1} \cong \mathbb{Z}$, i.e. making the diagram

$$\begin{array}{ccc} C_0 & \xrightarrow{\delta_0} & C_{-1} \\ \downarrow\wr & & \downarrow\wr \\ \mathbb{Z} & \xrightarrow{\delta_0'} & \mathbb{Z} \end{array}$$

commute. Note that $1 \in \mathbb{Z}$ corresponds to $\mathrm{id}_{\mathbb{Z}} \in C_0$, and $\delta_0(\mathrm{id}_{\mathbb{Z}}) = d_1$ (the inclusion $2\mathbb{Z} \hookrightarrow \mathbb{Z}$), which corresponds to $d_1(2) = 2 \in \mathbb{Z}$, i.e. $\delta_0'(1) = 2$, so $\delta_0'(a) = 2a$ for all $a \in \mathbb{Z}$. It follows that

$$\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = H_{-i}((C_\bullet, \delta_\bullet)) \cong \begin{cases} \mathbb{Z}/2\mathbb{Z}, & \text{if } i = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Similarly one finds that if $n > 0$, then $\mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$ and $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) = 0$ for $i \neq 1$.

EXAMPLE 21.3 We claim that if M and N are any R-modules, then $\mathrm{Ext}^0_R(M, N)$ is isomorphic (as an R-module) to $\mathrm{Hom}_R(M, N)$. To see this, note that $\mathrm{Ext}^0_R(M, N)$ is the homology in degree 0 of the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathrm{Hom}_R(P_0, N) \xrightarrow{\delta_0} \mathrm{Hom}_R(P_1, N) \xrightarrow{\delta_1} \cdots,$$

so $\mathrm{Ext}^0_R(M, N) = \ker(\delta_0)$. Recall however that $\mathrm{Hom}_R(-, N)$ is left-exact (Example 20.8), so the exactness of $P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} M \longrightarrow 0$ implies that of

$$0 \longrightarrow \mathrm{Hom}_R(M, N) \xrightarrow{\alpha} \mathrm{Hom}_R(P_0, N) \xrightarrow{\delta_0} \mathrm{Hom}_R(P_1, N)$$

(where $\alpha = F(\epsilon)$ is defined by composition with $P_0 \xrightarrow{\epsilon} M$). It follows that α is injective and has image $\ker(\delta_0) = H_0((C_\bullet, \delta_\bullet)) = \mathrm{Ext}^0_R(M, N)$. Note that this is consistent with part of the conclusion of Example 21.2, since $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = 0$.

Independence of the choice of resolution

Let us now return to the problem of showing that $\operatorname{Ext}^i_R(M,N)$ is independent[50] of the choice of the projective resolution of M. More precisely suppose we choose another projective resolution

$$\cdots \xrightarrow{d'_3} P'_2 \xrightarrow{d'_2} P'_1 \xrightarrow{d'_1} P'_0 \xrightarrow{\epsilon} M \longrightarrow 0$$

and apply the functor $\operatorname{Hom}_R(-,N)$ to obtain a chain complex $(C'_\bullet, \delta'_\bullet)$. We need to show that there is an isomorphism between $H_{-i}((C_\bullet,\delta_\bullet))$ and $H_{-i}(C'_\bullet,\delta'_\bullet)$; furthermore the isomorphism is in a certain sense "canonical," allowing us to identify them and view either as $\operatorname{Ext}^i_R(M,N)$.

The first step to showing an isomorphism is to construct a map between them at all. By Lemma 17.8, we know that if we can construct a chain map $\psi_\bullet\colon (C'_\bullet,\delta'_\bullet) \to (C_\bullet,\delta_\bullet)$, then we will obtain a map on homology (which in our case will be an isomorphism). But $(C_\bullet,\delta_\bullet)$ is obtained by applying $\operatorname{Hom}_R(-,N)$ to $(P_\bullet,d_\bullet)$ (and similarly for the other complex), so as $\operatorname{Hom}_R(-,N)$ is a contravariant functor, if we have a chain map $\varphi_\bullet\colon (P_\bullet,d_\bullet) \to (P'_\bullet,d_\bullet)$, then we obtain a chain map $\psi_\bullet$ by applying $\operatorname{Hom}_R(-,N)$ (functors preserve commutative diagrams, note the arrows have reversed). We make finding $\varphi_\bullet$ our first goal!

Step 1: Chain map between projective resolutions

Consider now our two projective resolutions of M, which we will sometimes abbreviate as $P_\bullet \xrightarrow{\epsilon} M$ and $P'_\bullet \xrightarrow{\epsilon'} M$.

We will inductively show that there are R-linear homomorphisms $\varphi_i : P_i \to P'_i$ (for all $i \geq 0$) such that the following diagram commutes:

$$\begin{array}{ccccccccccc}
\cdots & \longrightarrow & P_i & \xrightarrow{d_i} & P_{i-1} & \longrightarrow \cdots \longrightarrow & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & M \\
 & & \downarrow{\scriptstyle \varphi_i} & & \downarrow{\scriptstyle \varphi_{i-1}} & & \downarrow{\scriptstyle \varphi_1} & & \downarrow{\scriptstyle \varphi_0} & & \| \\
\cdots & \longrightarrow & P'_i & \xrightarrow{d'_i} & P'_{i-1} & \longrightarrow \cdots \longrightarrow & P'_1 & \xrightarrow{d'_1} & P'_0 & \xrightarrow{\epsilon'} & M,
\end{array} \tag{23}$$

i.e. $\epsilon = \epsilon' \circ \varphi_0$ and $\varphi_{i-1} \circ d_i = d'_i \circ \varphi_i$ for all $i \geq 1$.

We only have one tool to create maps! Namely, the triangle of Proposition 20.3 (iii). We will repeatedly create such triangles. The first "triangle" is the right hand square of (23), where we consider the top and right hand maps as one side. Note that $P'_0 \overset{\epsilon'}{M}$ is surjective as required by Proposition 20.3 (iii). So we obtain a map φ_0 such that $\epsilon' \circ \varphi_0 = \epsilon$.

To create the next "triangle", consider

$$\begin{array}{ccccc}
\cdots & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & \ker(P_0 \xrightarrow{\epsilon} M) \\
 & & \downarrow{\scriptstyle \varphi_1} & & \downarrow{\scriptstyle \varphi_0|_{\ker(\epsilon)}} \\
\cdots & \xrightarrow[d'_2]{} & P'_1 & \xrightarrow[d'_1]{} & \ker(P'_0 \xrightarrow{\epsilon'} M)
\end{array}$$

[50] In other words, that $\operatorname{Ext}^i_R(M,N)$ is well-defined. As an extreme example, we don't yet know that $\operatorname{Ext}^i_R(0,N) = 0$, what if we'd used an interesting resolution of zero?

Here we use that since $P_\bullet \to M$ is a complex, the image of d_1 must be contained in the kernel of ϵ and similarly for the lower row. We also use commutativity of the left hand square of (23) (which we just showed) to know that φ_0 takes $\ker(\epsilon)$ to $\ker(\epsilon')$. Now we can use that the lower complex is exact to deduce that $\ker(\epsilon') = \mathrm{im}(d_1')$. This ensures that the lower map is surjective and by Proposition 20.3 (iii), there is an R-module homomorphism φ_1 making the "triangle" commute.

We can now iterate the previous argument to define φ_i inductively. The argument is exactly the same taking the right hand side of the square to be $\ker(P_{i-1} \to P_{i-2}) \to \ker(P'_{i-1} \to P'_{i-2})$. We obtain R-module homomorphisms φ_i making all the squares commute, i.e. a chain map $\varphi_\bullet \colon (P_\bullet, d_\bullet) \to (P'_\bullet, d'_\bullet)$ (where we define φ_i to be zero for $i < 0$).

Step 2: The chain maps ψ, ψ'

Furthermore we claim that applying the functor $F = \mathrm{Hom}_R(-, N)$ to the morphisms φ_i yields a chain map[51] $\psi_\bullet : C'_\bullet \to C_\bullet$ (where the chain complexes $C_\bullet$ and $C'_\bullet$ are the ones obtained by applying F to $P_\bullet$ and $P'_\bullet$ as in the definition of $\mathrm{Ext}^i_R(M, N)$).

More precisely, recall that C_i is defined as $F(P_{-i}) = \mathrm{Hom}_R(P_{-i}, N)$, giving a chain complex $(C_\bullet, \delta_\bullet)$ (with $\delta_i = F(d_{-i+1}) : C_i \to C_{i-1}$), and similarly we define $(C'_\bullet, \delta'_\bullet)$ by $C'_i = F(P'_{-i})$ and $\delta'_i = F(d'_{-i+1})$. Therefore letting $\psi_i = F(\varphi_{-i}) : C'_i \to C_i$, we have

$$\begin{aligned} \psi_{i-1} \circ \delta_i &= F(\varphi_{-i+1}) \circ F(d_{-i+1}) = F(d_{-i+1} \circ \varphi_{-i+1}) \\ &= F(\varphi_{-i} \circ d'_{-i+1}) = F(d'_{-i+1}) \circ F(\varphi_{-i}) = \delta'_i \circ \psi_i \end{aligned}$$

for all $i \in \mathbb{Z}$; i.e. the squares in the diagram

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & C'_{i+1} & \xrightarrow{\delta'_{i+1}} & C'_i & \xrightarrow{\delta'_i} & C'_{i-1} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle \psi_{i+1}} & & \downarrow{\scriptstyle \psi_i} & & \downarrow{\scriptstyle \psi_{i-1}} & & \\ \cdots & \longrightarrow & C_{i+1} & \xrightarrow{\delta_{i+1}} & C_i & \xrightarrow{\delta_i} & C_{i-1} & \longrightarrow & \cdots \end{array}$$

commute, so that $\psi_\bullet$ is a chain map.

Applying Lemma 17.8, we obtain R-module homomorphisms $\psi_{i,*} \colon H_i((C'_\bullet, \delta'_\bullet)) \to H_i((C_\bullet, \delta_\bullet))$ defined by

$$\begin{array}{rccc} \psi_{i,*} \colon & Z'_i/B'_i & \longrightarrow & Z_i/B_i \\ & c + B'_i & \longmapsto & \psi_i(c) + B'_i. \end{array}$$

We will prove that the $\psi_{i,*}$ are isomorphisms, and furthermore this isomorphism is independent of the choice of the homomorphisms φ_i in the construction of the chain map.

As usual, we will prove that the homomorphism we just defined is an isomorphism by constructing an inverse. To that end, note that we can interchange the roles of the projective resolutions $P_\bullet \xrightarrow{\epsilon} M$ and $P'_\bullet \xrightarrow{\epsilon'} M$ in the construction of the chain map $\varphi_\bullet$ to obtain a chain map $\varphi'_\bullet : P'_\bullet \to P_\bullet$, and hence also $\psi'_\bullet : C_\bullet \to C'_\bullet$, inducing an R-linear homomorphism

$$\psi'_{i,*} : Z_i/B_i \longrightarrow Z'_i/B'_i.$$

Interlude: Why homotopies?

We would now like to prove that the composites $\psi'_{i,*} \circ \psi_{i,*}$ and $\psi_{i,*} \circ \psi'_{i,*}$ are both identity maps. This would follow easily if we knew that the φ_i and φ'_i were each other's

[51] Note that the direction has reversed since F is contravariant

inverses. This is unfortunately unlikely to be true[52]! Rather than prove the homomorphisms $\varphi_i' \circ \varphi_i$ and $\varphi_i \circ \varphi_i'$ are identity maps, we will prove that they are "homotopic[53]" to the identity, and this will be sufficient to know that $\psi_\bullet, \psi_\bullet'$ induce inverse maps on homology.

DEFINITION 21.4 Suppose that $(A_\bullet, e_\bullet)$ and $(A'_\bullet, e'_\bullet)$ are chain complexes, and that $\eta_\bullet : A_\bullet \to A'_\bullet$ is a chain map. We say that $\eta_\bullet$ is *null-homotopic* if there is a sequence of R-linear homomorphisms $h_i : A_i \to A'_{i+1}$ such that $\eta_i = (e'_{i+1} \circ h_i) + (h_{i-1} \circ e_i)$ for all $i \in \mathbb{Z}$.

We can visualize the maps via the diagram:

$$\begin{array}{ccccccccccccc}
\cdots & \longrightarrow & A_{i+2} & \xrightarrow{e_{i+2}} & A_{i+1} & \xrightarrow{e_{i+1}} & A_i & \xrightarrow{e_i} & A_{i-1} & \xrightarrow{e_{i-1}} & \cdots \\
 & & \downarrow{\scriptstyle \eta_{i+2}} & \swarrow{\scriptstyle h_{i+2}} & \downarrow{\scriptstyle \eta_{i+1}} & \swarrow{\scriptstyle h_{i+1}} & \downarrow{\scriptstyle \eta_i} & \swarrow{\scriptstyle h_i} & \downarrow{\scriptstyle \eta_{i-1}} & \swarrow{\scriptstyle h_{i-1}} & \\
\cdots & \longrightarrow & A'_{i+2} & \xrightarrow[e'_{i+2}]{} & A'_{i+1} & \xrightarrow[e'_{i+1}]{} & A'_i & \xrightarrow[e'_i]{} & A'_{i-1} & \xrightarrow[e'_{i-1}]{} & \cdots .
\end{array}$$

Recall that $\eta_\bullet$ is a chain map (Definition 17.6) if the squares commute; it is null-homotopic if there are maps h_i along the indicated diagonals so that each η_i is the sum of the two "routes around the parallelogram" $A_i \to A'_i$ (namely $e'_{i+1} \circ h_i$ and $h_{i-1} \circ e_i$). The significance is the following:

LEMMA 21.5 *If $\eta_\bullet : A_\bullet \to A'_\bullet$ is a null-homotopic chain map, then the induced map $\eta_{i,*} : H_i(A_\bullet) \to H_i(A'_\bullet)$ is 0 for all $i \in \mathbb{Z}$.*

Proof. Recall that $\eta_{i,*}$ is the R-linear map:

$$\begin{array}{rccc}
\eta_{i,*}: & Z_i/B_i & \longrightarrow & Z'_i/B'_i \\
 & a + B_i & \longmapsto & \eta_i(a) + B'_i,
\end{array}$$

where $Z_i = \ker(e_i)$, $Z'_i = \ker(e'_i)$, $B_i = \operatorname{im}(e_{i+1})$ and $B'_i = \operatorname{im}(e'_{i+1})$. Thus for $\eta_{i,*}$ to be 0 means that $\eta_i(Z_i) \subset B'_i$.

Suppose then $a \in Z_i$, i.e. $e_i(a) = 0$. The existence of the homomorphisms h_i as in Definition 21.4 thus implies that

$$\eta_i(a) = e'_{i+1}(h_i(a)) + h_{i-1}(e_i(a)) = e'_{i+1}(h_i(a)) \in \operatorname{im}(e'_{i+1}) = Z'_i,$$

using that $a \in \ker(e_i)$, as required. □

Suppose now that $\eta_\bullet$ and $\zeta_\bullet$ are two chain maps $(A_\bullet, e_\bullet) \to (A'_\bullet, e'_\bullet)$. Since $\eta_{i-1} \circ e_i = e'_i \circ \eta_i$ and $\zeta_{i-1} \circ e_i = e'_i \circ \zeta_i$, it follows that

$$(\eta_{i-1} - \zeta_{i-1}) \circ e_i = (\eta_{i-1} \circ e_i) - (\zeta_{i-1} \circ e_i) = (e'_i \circ \eta_i) - (e'_i \circ \zeta_i) = e'_i \circ (\eta_i - \zeta_i),$$

so that the sequence $\eta_i - \zeta_i$ is also a chain map $(A_\bullet, e_\bullet) \to (A'_\bullet, e'_\bullet)$, which we denote $\eta_\bullet - \zeta_\bullet$.

DEFINITION 21.6 Suppose that $\eta_\bullet : A_\bullet \to A'_\bullet$ and $\zeta_\bullet : A_\bullet \to A'_\bullet$ are chain maps $A_\bullet \to A'_\bullet$. We say that $\eta_\bullet$ is *homotopic* to $\zeta_\bullet$ if $\eta_\bullet - \zeta_\bullet$ is null-homotopic.

[52] What if we took $(P'_\bullet, d'_\bullet)$ to be $(P_\bullet, d_\bullet)$ but with $P'_0 = P_0 \oplus R$ and $P'_1 = P_1 \oplus R$ and $d'_1 = d_1 \oplus \mathrm{id}$ using that $0 \to R \xrightarrow{\mathrm{id}} R \to 0$ is exact

[53] Homotopies come originally from topology, where two maps are homotopic when they in some strong sense nearly the same, but not necessarily equal. This will be true for homotopic chain maps also! The topological and algebraic notions can be seen to be compatible once we use chain complexes to study topology.

It is easy to see that the property of being homotopic is an equivalence relation on chain maps $A_\bullet \to A'_\bullet$. For example if $\eta_\bullet - \zeta_\bullet$ is null-homotopic, then replacing h_i by $-h_i$ in Definition 21.4 shows that so is $\zeta_\bullet - \eta_\bullet$.

COROLLARY 21.7 *Suppose that $\eta_\bullet : A_\bullet \to A'_\bullet$ and $\zeta_\bullet : A_\bullet \to A'_\bullet$ are chain maps. If $\eta_\bullet$ is homotopic to $\zeta_\bullet$, then $\eta_{i,*} = \zeta_{i,*}$ for all $i \in \mathbb{Z}$ (where $\eta_{i,*}$ and $\zeta_{i,*} : H_i(A_\bullet) \to H_i(A'_\bullet)$ are the homomorphisms induced by $\eta_\bullet$ and $\zeta_\bullet$ on homology).*

Proof. Since $\eta_\bullet - \zeta_\bullet$ is null-homotopic, the induced homomorphism

$$(\eta_i - \zeta_i)_* : H_i(A_\bullet) \to H_i(A'_\bullet)$$

is 0, but it is immediate from its definition that this is the same as $\eta_{i,*} - \zeta_{i,*}$, and hence $\eta_{i,*} = \zeta_{i,*}$ □

Step 3: Construction of the homotopies

Returning now to the task of proving that $\psi_{i,*}$ and $\psi'_{i,*}$ are each other's inverses, let us first consider the composite $\psi_{i,*} \circ \psi'_{i,*}$. Letting $\theta_i = \psi_i \circ \psi'_i$, we have that $\theta_\bullet$ is a chain map $C_\bullet \to C_\bullet$, and $\theta_{i,*} = \psi_{i,*} \circ \psi'_{i,*}$. By Corollary 21.7, it suffices to prove that $\theta_\bullet$ is homotopic to $\mathrm{id}_{C_\bullet}$ (where $\mathrm{id}_{C_\bullet}$ is the chain map defined by the identity on each C_i, which clearly induces the identity on $H_i(C_\bullet)$).

To prove that $\theta_\bullet - \mathrm{id}_{C_\bullet}$ is null-homotopic, we claim that it suffices to prove that $\zeta_\bullet - \mathrm{id}_{P_\bullet}$ is null-homotopic, where $\zeta_\bullet = \varphi'_\bullet \circ \varphi_\bullet$ (i.e. $\zeta_i = \varphi'_i \circ \varphi_i$ for all i). Recall that $(C_\bullet, \delta_\bullet)$ is defined by applying the $F = \mathrm{Hom}_R(-, N)$ to $(P_\bullet, d_\bullet)$, and similarly the chain maps $\psi_\bullet$ and $\psi'_\bullet$ are so obtained from $\varphi_\bullet$ and $\varphi'_\bullet$. So if $h_i : P_i \to P_{i+1}$ is as in Definition 21.4 with $(A_\bullet, e_\bullet) = (P_\bullet, d_\bullet)$ and $\eta_\bullet = \zeta_\bullet - \mathrm{id}_{P_\bullet}$, then

$$\begin{aligned}
\theta_i - \mathrm{id}_{C_i} &= (\psi_i \circ \psi'_i) - \mathrm{id}_{C_i} = (F(\varphi_{-i}) \circ F(\varphi'_{-i})) - F(\mathrm{id}_{P_{-i}}) \\
&= F(\varphi'_{-i} \circ \varphi_{-i}) - F(\mathrm{id}_{P_{-i}}) = F(\zeta_{-i}) - F(\mathrm{id}_{P_{-i}}) \\
&= F(\zeta_{-i} - \mathrm{id}_{P_{-i}}) = F((d_{-i+1} \circ h_{-i}) + (h_{-i-1} \circ d_{-i})) \\
&= F(d_{-i+1} \circ h_{-i}) + F(h_{-i-1} \circ d_{-i}) \\
&= (F(h_{-i}) \circ F(d_{-i+1})) + (F(d_{-i}) \circ F(h_{-i-1})) \\
&= (g_{i-1} \circ \delta_i) + (\delta_{i+1} \circ g_i),
\end{aligned}$$

where $g_i = F(h_{-i-1}) : F(P_{-i}) = C_i \to C_{i+1} = F(P_{-i-1})$, showing that $\theta_\bullet - \mathrm{id}_{C_\bullet}$ is null-homotopic. "Functors preserve equalities" (and being null-homotopic is an equality).

Finally, to prove that $\zeta_\bullet - \mathrm{id}_{P_\bullet}$ is null-homotopic (i.e. that $\zeta_\bullet$ is homotopic to $\mathrm{id}_{P_\bullet}$), we will appeal to the following lemma:

LEMMA 21.8 *Suppose that $f : A \to A'$ is a homomorphism of R-modules, and that $Q_\bullet \xrightarrow{\varepsilon} A$ and $Q'_\bullet \xrightarrow{\varepsilon'} A'$ are projective resolutions[54]. Then there is a chain map $\xi_\bullet : Q_\bullet \to Q'_\bullet$ such that $f \circ \varepsilon = \varepsilon' \circ \xi_0$, i.e. the squares in the diagram*

$$\begin{array}{ccccccccccc}
\cdots & \longrightarrow & Q_i & \xrightarrow{e_i} & Q_{i-1} & \longrightarrow \cdots \longrightarrow & Q_1 & \xrightarrow{e_1} & Q_0 & \xrightarrow{\varepsilon} & A \\
 & & \downarrow{\scriptstyle \xi_i} & & \downarrow{\scriptstyle \xi_{i-1}} & & \downarrow{\scriptstyle \xi_1} & & \downarrow{\scriptstyle \xi_0} & & \downarrow{\scriptstyle f} \\
\cdots & \longrightarrow & Q'_i & \xrightarrow{e'_i} & Q'_{i-1} & \longrightarrow \cdots \longrightarrow & Q'_1 & \xrightarrow{e'_1} & Q'_0 & \xrightarrow{\varepsilon'} & A'
\end{array} \tag{24}$$

commute. Furthermore any two such chain maps $\xi_\bullet$ and $\xi'_\bullet$ are homotopic.

[54] As can be checked in the proof, it actually suffices that $Q_\bullet$ is simply a complex of projective modules and $Q'_\bullet \to A'$ a resolution by not necessarily projectives

Proof. The proof that such a chain maps $\xi_\bullet$ exist is exactly as in the contruction of the chain map $\varphi_\bullet$ in (23); simply replace $\mathrm{id}_M : M \to M$ by $f : A \to A'$, so that $\epsilon : P_0 \to M$ is replaced by $f \circ \varepsilon : Q_0 \to A'$. We omit the details.

We still need to prove that if $\xi_\bullet$ and $\xi'_\bullet$ are two such chain maps, then $\eta_\bullet = \xi_\bullet - \xi'_\bullet$ is null-homotopic. So we need to construct a sequence of R-linear homomorphisms $h_i : Q_i \to Q'_{i+1}$ such that $\eta_i = e'_{i+1} \circ h_i = h_{i-1} \circ e_i$ for all $i \in \mathbb{Z}$. This is a slightly more advanced version of the method of Step 1. In particular, we will once again appeal to Proposition 20.3 (iii) to construct maps given triangles.

Since $Q_i = 0$ for $i < 0$, we must have $h_i = 0$ for $i < 0$ (and the desired equation is automatically satisfied for such i).

Next note that $\varepsilon' \circ \eta_0 = (\varepsilon' \circ \xi_0) - (\varepsilon' \circ \xi'_0) = (f \circ \varepsilon) - (f \circ \varepsilon) = 0$, so $\mathrm{im}(\eta_0) \subset \ker(\varepsilon')$. We then have a diagram:

$$\begin{array}{ccccc} & & & & Q_0 \\ & & & \overset{\exists\, h_0}{\swarrow} & \downarrow{\scriptstyle \eta_0} \\ \cdots & \longrightarrow & Q'_1 & \xrightarrow[e'_1]{} & \ker(\varepsilon') \end{array}$$

With h_0 existing by projectivity of Q_0 and the fact that $Q'_\bullet$ is exact so that $\ker(\varepsilon') = \mathrm{im}(e'_1)$ and the lower map is surjective. Note that since $h_{-1} = 0$, the equation $\eta_0 = (h_{-1} \circ e_0) + (e'_1 \circ h_0)$ is satisfied.

We now carry on inductively to define the homomorphisms h_i for $i > 0$. More precisely, suppose that $n > 0$ and that $h_0, h_1, \ldots, h_{n-1}$ have been defined so that

$$\eta_i = (e'_{i+1} \circ h_i) + (h_{i-1} \circ e_i) \quad \text{for } i = 0, \ldots, n-1.$$

We wish to define $h_n : Q_n \to Q'_{n+1}$ so that $\eta_n = (e'_{n+1} \circ h_n) + (h_{n-1} \circ e_n)$, or equivalently $e'_{n+1} \circ h_n = \eta_n - (h_{n-1} \circ e_n)$. Since $e'_n \circ \eta_n = \eta_{n-1} \circ e_n$ and $e'_n \circ h_{n-1} = \eta_{n-1} - (h_{n-2} \circ e_{n-1})$, we have

$$\begin{aligned} e'_n \circ (\eta_n - (h_{n-1} \circ e_n)) &= (e'_n \circ \eta_n) - ((e'_n \circ h_{n-1}) \circ e_n) \\ &= (\eta_{n-1} \circ e_n) - ((\eta_{n-1} - (h_{n-2} \circ e_{n-1})) \circ e_n) \\ &= (\eta_{n-1} \circ e_n) - (\eta_{n-1} \circ e_n) + (h_{n-2} \circ e_{n-1} \circ e_n) \\ &= 0. \end{aligned}$$

Therefore $\mathrm{im}(\eta_n - (h_{n-1} \circ e_n)) \subset \ker(e'_n)$. We then have a triangle

$$\begin{array}{ccccc} & & & & Q_n \\ & & & \overset{\exists\, h_n}{\swarrow} & \downarrow{\scriptstyle \eta_n - (h_{n-1} \circ e_n)} \\ \cdots & \longrightarrow & Q'_{n+1} & \xrightarrow[e'_{n+1}]{} & \ker(e'_n) \end{array}$$

Here h_n exists as Q_n is projective and the lower sequence is exact so that $\ker(e'_n) = \mathrm{im}(e'_{n+1})$ and the lower map of the triangle is surjective. We then have

$$(e'_{n+1} \circ h_n) = \eta_n - (h_{n-1} \circ e_n),$$

i.e.

$$\eta_n = (e'_{n+1} \circ h_n) + (h_{n-1} \circ e_n),$$

So $\eta_\bullet$ really is null-homotopic via the h_i. □

Recall that we have two choices of chain maps $P_\bullet \to P_\bullet$ such that the diagram (24) commutes when f is $\mathrm{id}_M : M \to M$. Namely, the identity map $\mathrm{id}_{P_\bullet}$ and $\zeta_\bullet := \varphi'_\bullet \circ \varphi_\bullet$.

We now apply the last part of Lemma 21.8 to see that these two choices are homotopic. It follows that $\theta_\bullet = \psi_\bullet \circ \psi'_\bullet$ is homotopic to $\mathrm{id}_{C_\bullet}$ (functors preserve equalities of compositions, additive functors preserve sums). Corollary 21.7 now implies that $\psi_{i,*} \circ \psi'_{i,*} = \theta_{i,*} = \mathrm{id}_{C_i,*}$ is the identity on $H_i(C_\bullet)$ for all $i \in \mathbb{Z}$ as desired!

Reversing the roles of $P'_\bullet \to M$ and $P_\bullet \to M$, shows that $\psi'_\bullet \circ \psi_\bullet$ is homotopic to $\mathrm{id}_{C'_\bullet}$ so that $\psi_{i,*}\colon H_i(C'_\bullet) \to H_i(C_\bullet)$ and $\psi'_{i,*}\colon H_i(C_\bullet) \to H_i(C'_\bullet)$ are mutually inverse isomorphisms as desired!

The exact same argument, using the resolution $P'_\bullet \xrightarrow{\epsilon'} M$ instead of $P_\bullet \xrightarrow{\epsilon} M$, shows that $\psi'_\bullet \circ \psi_\bullet$ is homotopic to $\mathrm{id}_{C'_\bullet}$, so $\psi'_{i,*} \circ \psi_{i,*}$ is the identity on $H_i(C'_\bullet)$.

We have now completed the proof that $\psi_{i,*} : H_i(C'_\bullet) \to H_i(C_\bullet)$ is an isomorphism for all $i \in \mathbb{Z}$. So for any two choices of projective resolution of M, the R-modules we wish to use as the definition of $\mathrm{Ext}^i_R(M,N)$ (namely $H_{-i}(C_\bullet)$ and $H_{-i}(C'_\bullet)$) are isomorphic.

Furthermore another application of Lemma 21.8 shows that any two chain maps $\varphi_\bullet$ as in (23) are homotopic. It follows, as in the discussion preceding Lemma 21.8, that so are the resulting chain maps $\psi_\bullet$, and hence the isomorphisms $\psi_{i,*}$ are independent of the choice of $\varphi_\bullet$. Identifying $H_{-i}(C'_\bullet)$ with $H_{-i}(C_\bullet)$ via this "canonical" isomorphism, we conclude that $\mathrm{Ext}^i_R(M,N)$ is well-defined[55].

22 Functoriality of Ext

Recall that if M and N are R-modules, then $\mathrm{Hom}_R(M,-)$ defines a functor from the category R-Mod to itself, and similarly $\mathrm{Hom}_R(-,N)$ defines a contravariant functor. We will now explain how the same is true with Hom replaced by Ext^i.

For each homomorphism $g : N \to N'$ of R-modules, we will associate an R-linear homomorphism $\mathrm{Ext}^i_R(M,N) \to \mathrm{Ext}^i_R(M,N')$. It's actually not hard at all to define the map. The issue is that we defined $\mathrm{Ext}^i_R(M,N)$ with respect to a choice of projective resolution of M and we need to show independence of this choice.

The map is constructed as follows. Let $P_\bullet \xrightarrow{\epsilon} M$ be a projective resolution of M. Then the homology of $(C_\bullet, \delta_\bullet)$ with $C_i = \mathrm{Hom}_R(P_{-i}, N)$ computes $\mathrm{Ext}^i_R(M,N)$ (it is its $-i^{\text{th}}$) homology whilst the homology of $(C'_\bullet, \delta'_\bullet)$ computes $\mathrm{Ext}^i_R(M,N')$. Note the boundary maps in both complexes are given by applying $\mathrm{Hom}_R(-,N), \mathrm{Hom}_R(-,N')$ to the boundary maps in the complex $P_\bullet \to M$. For each i, applying the functor $\mathrm{Hom}_R(P_i,-)$ to $g : N \to N'$ therefore yields an R-linear homomorphism $g_* : C_i \to C'_i$; more explicitly, we have maps

$$\begin{array}{ccccccc} g_*\colon & C_i & = & \mathrm{Hom}_R(P_{-i},N) & \longrightarrow & \mathrm{Hom}_R(P_{-i},N') & = & C'_i \\ & & & f & \longmapsto & g \circ f. & & \end{array}$$

[55] We have seen that the resolution defining $\mathrm{Hom}_R(P_\bullet, N)$ is not well-defined but have the much weaker statement that its homology $\mathrm{Ext}^i_R(M,N)$ is. In fact, there is a notion of equivalence on complexes that is much weaker than equality but much stronger than "having the same homology" on complexes such that the equivalence class of $\mathrm{Hom}_R(P_\bullet, N)$ is well-defined. Investigating this leads directly to the notion of "derived categories".

We claim that this sequence of homomorphisms (for $i \in \mathbb{Z}$) defines a chain map $C_\bullet \to C'_\bullet$, i.e. that the squares in the diagram

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{i+1} & \xrightarrow{\delta_{i+1}} & C_i & \xrightarrow{\delta_i} & C_{i-1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle g_*} & & \\
\cdots & \longrightarrow & C'_{i+1} & \xrightarrow{\delta'_{i+1}} & C'_i & \xrightarrow{\delta'_i} & C'_{i-1} & \longrightarrow & \cdots
\end{array}$$

commute. This is because the squares are given by:

(25)

$$\begin{array}{ccc}
\mathrm{Hom}_R(P_{-i}, N) & \xrightarrow{\delta_i} & \mathrm{Hom}_R(P_{-(i-1)}, N) \\
\downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle g_*} \\
\mathrm{Hom}_R(P_{-i}, N') & \xrightarrow{\delta'_i} & \mathrm{Hom}_R(P_{-(i-1)}, N')
\end{array}
\qquad\qquad
\begin{array}{ccc}
\psi & \xmapsto{\delta_i} & \psi \circ d_{-(i-1)} \\
\downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle g_*} \\
g \circ \psi & \xmapsto{\delta'_i} & g \circ \psi \circ d_{-(i-1)}.
\end{array}$$

We therefore obtain an induced map on homology in degree $-i$, which is the R-linear homomorphism

$$g_* : \mathrm{Ext}^i_R(M,N) = H_{-i}(C_\bullet) \longrightarrow H_{-i}(C'_\bullet) = \mathrm{Ext}^i_R(M,N'). \tag{26}$$

We now need to show that this homomorphism g_* is independent of the choice of resolution $P_\bullet \to M$. More precisely, suppose we choose a different[56] projective resolution $Q_\bullet \xrightarrow{\varepsilon} M$, and let $D_\bullet$ denote the chain complex obtained by applying $\mathrm{Hom}_R(-,N)$ to $Q_\bullet$, so that $\mathrm{Ext}^i_R(M,N)$ is identified with $H_{-i}(D_\bullet)$ via its canonical isomorphism with $H_{-i}(C_\bullet)$. Similarly $\mathrm{Ext}^i_R(M,N')$ is identified with $H_{-i}(D'_\bullet)$ via its canonical isomorphism with $H_{-i}(C'_\bullet)$, where $D'_\bullet$ is defined by applying $\mathrm{Hom}_R(-,N')$ to $Q_\bullet$, and the construction above (with $P_\bullet$ replaced by $Q_\bullet$) yields a homomorphism

$$g_* : H_{-i}(D_\bullet) \longrightarrow H_{-i}(D'_\bullet). \tag{27}$$

For g_* to be independent of the choice of resolution means that the versions defined in (26) and (27) correspond to each other under the canonical isomorphisms of §21 $H_{-i}(D_\bullet) \xrightarrow{\sim} H_{-i}(C_\bullet)$ and $H_{-i}(D'_\bullet) \xrightarrow{\sim} H_{-i}(C'_\bullet)$ that showed Ext^i was independent of the choice of resolution, i.e. the diagram

$$\begin{array}{ccc}
H_{-i}(D_\bullet) & \xrightarrow{g_*} & H_{-i}(D'_\bullet) \\
\downarrow\wr & & \downarrow\wr \\
H_{-i}(C_\bullet) & \xrightarrow{g_*} & H_{-i}(C'_\bullet)
\end{array}$$

commutes. Recall that the vertical isomorphisms are induced by the chain maps $\psi_\bullet : D_\bullet \to C_\bullet$ and $\psi'_\bullet : D'_\bullet \to C'_\bullet$ obtained by applying $\mathrm{Hom}_R(-,N)$ and $\mathrm{Hom}_R(-,N')$ to the chain map $\varphi_\bullet : P_\bullet \to Q_\bullet$ constructed in (23). The desired commutativity will therefore follow from that of

$$\begin{array}{ccccccc}
\mathrm{Hom}_R(Q_i,N) & = & D_{-i} & \xrightarrow{g_*} & D'_{-i} & = & \mathrm{Hom}_R(Q_i,N') \\
 & & \downarrow{\scriptstyle \psi_{-i}} & & \downarrow{\scriptstyle \psi'_{-i}} & & \\
\mathrm{Hom}_R(P_i,N) & = & C_{-i} & \xrightarrow{g_*} & C'_{-i} & = & \mathrm{Hom}_R(P_i,N'),
\end{array}$$

which holds since

$$\psi'_{-i}(g_*(f)) = \psi'_{-i}(f \circ g) = \varphi_i \circ f \circ g = g_*(\varphi_i \circ f) = g_*(\psi_{-i}(f)).$$

[56] We previously denoted this $P'_\bullet \xrightarrow{\epsilon'} M$, yielding a chain complex $C'_\bullet$, which has a different meaning here, so we have changed the notation.

We have now shown that $g_* : \mathrm{Ext}^i_R(M,N) \longrightarrow \mathrm{Ext}^i_R(M,N')$ is well-defined. We claim that this makes $\mathrm{Ext}^i_R(M,-)$ a functor; more precisely we have a functor $F : R\text{-Mod} \to R\text{-Mod}$ defined by $F(N) = \mathrm{Ext}^i_R(M,N)$ and $F(g) = g_*$ (as in (26)) for $g \in \mathrm{Hom}_R(N,N')$. For this, note that if $g = \mathrm{id}_N$, then the chain map $C_\bullet \to C_\bullet$ defined by g_* (in each degree) is also the identity, and hence so is g_*. Furthermore if $g \in \mathrm{Hom}_R(N,N')$ and $g' \in \mathrm{Hom}_R(N',N'')$, then the composite of the chain maps $C_\bullet \to C'_\bullet \to C''_\bullet$ inducing $F(g)$ and $F(g')$ is given (in each degree) by $f \mapsto g'_*(g_*(f)) = (g' \circ g)_*(f)$ (where $f \in C_i = \mathrm{Hom}_R(P_{-i},N)$, and $C''_i = \mathrm{Hom}_R(P_{-i},N'')$), and this is the same as the chain map $C_\bullet \to C''_\bullet$ inducing $F(g' \circ g)$.

Long exact sequence of Ext

Starting with a short exact sequence of R-modules, we will apply Lemma 18.7 to construct long exact sequences of Ext-modules.

Let M be an R-module, $P_\bullet \xrightarrow{\epsilon} M$ a projective resolution, and

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

an exact sequence of R-modules. For any $i \in \mathbb{Z}$, if we apply $\mathrm{Hom}_R(P_{-i},-)$ to this sequence, we obtain a sequence

$$0 \longrightarrow \mathrm{Hom}_R(P_{-i},N) \xrightarrow{g_*} \mathrm{Hom}_R(P_{-i},N') \xrightarrow{g'_*} \mathrm{Hom}_R(P_{-i},N'') \longrightarrow 0.$$

Since P_{-i} is projective, this sequence is exact (this is the definition of projective). The terms in this sequence are the terms of the complexes computing $(C_\bullet,\delta_\bullet),(C'_\bullet,\delta'_\bullet),(C''_\bullet,\delta''_\bullet)$ computing $\mathrm{Ext}^i_R(M,N)$, $\mathrm{Ext}^i_R(M,N')$, $\mathrm{Ext}^i_R(M,N'')$ respectively. Moreover, the maps g_*, g'_* are chain maps between these complexes (see (25)). In other words, we have a short exact sequence of chain complexes. So we can apply Lemma 18.7 to obtain a long exact sequence of Ext-groups.

COROLLARY 22.1 *Suppose that M is an R-module. If*

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

is an exact sequence of R-modules, then there is a long exact sequence:

$$\begin{array}{l}
0 \longrightarrow \mathrm{Hom}_R(M,N) \xrightarrow{g_*} \mathrm{Hom}_R(M,N') \xrightarrow{g'_*} \mathrm{Hom}_R(M,N'') \xrightarrow{\partial_0} \\
\mathrm{Ext}^1_R(M,N) \xrightarrow[g_*]{} \mathrm{Ext}^1_R(M,N') \xrightarrow{g'_*} \mathrm{Ext}^1_R(M,N'') \xrightarrow{\partial_1} \\
\mathrm{Ext}^2_R(M,N) \xrightarrow[g_*]{} \mathrm{Ext}^2_R(M,N') \xrightarrow{g'_*} \mathrm{Ext}^2_R(M,N'') \longrightarrow \\
\cdots
\end{array}$$

Proof. This all follows from our existing work. Note in particular, that $\mathrm{Ext}^0_R(M,-) = \mathrm{Hom}_R(M,-)$ by Example 21.3. □

EXAMPLE 22.2 Let $R = \mathbb{Z}$ and $M = \mathbb{Z}/2\mathbb{Z}$, and consider the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{g} \mathbb{Z} \xrightarrow{g'} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

(of $\mathbb{Z}$-modules), where $g(a) = 2a$ and $g'(a) = \overline{a}$.

Applying Corollary 22.1 gives a long exact sequence

$$\begin{array}{l} 0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}) \xrightarrow{g_*} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}) \xrightarrow{g'_*} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{\partial_0} \\ \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}) \xrightarrow[g_*]{} \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}) \xrightarrow{g'_*} \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{\partial_1} \\ \mathrm{Ext}^2_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}) \longrightarrow \cdots \end{array}$$

Recall that we computed $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z})$ in Example 21.2 to be 0 if $i \neq 1$, and $\mathrm{Ext}^1(\mathbb{Z}/2\mathbb{Z},\mathbb{Z})$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. Note also that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z})$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, so the above sequence becomes

$$\begin{array}{l} 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{\partial_0} \\ \mathbb{Z}/2\mathbb{Z} \xrightarrow[g_*]{} \mathbb{Z}/2\mathbb{Z} \xrightarrow{g'_*} \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{\partial_1} \\ 0 \longrightarrow 0 \longrightarrow \mathrm{Ext}^2_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{\partial_1} \\ 0 \longrightarrow 0 \longrightarrow \cdots \end{array}$$

The exactness implies that ∂_0 must be injective and actually an isomorphisms since every injective map $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is injective. We also find completely for free that there is an isomorphism $\mathbb{Z}/2\mathbb{Z} \xrightarrow{g'_*} \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z})$ and that all $\mathrm{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) = 0$ for $i \geq 2$.

23 Yoneda extensions (not examinable)

We are now finally ready to give the link between $\mathrm{Ext}^1_R(M,N)$ and extensions.

DEFINITION 23.1 Given a short exact sequence

$$0 \longrightarrow N \longrightarrow X \longrightarrow M \longrightarrow 0$$

is also called an *extension of M by N*. We define an equivalence relation on extensions of M by N by saying that they are equivalent whenever there is a chain map between them with outer maps the identity:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \longrightarrow & X & \longrightarrow & M & \longrightarrow & 0 \\ & & \| & & \downarrow & & \| & & \\ 0 & \longrightarrow & N & \longrightarrow & X' & \longrightarrow & M & \longrightarrow & 0 \end{array} \tag{28}$$

This is obviously transitive and reflexive. To show that it is symmetric requires knowing that the map $X \to X'$ above is an isomorphism. This follows from the snake lemma (check!). We then denote the set of equivalence classes of extensions of M by N by $\mathcal{E}xt_R(M,N)$. It is called the *Yoneda* set of extensions.

We claim that there is a canonical bijection $\mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$ from Yoneda extensions to our extensions. This means that elements of $\mathrm{Ext}^1_R(M,N)$ actually classify extensions and, conversely, this gives a route to actually compute the set $\mathcal{E}xt_R(M,N)$.

Note that if $\mathrm{Ext}^1_R(M,N) = 0$ this would then give that there is only one equivalence class of extensions. Since we can always consider the "split" extension

$$0 \longrightarrow N \longleftarrow M \oplus N \longrightarrow M \longrightarrow 0,$$

this means that whenever $\mathrm{Ext}^1_R(M,N) = 0$, all extensions of M by N are split. This is a situation we have seen many times, for example by Example 21.1 we know that $\mathrm{Ext}^1_R(P,N) = 0$ whenever P is a projective R-module. So all extensions of the form

$$0 \longrightarrow N \longrightarrow X \longrightarrow P \longrightarrow 0$$

are isomorphic to the split extension $N \oplus P$. In particular, there are no non-trivial extensions

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow X \longrightarrow \mathbb{Z} \longrightarrow 0$$

and no non-trivial extensions of vector spaces (this in particular is a comforting fact!). (Exercise, prove this directly.)

We first construct a map from $\mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$. The idea is to consider the long exact sequence of Ext corresponding to applying Corollary 22.1 with $M = M$ and the above short exact sequence. The long exact sequence we obtain begins:

$$0 \to \mathrm{Hom}_R(M,N) \to \mathrm{Hom}_R(M,X) \to \mathrm{Hom}_R(M,M) \xrightarrow{\partial_0} \mathrm{Ext}^1_R(M,N) \to \cdots. \tag{29}$$

But there is a privileged element of $\mathrm{Hom}_R(M,M)$, namely id_M. We then set

$$\begin{array}{rccc} \Phi\colon & \mathcal{E}xt_R(M,N) & \longrightarrow & \mathrm{Ext}^1_R(M,N) \\ & (0 \to M \to X \to N \to 0) & \longmapsto & \partial_0(\mathrm{id}_M). \end{array}$$

(Note the notation ∂_0 hides the fact that it depends on the choice of short exact sequence.)

We need to check that this is well-defined, i.e. independent of the choice of extension within an equivalence class. To do so, we introduce a technique that we will utilise repeatedly in this section.

Specifically, we will use that given any map of short exact sequences

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \xrightarrow{f_1} & B_1 & \xrightarrow{g_1} & C_1 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d_A} & & \downarrow{\scriptstyle d_B} & & \downarrow{\scriptstyle d_C} & & \\ 0 & \longrightarrow & A_0 & \xrightarrow{f_0} & B_0 & \xrightarrow{g_0} & C_0 & \longrightarrow & 0 \end{array} \tag{30}$$

we obtain separately from each row a short exact sequence of chain complexes

$$0 \to \mathrm{Hom}_R(P_\bullet, A_i) \to \mathrm{Hom}_R(P_\bullet, B_i) \to \mathrm{Hom}_R(P_\bullet, C_i) \to 0 \qquad \text{for } i = 0, 1, \tag{31}$$

where here $P_\bullet \to M$ is any chosen projective resolution of M. The homology each of these short exact sequences of chain complexes yields a long exact sequence of homology, which by definition computes Ext. These are the rows of the following diagram:

(32)

$$\begin{array}{ccccccccccc} 0 & \to & \mathrm{Hom}_R(M,A_1) & \to & \mathrm{Hom}_R(X,B_1) & \to & \mathrm{Hom}_R(M,C_1) & \to & \mathrm{Ext}^1_R(M,A_1) & \to & \cdots \\ & & \downarrow{\scriptstyle d_{A,*}} & & \downarrow{\scriptstyle d_{B,*}} & & \downarrow{\scriptstyle \alpha \mapsto d_C \circ \alpha} \quad {\scriptstyle d_{C,*}} & (*) & \| & & \\ 0 & \to & \mathrm{Hom}_R(M,A_0) & \to & \mathrm{Hom}_R(M,B_0) & \to & \mathrm{Hom}_R(M,C_0) & \to & \mathrm{Ext}^1_R(M,A_0) & \to & \cdots \end{array}$$

Now the vertical maps in (30) acting in the second coordinate induce chain maps between the chain complexes appearing in the short exact sequences of chain complexes (31) making a 3D commutative diagram! But we know chain maps induce maps on homology and we obtain the vertical maps in (32). Note in degree zero these are simply given by post-composition in the most obvious way. For the rest of this section, we will only need that given a map of short exact sequences as in (30), the square labelled $(*)$ commutes.

Returning to well-definedness, we apply the construction to a map of sequences (28) defining an equivalence. The square $(*)$ then becomes:

$$\begin{array}{ccc} \mathrm{Hom}_R(M,M) & \xrightarrow{\partial_X} & \mathrm{Ext}^1_R(M,N) \\ \| & & \| \\ \mathrm{Hom}_R(M,M) & \xrightarrow{\partial_{X'}} & \mathrm{Ext}^1_R(M,N) \end{array} \qquad \begin{array}{ccc} \mathrm{id}_M & \longmapsto & \partial_X(\mathrm{id}_M) \\ \| & & \| \\ \mathrm{id}_M & \longmapsto & \partial_{X'}(\mathrm{id}_M) \end{array}$$

Here, we stress that the boundary maps are for the different short exact sequences and need not be, a priori, equal. We then have

$$\Phi(0 \to N \to X \to M \to 0) := \partial_X(\mathrm{id}_M) = \partial_{X'}(\mathrm{id}_M) =: \Phi(0 \to N \to X' \to M \to 0),$$

as desired.

If the map $\Phi\colon \mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$ really was a bijection, this would give $\mathcal{E}xt_R(M,N)$ the structure of an abelian group (with identity inverses etc.). In fact, we can see directly that $\mathcal{E}xt_R(M,N)$ has a group structure.

DEFINITION 23.2 Suppose we are given two extensions

$$0 \longrightarrow N \xrightarrow{i_X} X \xrightarrow{\pi_X} M \longrightarrow 0, \tag{33}$$

$$0 \longrightarrow N \xrightarrow{i_Y} Y \xrightarrow{\pi_Y} M \longrightarrow 0. \tag{34}$$

Consider

$$Z_0 := \{(\bar{x}, \bar{y}) \mid \pi_X(x) = \pi_Y(y)\} \subseteq X \oplus Y.$$

Note that for any $n \in N$, we have that $(i_X(n), -i_Y(n)) \in Z_0$ as $\pi_X(i_X(n)) = 0 = \pi_Y(i_Y(n))$. So it makes sense to consider

$$Z := Z_0/\{(i_X(n), -i_Y(n)) \mid n \in N\}.$$

Note that this admits a map from N given by $n \mapsto [(i_X(n), 0)]$ (note $(n,0) \in Z_0$ since $\pi_X(i_X(n)) = 0$). (By our choice of quotient, this is exactly the same as $n(\mapsto [(0, i_Y(n))]$.) We also have a quotient map $Z \to M$ given by $(x,y) \mapsto \pi_X(x)$ (which is equal to $\pi_Y(y)$ by definition of Z_0), which is well-defined as $(i_X(n), -i_Y(n)) \mapsto 0$.

We claim that with these maps

$$0 \longrightarrow N \longrightarrow Z \longrightarrow M \longrightarrow 0 \tag{35}$$

is a short exact sequence. We leave this as an exercise. This will be our group multiplication and is called the *Baer sum* of the sequences (33), (34).

EXAMPLE 23.3 Consider the sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2} \mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0. \tag{36}$$

We now compute the Baer sum of this sequence with itself. We have that

$$Z_0 = \{(x,y) \in \mathbb{Z}^2 \mid x - y \text{ is even}\}.$$

It is easy to see that this is generated by $(2,0), (1,-1)$ as a submodule of $\mathbb{Z}^2$. Now in

$$Z = \{(x,y) \in \mathbb{Z}^2 \mid x - y \text{ is even}\}/\{(2x, -2y)\},$$

the denominator is generated by $(2,-2)$. We then find that

$$Z = \langle (2,0), (1,-1) \rangle / \langle (2,-2) \rangle = \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}.$$

Moreover, since the projection for the Baer sum is given by $[(x,y)] \mapsto \pi_X(x)$ and $\overline{2} \equiv 0 \pmod 2$, we find that the Baer sum is given by

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \xrightarrow{\mathrm{pr}_2} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

In other words, the Baer sum of (36) with itself is the split extension. This will make (36) its own inverse under the group law.

LEMMA 23.4 *The Baer sum takes equivalence classes to equivalence classes and gives the set of Yoneda extensions $\mathcal{E}xt_R(M,N)$ the structure of a commutative abelian group.*

Proof. Well-definedness is just a matter of checking that the construction of the extension (35) is "functorial", that is given a map of short exact sequences we get an induced map of their sums such that maps identities to identities etc. as in Definition 13.2. Since the quotients and subspaces defining the Baer sum are all functorial, so is the sum. (Apply the functoriality to the isomorphism of Definition 23.1.

It is clear that the Baer sum operation is commutative. Most of the group law conditions are formal and straightforward, if a little painful. We omit this and simply remark that the inverse of

$$0 \longrightarrow N \xrightarrow{i} X \xrightarrow{\pi} M \longrightarrow 0$$

is given by

$$0 \longrightarrow N \xrightarrow{i} X \xrightarrow{-\pi} M \longrightarrow 0.$$

(Note that Example 23.3 is a special case of this, and proving this version is only slightly harder than that example.) □

PROPOSITION 23.5 *With respect to the Baer sum. the map $\Phi\colon \mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$ is a group homomorphism.*

Proof. This is a pretty application of functoriality of the long exact sequence associated to a short exact sequence of chain complexes (Lemma 18.7) discussed above.

Now suppose we have extensions of M by N as in (33), (34). There is a map of short exact sequences:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & N\oplus N & \longrightarrow & Z_0 & \longrightarrow & M & \longrightarrow & 0 \\
 & & \| & & \downarrow & & \downarrow{\scriptstyle \Delta} & & \\
0 & \longrightarrow & N\oplus N & \xrightarrow{(i_X,i_Y)} & X\oplus Y & \xrightarrow{(\pi_X,\pi_Y)} & M\oplus M & \longrightarrow & 0
\end{array}$$

where here $\Delta\colon M \to M\oplus M$ is the diagonal map $m \mapsto (m,m)$. Label the boundaries for the upper and lower sequence by $\partial_{Z_0}, \partial_{X\oplus Y}$ respectively. So by $(*)$ above, we have a commutative square:

$$\begin{array}{ccc}
\mathrm{Hom}_R(M,M) & \xrightarrow{\partial_{Z_0}} & \mathrm{Ext}^1_R(M,N\oplus N) \\
{\scriptstyle \alpha\mapsto\Delta\circ\alpha}\downarrow & & \| \\
\mathrm{Hom}_R(M,M\oplus M) & \xrightarrow{\partial_{X\oplus Y}} & \mathrm{Ext}^1_R(M,N\oplus N)
\end{array}
\qquad\qquad
\begin{array}{ccc}
\mathrm{id}_M & \longmapsto & \partial_{Z_0}(\mathrm{id}_M) \\
\downarrow & & \| \\
\Delta & \longmapsto & \partial_{X\oplus Y}(\Delta)
\end{array}$$

The upshot being that $\partial_{Z_0}(\mathrm{id}_M) = \partial_{X\oplus Y}(\Delta)$. Now let $p\colon N\oplus N \to N$ denote the sum map (alternatively this can be considered as the quotient of $N\oplus N$ by $\{(n,-n) \mid n\in N\}$ canonically identified with N). Then we have another map of short exact sequences:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & N\oplus N & \longrightarrow & Z_0 & \longrightarrow & M & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle p} & & \downarrow & & \| & & \\
0 & \longrightarrow & N & \xrightarrow{i_Z} & Z & \xrightarrow{\pi_Z} & M & \longrightarrow & 0
\end{array}$$

This induces:

$$\begin{array}{ccc} \mathrm{Hom}_R(M,M) & \xrightarrow{\partial_{Z_0}} & \mathrm{Ext}^1_R(M,N\oplus N) \\ \| & & \downarrow{\scriptstyle p_*} \\ \mathrm{Hom}_R(M,M) & \xrightarrow{\partial_{X\oplus Y}} & \mathrm{Ext}^1_R(M,N) \end{array} \qquad \begin{array}{ccc} \mathrm{id}_M & \longmapsto & \partial_{Z_0}(\mathrm{id}_M) \\ \| & & \searrow \\ \mathrm{id}_M & \longmapsto & \partial_Z(\mathrm{id}_M) = p_*(\partial_{Z_0}(\mathrm{id}_M)) \end{array}$$

We now have

$$p_*(\partial_{X\oplus Y}(\Delta)) = p_*(\partial_{Z_0}(\mathrm{id}_M)) = \partial_Z(\mathrm{id}_M) =: \Phi(0\to N\to Z\to M\to 0).$$

We therefore want to show that

$$\begin{aligned} p_*(\partial_{X\oplus Y}(\Delta)) = \partial_X(\mathrm{id}_M) + \partial_Y(\mathrm{id}_M) =: \Phi(0\to N\to X\to M\to 0) \\ + \Phi(0\to N\to Y\to M\to 0) \end{aligned}$$

To see this, first note that $\Delta = i_1 + i_2$ as elements of $\mathrm{Hom}_R(M, M\oplus M)$. We know that both p_* and $\partial_{X\oplus Y}$ are group homomorphisms, so we need only show that $p_*(\partial_{X\oplus Y}(i_1)) = \partial_X(\mathrm{id}_M)$ (the same being true of i_2 by symmetry).

The last piece is given by the functoriality of the map:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \longrightarrow & X & \longrightarrow & M & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle i_1} & & \downarrow{\scriptstyle i_1} & & \downarrow{\scriptstyle i_1} & & \\ 0 & \longrightarrow & N\oplus N & \longrightarrow & X\oplus Y & \longrightarrow & M\oplus M & \longrightarrow & 0 \end{array}$$

giving

$$\begin{array}{ccc} \mathrm{Hom}_R(M,M) & \xrightarrow{\partial_X} & \mathrm{Ext}^1_R(M,N) \\ \downarrow{\scriptstyle i_{1,*}} & & \downarrow{\scriptstyle i_{1,*}} \\ \mathrm{Hom}_R(M,M\oplus M) & \xrightarrow{\partial_{X\oplus Y}} & \mathrm{Ext}^1_R(M,N\oplus N) \end{array} \qquad \begin{array}{ccc} \mathrm{id}_M & \longmapsto & \partial_X(\mathrm{id}_M) \\ \downarrow & & \searrow \\ i_1 & \longmapsto & \partial_{X\oplus Y}(i_1) = i_{1,*}(\partial_X(\mathrm{id}_M)) \end{array}$$

We then have $p_*(\partial_{X\oplus Y}(i_1)) = p_*(i_{1,*}(\partial_X(\mathrm{id}_M))$. But we know that $p\circ i_1 = \mathrm{id}_N$. So $p_*\circ i_{1,*} = \mathrm{id}$ and $p_*(i_{1,*}(\partial_X(\mathrm{id}_M)) = \partial_X(\mathrm{id}_M)$ as desired. □

We are nearly ready to prove that Φ is an isomorphism. We just need one additional fact:

LEMMA 23.6 *For any R-module M, there is an inclusion $M\hookrightarrow I$ of R-modules with I an injective R-module, i.e. "R-mod" has enough injectives.*

REMARK 23.7 Recall that injective R-modules are those such that $\mathrm{Hom}_R(-,I)$ is exact (Example 20.8). In this sense, they are the opposite of projective modules. This makes them satisfy the opposite of Proposition 20.3 with the argument verbatim but with arrows reversed.

We also have the analogue of Lemma 20.12 so that every R-module M admits an inclusion $M\hookrightarrow I$ embedding it within an injective module (cf. every module admits a surjection $P\twoheadrightarrow M$ from a projective). Curiously, this is not at all just given by reversing arrows in the argument of Lemma 20.12 and the existence of injective and projective resolutions are really separate properties of R-mod. (Note that free modules happen to be projective, unfortunately there is no corresponding result for injectives.) In any case, we will use this fact in the next proof, deferring its proof to the next subsection for the interested reader.

THEOREM 23.8 *The assignment* $\Phi\colon \mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$ *is an isomorphism, i.e.* $\mathrm{Ext}^1_R(M,N)$ *classifies extensions.*

Proof. Suppose that $\partial_0(\mathrm{id}_M) = 0$. Then by exactness of (29), we have that id_M lies in the image of $\mathrm{Hom}_R(M,X)$. Write $g\colon X \to M$ for the map in the original short exact sequence. This then says that $\mathrm{id}_M = g \circ s$ for some $s \in \mathrm{Hom}_R(M,X)$. In other words, there is a section

$$0 \longrightarrow N \longrightarrow X \underset{g}{\overset{s}{\rightleftarrows}} M \longrightarrow 0\ .$$

This forces $0 \to N \to X \to M \to 0$ to be isomorphic to the split extension (exercise, take $X \to N \oplus M$ to be $(\mathrm{id} - s \circ g, g)$).

For surjectivity, we must magic up a short exact sequence given an element $\xi \in \mathrm{Ext}^1_R(M,N)$. This is slightly more involved. First fix an injection $N \overset{i}{\hookrightarrow} I$ with I injective (see above remark). We can complete this into a short exact sequence

$$0 \longrightarrow N \overset{i}{\hookrightarrow} I \overset{\pi}{\longrightarrow} \mathrm{coker}\,(i) \longrightarrow 0. \tag{37}$$

For us, we will use this sequence as a kind of "universal[57]" extension starting with N. We need only cut out our desired extension from this!

Now consider the long exact sequence arising from applying Corollary 22.1 to M and the short exact sequence (37):

(38)

$$0 \to \mathrm{Hom}_R(M,N) \to \mathrm{Hom}_R(M,I) \to \mathrm{Hom}_R(M,\mathrm{coker}\,(i)) \overset{\partial_0}{\to} \mathrm{Ext}^1_R(M,N) \to \mathrm{Ext}^1_R(M,I) \to \cdots.$$

We claim that, because I is injective, we have that $\mathrm{Ext}^1_R(M,I)$ automatically vanishes. Indeed, $\mathrm{Ext}^1_R(M,I) = H_i(\mathrm{Hom}_R(P_\bullet, I))$ for a projective resolution $P_\bullet \to I$. But by definition injectives are such that $\mathrm{Hom}_R(-,I)$ is exact. Since $P_\bullet$ is exact in degrees $i \geq 1$ (see the start of §21), we have that $H_1(\mathrm{Hom}_R(P_\bullet, I)) = 0$.

Since $\mathrm{Ext}^1_R(M,I) = 0$ and (38) is exact, we find that $\mathrm{Hom}_R(M,\mathrm{coker}\,(i)) \overset{\partial_0}{\to} \mathrm{Ext}^1_R(M,N)$ is surjective. Let $\psi\colon M \to \mathrm{coker}\,(i)$ be any lift of ξ. We claim that our desired extension is given by the upper sequence of

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \xrightarrow{i_1} & X & \xrightarrow{\mathrm{pr}_2} & M & \longrightarrow & 0 \\ & & \| & & \downarrow & & \downarrow{\scriptstyle \psi} & & \\ 0 & \longrightarrow & N & \xrightarrow{i} & I & \xrightarrow{\pi} & \mathrm{coker}\,(i) & \longrightarrow & 0, \end{array} \tag{39}$$

where here X is the submodule of $I \oplus M$ defined by

$$X = \{(x,y) \mid \pi(x) = \psi(y)\} \subseteq I \oplus M.$$

The map $i_1\colon N \to X$ is then $n \mapsto (i(n), 0)$, which really does land in X as $\pi(i(n)) = 0 = \psi(0)$ and the map $\mathrm{pr}_2\colon X \to M$ is $(x,y) \mapsto y$. We should check that this really does define a short exact sequence, certainly i_1 is injective and $\mathrm{pr}_2 \circ i_1 = 0$. To see that pr_2 is surjective, fix $y \in M$ and note that since π is surjective, there exists $x \in I$ such that $\pi(x) = \psi(y)$. We then have that $(x,y) \in X$ and $\mathrm{pr}_2((x,y)) = y$. Finally, if $(x,y) \in \ker(\mathrm{pr}_2)$, then $y = 0$. That this element $(x,0)$ lies in X ensures that $\pi(x) = \psi(0) = 0$. So by exactness of the lower sequence, $x = i(n)$ for some n, as desired.

[57] Technically, using the word universal means that there should be one and only one way of cutting out every desired extension. This won't be the case and we are claiming every extension can be cut out but not necessarily uniquely. This is property is sometimes called "versal".

It remains to show that the extension defined in (39) really returns ξ under the map Φ. This is one last application of the functoriality of the Ext long exact sequence. Specifically, from the diagram 39, we obtain a map of long exact sequences:

$$\begin{array}{ccccccccc}
0 & \to & \mathrm{Hom}_R(M,N) & \to & \mathrm{Hom}_R(X,N) & \longrightarrow & \mathrm{Hom}_R(M,M) & \longrightarrow & \mathrm{Ext}^1_R(M,N) \to \cdots \\
 & & \| & & \downarrow & & {\scriptstyle \alpha\mapsto\psi\circ\alpha}\downarrow{\scriptstyle \psi_*} & & \| \\
0 & \to & \mathrm{Hom}_R(M,N) & \to & \mathrm{Hom}_R(M,I) & \to & \mathrm{Hom}_R(M,\mathrm{coker}\,(i)) & \to & \mathrm{Ext}^1_R(M,N) \to \cdots
\end{array}$$

The commutativity of final square then says that

$$\begin{array}{ccc}
\mathrm{id}_M & \longmapsto & \partial_0(\mathrm{id}_M) =: \Phi(0 \to N \to X \to M \to 0) \\
{\scriptstyle \psi_*}\downarrow & & \| \\
\psi & \longmapsto & \xi
\end{array}$$

and we win! □

There is also a notion of higher-order extensions, which are similarly classified by $\mathrm{Ext}^i_R(M,N)$.

EXAMPLE 23.9 Recall that in Example 22.2, ∂_0 was an isomorphism. In particular, $\partial_0(\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}}) \neq 0$, which is consistent with the fact the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{g} \mathbb{Z} \xrightarrow{g'} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

does not split (since, for example, $\mathbb{Z}$ is not isomorphic to $\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$).

REMARK 23.10 We have established a canonical bijection $\Phi\colon \mathcal{E}xt_R(M,N) \to \mathrm{Ext}^1_R(M,N)$. Note the left hand side is an abelian group. This means that there is a canonical group structure on $\mathcal{E}xt_R(M,N)$ also. We can in fact see this group structure explicitly on pairs of extensions. It is called the "Baer sum". We then have inverses and the identity is given by the split extension.

Enough injectives

In this subsection, we prove that any R-module M admits an injection $i\colon M \hookrightarrow I$ with I an injective R-module.

LEMMA 23.11 *For any $\mathbb{Z}$-module M (i.e. abelian group), there is an inclusion $M \hookrightarrow I$ with I an injective $\mathbb{Z}$-module, i.e. "*Ab *has enough injectives".*

Proof. Fix $m \in M$ non-zero and consider the submodule $\langle m \rangle$. This is a cyclic abelian group so either $\mathbb{Z}$ or $\mathbb{Z}/n\mathbb{Z}$ for some n. In either case, there is some map $f_m\colon \langle m \rangle \to \mathbb{Q}/\mathbb{Z}$ which is non-zero on m (it might not be injective!). We can then apply the extension property of injectives:

$$\begin{array}{ccccc}
0 & \longrightarrow & \langle m \rangle & \hookrightarrow & M \\
 & & {\scriptstyle f_m}\downarrow & \swarrow_{\exists\, \widetilde{f}_x} & \\
 & & I & &
\end{array}$$

to extend f to a map $\widetilde{f}\colon M \to \mathbb{Q}/\mathbb{Z}$ that is still non-zero on m. We now do this for all non-zero m simultaneously to obtain (by the universal property of direct products) a

map

$$\left(\prod_{m\in M\setminus\{0\}} \widetilde{f_m}\right) : M \longrightarrow \prod_{m\in M\setminus\{0\}} \mathbb{Q}/\mathbb{Z}$$

that is non-zero on all elements $m \in M \setminus \{0\}$ at once. In other words, it is an injective map.

It remains to show that $\prod_{m\in M\setminus\{0\}} \mathbb{Q}/\mathbb{Z}$ is injective, but this is a special case of the following lemma. □

LEMMA 23.12 *An arbitrary direct product of injective R-modules is injective.*

Proof. This is formal. Suppose that we have an injection $A \hookrightarrow B$ and a diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & A & \hookrightarrow & B \\ & & \downarrow{\scriptstyle g} & & \\ & & \prod_{s\in S} I_s & & \end{array}$$

with I_s injective. Then we have for every s a diagram and map h_s

$$\begin{array}{ccccc} 0 & \longrightarrow & A & \hookrightarrow & B \\ & & \downarrow{\scriptstyle \mathrm{pr}_s\circ g} & \swarrow{\scriptstyle \exists\, h_s} & \\ & & I_s & & \end{array} \tag{40}$$

So by the universal property of direct products, we obtain a map $(\prod_{s\in S} h_s) : B \to \prod_{s\in S} I_s$. We of course claim that this makes the following diagram commute:

$$\begin{array}{ccccc} 0 & \longrightarrow & A & \hookrightarrow & B \\ & & \downarrow{\scriptstyle g} & \swarrow{\scriptstyle \prod_{s\in S} h_s} & \\ & & \prod_{s\in S} I_s & & \end{array}$$

But this follows as by the universal property, there is a unique map $A \to \prod_{s\in S} I_s$ with all projections given by $\mathrm{pr}_s \circ g$, but this is true of $A \hookrightarrow B \to \prod_{s\in S} h_s$ by (40). □

Proof of Lemma 23.6. We bootstrap from Lemma 23.11. First consider M as a $\mathbb{Z}$-module (i.e. as its underlying abelian group). Then by Lemma 23.11, we know that there is an injection of $\mathbb{Z}$-modules $j : M \hookrightarrow J$ with J injective as a $\mathbb{Z}$-module. Using this j, we then have

$$M = \mathrm{Hom}_R(R, M) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, M) \xhookrightarrow{j_*} \mathrm{Hom}_{\mathbb{Z}}(R, J).$$

We can consider $\mathrm{Hom}_{\mathbb{Z}}(R, -)$ as an R-module by acting on the first factor (note R is commutative). When we do so, the map j_* is a map of R-modules. The map $\mathrm{Hom}_R(R, M) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, M)$ is also, so we need only show that $\mathrm{Hom}_{\mathbb{Z}}(R, J)$ is injective as an R-module.

This is actually a standard application of a key general result of category theory[58], which we can easily unwind here. This relies on the fact that there is a natural bijection

$$\mathrm{Hom}_R(T, \mathrm{Hom}_{\mathbb{Z}}(R, J)) \longleftrightarrow \mathrm{Hom}_{\mathbb{Z}}(T, J),$$

for any R-module T (on the right hand side, we consider T as just its underlying abelian group). This is given by sending $(t \mapsto \psi)$ to $(t \mapsto \psi(1))$ and $\varphi \in \mathrm{Hom}_{\mathbb{Z}}(T, J)$ to $(t \mapsto (r \mapsto (\varphi(rt))))$. The claim that this is "natural" for us refers to the fact that this bijection respects R-module homomorphisms $T \to T'$ (check all these claims).

[58] "functors with an exact left adjoint preserve injectives"

Now, $\mathrm{Hom}_{\mathbb{Z}}(R,J)$ is injective if and only if for all $i\colon N \hookrightarrow M$, we have that

$$\mathrm{Hom}_R(M,\mathrm{Hom}_{\mathbb{Z}}(R,J)) \xrightarrow{i^*} \mathrm{Hom}_R(N,\mathrm{Hom}_{\mathbb{Z}}(R,J)$$

is surjective (the injective version of Proposition 20.3). By the above we have a commutative diagram:

$$\begin{array}{ccc} \mathrm{Hom}_R(M,\mathrm{Hom}_{\mathbb{Z}}(R,J)) & \xrightarrow{i^*} & \mathrm{Hom}_R(N,\mathrm{Hom}_{\mathbb{Z}}(R,J)) \\ \updownarrow & & \updownarrow \\ \mathrm{Hom}_{\mathbb{Z}}(M,J) & \xrightarrow{\quad i^* \quad} & \mathrm{Hom}_{\mathbb{Z}}(N,J) \end{array}$$

(The diagram commuting by naturality of the bijection). But, by assumption J was injective as a $\mathbb{Z}$-module, so since $N \hookrightarrow M$ as a $\mathbb{Z}$-modules, we have that the lower arrow is surjective (analogue of Proposition 20.3 so that the upper arrow is also. □

Note that similarly to Lemma 20.12, having enough injectives immediately shows that injective resolutions exist also.

REMARK 23.13 From what we have seen in this course, it is tempting to think that projective resolutions are "better" than injective ones ($\mathbb{Z}$ is a much nicer module that $\mathbb{Q}/\mathbb{Z}$). Unfortunately, in several categories[59] important for geometry, injectives exist in higher generality than projectives and there is no option but to consider injective resolutions.

[59] Here we have to extend the notion of injective/projective, but this is not so hard.

Part 4

Additional Topics (not lectured)

24 Group actions on topological spaces

We will now focus on the setting where $\mathcal{C} = \mathsf{Top}$, the category of topological spaces. We first review some basic notions from topology.

DEFINITION 24.1 A *topological space* is a set X, together with a subset $\mathcal{U} \subset \{\text{subsets of } X\}$ such that

(i) $\emptyset, X \in \mathcal{U}$;

(ii) if $V, V' \in \mathcal{U}$, then $V \cap V' \in \mathcal{U}$;

(iii) if $\mathcal{A} \subset \mathcal{U}$, then $\bigcup_{V \in \mathcal{A}} V \in \mathcal{U}$.

The set $\mathcal{U}$ is called a *topology* on X, and the elements of $\mathcal{U}$ are called the *open subsets* of X. We say a subset $E \subset X$ is *closed* if its complement in X is open.

Expressed in terms of open subsets, (i) says that $\emptyset$ and X are open, (ii) says that open subsets are stable under taking *finite* intersections, and (iii) says that they are stable under taking *arbitrary* unions.

EXAMPLE 24.2 Let $X = \mathbb{R}^n$ and let $\mathcal{U}$ be the set of subsets of $\mathbb{R}^n$ which are open in the usual sense, i.e., a subset V of $\mathbb{R}^n$ is *open* if it has the property that for every $\mathbf{x} \in V$, there is $\varepsilon > 0$ such that $B(\mathbf{x}, \varepsilon) \subset V$, where

$$B(\mathbf{x}, \varepsilon) = \{\, \mathbf{y} \in \mathbb{R}^n \mid |\mathbf{x} - \mathbf{y}| < \varepsilon \,\}$$

is the open ball of radius ε around $\mathbf{x}$.

Condition (i) is obviously satisfied. Condition (ii) is satisfied since if V and V' are open and $\mathbf{x} \in V \cap V'$, then there are $\varepsilon, \varepsilon' > 0$ such that $B(\mathbf{x}, \varepsilon) \subset V$ and $B(\mathbf{x}, \varepsilon') \subset V'$, so letting $\varepsilon'' = \min(\varepsilon, \varepsilon')$, we have $B(\mathbf{x}, \varepsilon'') \subset V \cap V'$. Finally (iii) is satisfied since if $\mathcal{A} \subset \mathcal{U}$ and $\mathbf{x} \in \bigcup_{V \in \mathcal{A}} V$, then $\mathbf{x} \in V'$ for some $V' \in \mathcal{A}$; since V' is open, we have $B(\mathbf{x}, \varepsilon) \subset V' \subset \bigcup_{V \in \mathcal{A}} V$ for some $\varepsilon > 0$.

EXAMPLE 24.3 Let X be any set, and let $\mathcal{U}$ denote the set of *all* subsets of X. Then $\mathcal{U}$ clearly satisfies (i)-(iii), so it defines a topology on X, called the *discrete topology*.

On the other hand, let $\mathcal{U}' = \{\emptyset, X\}$. Then $\mathcal{U}'$ also defines a topology on X, called the *indiscrete* (or *trivial*) topology.

DEFINITION 24.4 If X is a topological space, and $x \in X$, then an open subset of X containing x is called a *neighborhood* of x (in X). We say that X is *Hausdorff* if every pair of distinct points in X have disjoint neighborhoods in X; i.e., if $x, y \in X$ and $x \neq y$, then there are open subsets $U, V \subset X$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

For example, $\mathbb{R}^n$ is Hausdorff, as is any set with the discrete topology, but X with the indiscrete topology is not Hausdorff (if $\#X > 1$).

Note that in Example 24.3, we had the same set X on which we defined two *different* topologies (at least if $\#X > 1$). These are therefore two different topological spaces, so we should really use notation like $(X, \mathcal{U})$ to distinguish it from $(X, \mathcal{U}')$ (just as we should use $(G, *)$ to denote a group). However the topology will usually be clear from the context, so we omit it from the notation (just as we usually omit explicit reference to the binary operation in the notation for a group).

EXAMPLE 24.5 Suppose that $(X, \mathcal{U})$ is a topological space and X' is a subset of X. We define the *subspace topology* on X' to be $\mathcal{U}' = \{\, V \cap X' \mid V \in \mathcal{U} \,\}$. This satisfies (i) since $\emptyset = \emptyset \cap X'$ and $X' = X \cap X'$, (ii) since $(V \cap X') \cap (V' \cap X') = (V \cap V') \cap X'$ and (iii) since

if $\mathcal{A}' \subset \mathcal{U}'$, then $\mathcal{A}' = \{ V \cap X' \mid V \in \mathcal{A} \}$ for some $\mathcal{A} \subset \mathcal{U}$, so

$$\bigcup_{V \cap X' \in \mathcal{A}'} (V \cap X') = \bigcup_{V \in \mathcal{A}} (V \cap X') = \left(\bigcup_{V \in \mathcal{A}} V \right) \cap X' \in \mathcal{U}'.$$

Therefore $(X', \mathcal{U}')$ is a topological space (but we would usually just refer to X' as a subspace of X, without explicit reference to $\mathcal{U}$ and $\mathcal{U}'$). So for example the unit circle

$$S^1 = \{ \mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x}| = 1 \} = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \}$$

is a subspace of $\mathbb{R}^2$.

EXAMPLE 24.6 Suppose that $(X, \mathcal{U})$ is a topological space, and that $\sim$ is an equivalence relation on the set X. Let X' denote the set of equivalence classes in X (often denoted $X/\sim$). Let $\pi : X \to X'$ denote the function sending x to its equivalence class (sometimes called the *quotient map*). We define the *quotient topology* on X' by

$$\mathcal{U}' = \{ V \subset X' \mid \pi^{-1}(V) \in \mathcal{U} \}.$$

This is in fact a topology on X' since (i) $\pi^{-1}(\emptyset) = \emptyset$ and $\pi^{-1}(X') = X$, so $\emptyset, X' \in \mathcal{U}'$; (ii) if $V, V' \in \mathcal{U}'$, then $\pi^{-1}(V)$ and $\pi^{-1}(V')$ are open in X, and therefore so is $\pi^{-1}(V \cap V') = \pi^{-1}(V) \cap \pi^{-1}(V')$, implying that $V \cap V' \in \mathcal{U}'$; (iii) if $\mathcal{A} \subset \mathcal{U}'$, then $\pi^{-1}(V) \in \mathcal{U}$ for every $V \in \mathcal{A}$, which implies that $\pi^{-1}(\bigcup_{V \in \mathcal{A}} V) = \bigcup_{V \in \mathcal{A}} \pi^{-1}(V)$ is also open in X, and hence $\bigcup_{V \in \mathcal{A}} V \in \mathcal{U}'$. The resulting topological space $(X', \mathcal{U}')$ is called the *quotient space*.

For example, let $X = \mathbb{R}$, and let $X' = \mathbb{R}/\mathbb{Z}$ (the equivalence relation being $x \sim y$ if $x - y \in \mathbb{Z}$). Then the open subsets of $\mathbb{R}/\mathbb{Z}$ are precisely those whose premiage in $\mathbb{R}$ is open.

EXAMPLE 24.7 Suppose that $(X_1, \mathcal{U}_1)$ and $(X_2, \mathcal{U}_2)$ are topological spaces. We define the *product topology* $\mathcal{U}$ on $X_1 \times X_2$ as follows: if $V \subset X \times X'$, then V is open if it is a union of sets of the form $V_1 \times V_2$, where $V_1 \in \mathcal{U}_1$ and $V_2 \in \mathcal{U}_2$, i.e., if

$$V = \bigcup_{(V_1, V_2) \in \mathcal{A}} (V_1 \times V_2)$$

for some $\mathcal{A} \subset \mathcal{U}_1 \times \mathcal{U}_2$. Note that the set $\mathcal{A}$ can be infinite. Then (i) and (iii) are clearly satisfied, and (ii) holds since if $V = \bigcup_{(V_1, V_2) \in \mathcal{A}} (V_1 \times V_2)$ and $V' = \bigcup_{(V_1', V_2') \in \mathcal{A}'} (V_1' \times V_2')$, then

$$V \cap V' = \bigcup_{(V_1'', V_2'') \in \mathcal{A}''} (V_1'' \times V_2''),$$

where $(V_1'', V_2'') \in \mathcal{A}''$ if $V_1'' = V_1 \cap V_1'$ and $V_2'' = V_2 \cap V_2'$ for some $(V_1, V_2) \in \mathcal{A}$ and $(V_1', V_2') \in \mathcal{A}'$.

It might have been tempting to try to define $\mathcal{U}$ to consist of subsets of $X_1 \times X_2$ of the form $V_1 \times V_2$. This would satisfy (i) and (ii), but not (iii), so it would not define a topology on $X_1 \times X_2$. On the other hand, it is not hard to see for example that if $\mathbb{R}$ is given the usual topology (as in Example 24.2 with $n = 1$), then the product topology on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ coincides with its usual topology (i.e., Example 24.2 again, with $n = 2$). The point is that $V \subset \mathbb{R}^2$ is open for the usual topology if it contains an open disk centered at each point of V, and open for the product topology if it contains an open rectangle centered at each point of V, and these two conditions are equivalent.

To define a category whose objects are topological spaces, we still need to define the morphisms (and composition).

DEFINITION 24.8 Suppose that X and Y are topological spaces. We say that a function $f : X \to Y$ is *continuous* if for every open subset V of Y, its preimage

$$f^{-1}(V) = \{ x \in X \mid f(x) \in V \}$$

is an open subset of X.

EXAMPLE 24.9 Recall the "traditional" notion[60] of continuity for functions $f : \mathbb{R}^m \to \mathbb{R}^n$: for every $\mathbf{x} \in \mathbb{R}^m$ and $\delta > 0$, there exists $\varepsilon > 0$ such that if $\mathbf{x}' \in \mathbb{R}^m$ and $|\mathbf{x}' - \mathbf{x}| < \varepsilon$, then $|f(\mathbf{x}') - f(\mathbf{x})| < \delta$, i.e.,

$$B(\mathbf{x}, \varepsilon) \subset f^{-1}(B(f(\mathbf{x}), \delta)).$$

Suppose f is continuous in the traditional sense, and V is an open subset of $\mathbb{R}^n$. If $\mathbf{x} \in f^{-1}(V)$, then $f(\mathbf{x}) \in V$, and since V is open, we have $B(f(\mathbf{x}), \delta) \subset V$ for some $\delta > 0$. So there exists $\varepsilon > 0$ such that

$$B(\mathbf{x}, \varepsilon) \subset f^{-1}(B(f(\mathbf{x}), \delta)) \subset f^{-1}(V),$$

and therefore $f^{-1}(V)$ is an open subset of $\mathbb{R}^m$. This proves that f is continuous in the sense of Definition 24.8.

Conversely suppose f is continuous in the sense of Definition 24.8, and let $\mathbf{x} \in \mathbb{R}^m$ and $\delta > 0$. Then $V = B(f(\mathbf{x}), \delta)$ is an open subset of $\mathbb{R}^n$, and therefore $f^{-1}(V)$ is an open subset of $\mathbb{R}^m$. Since $\mathbf{x} \in f^{-1}(V)$, we must have $B(\mathbf{x}, \varepsilon)$ for some $\varepsilon > 0$. Therefore f is continuous in the traditional sense.

EXAMPLE 24.10 If X is any set with the discrete topology and Y is any topological space, then every function $f : X \to Y$ is continuous. (Every subset of X is open, so $f^{-1}(V)$ will be open for every subset V of Y.)

Similary if X is any topological space and Y is any set with the indiscrete topology, then every function $f : X \to Y$ is continuous (since $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(Y) = X$ are open subsets of X).

EXAMPLE 24.11 If X is any topological space, and $X' \subset X$ is given the subspace topology, then the inclusion $i : X' \hookrightarrow X$ is continuous. (This is immediate from definition of the subspace topology.)

EXAMPLE 24.12 Suppose that X is a topological space, $\sim$ is an equivalence relation on X, and the set of equivalence classes $X' = X/\sim$ is given the quotient topology (see Example 24.6). Then the quotient map $\pi : X \to X'$ (sending x to its equivalence class) is continuous (immediately from the definition of the quotient topology).

EXAMPLE 24.13 Suppose that X, Y and Z are topological spaces and $f : X \to Y$ and $g : Y \to Z$ are continuous functions. Then $g \circ f : X \to Z$ is continuous: if V is an open subset of Z, then $g^{-1}(V)$ is an open subset of Y, and therefore $f^{-1}(g^{-1}(V))$ is an open subset of X, but note that

$$\begin{aligned} f^{-1}(g^{-1}(V)) &= \{x \in X \mid f(x) \in g^{-1}(V)\} \\ &= \{x \in X \mid g(f(x)) \in V\} = (g \circ f)^{-1}(V). \end{aligned}$$

DEFINITION 24.14 We define Top to be the category whose objects are topological spaces, morphisms are continuous functions (i.e., if X and Y are topological spaces, then $\mathrm{Hom}_{\mathsf{Top}}(X, Y)$ is the set of continuous functions $f : X \to Y$), and composition is the usual composition of functions.

We have already noted in Example 24.13 that composites of continuous functions are continuous. Since (i) obviously holds and composition of functions is associative, the only remaining thing to note is that the identity function on any topological space is continuous, so Top is indeed a category.

[60] I was going to call this the "usual" notion of continuity, but it would be better to start thinking of Definition 24.8 as the usual notion.

DEFINITION 24.15 An isomorphism in the category of topological spaces is called a *homeomorphism*. Thus if X and Y are topological spaces, then a continuous function $f : X \to Y$ is a homeomorphism if it has a continous inverse function $g : Y \to X$. In this case we say X and Y are *homeomorphic*.

EXAMPLE 24.16 Consider the complex plane $\mathbb{C}$ with its usually topology, i.e., V is open if for each $z \in V$, it contains an open disk of the form $B(z,\varepsilon) = \{ w \in \mathbb{C} \,|\, |z-w| < \varepsilon \}$ for some $\varepsilon > 0$. Then the function $f : \mathbb{R}^2 \to \mathbb{C}$ defined by $f(x,y) = x + iy$ is a homeomorphism, so $\mathbb{C}$ is homeomorphic to $\mathbb{R}^2$.

EXAMPLE 24.17 Let X be the open interval $\{ x \in \mathbb{R} \,|\, |x| < 1 \}$ (with the subspace topology as a subset of $\mathbb{R}$). The function $f : X \to \mathbb{R}$ defined by $f(x) = \tan(\pi x)$ is continuous, and so is its inverse function $g : \mathbb{R} \to X$ defined by $g(y) = \pi^{-1} \tan^{-1}(x)$. Therefore f is a homeomorphism, and X is homeomorphic to $\mathbb{R}$.

Note that if $f : X \to Y$ is a homeomorphism, then it has an inverse function, so it is bijective, but not every bijective continuous map is a homeomorphism.

EXAMPLE 24.18 Let X be any set with the discrete topology, and let Y denote the same set as X, but with the indiscrete topology (see Example 24.3). Then the identity function $f : X \to Y$ (defined by $f(x) = x$) is continuous, and obviously bijective, but its inverse function $g : Y \to X$ (also the identity function) is not continuous (assuming $\#X > 1$): Indeed suppose that V is any subset of X other than $\emptyset$ or X; then V is an open subset of X, but $g^{-1}(V) = V$ is not an open subset of Y. Therefore f is not a homeomorphism.

Suppose that $f : X \to Y$ is bijective, and let $g : Y \to X$ be its inverse map. For g to be continuous means that $g^{-1}(V)$ is open for every open subset V of X, but note that $g^{-1}(V) = f(V)$. Therefore g is continuous if and only if f has the property that it sends open sets to open sets.

DEFINITION 24.19 Suppose that $f : X \to Y$ is a function, where X and Y are topological spaces. We say that f is an *open map* (or just *open*) if for every open subset V of X, $f(V)$ is an open subset of Y.

Therefore if $f : X \to Y$ is bijective, then it is a homeomorphism if and only if it is both continuous and open, or equivalently if it has the property that V is an open subset of X if and only if $f(V)$ is an open subset of Y.

EXAMPLE 24.20 For another example of a continuous bijection whose inverse isn't continuous, let $X = [0,1) = \{ x \in \mathbb{R} \,|\, 0 \le x < 1 \}$ as a subspace of $\mathbb{R}$, and let $Y = S^1 = \{ (x,y) \in \mathbb{R}^2 \,|\, x^2 + y^2 = 1 \}$ as a subspace of $\mathbb{R}^2$. Then the function

$$\begin{array}{rccc} f: & X & \to & Y \\ & x & \mapsto & (\cos(2\pi x), \sin(2\pi x)) \end{array}$$

is continuous and bijective, but it is not open: let $V = [0, 1/2)$. Then V is an open subset of X (for example since $V = (-1/2, 1/2) \cap X$), but $f(V)$ is not open in S^1 since $\mathbf{x} = (1,0) = f(0) \in f(V)$, but there is no neighborhood of $\mathbf{x}$ in S^1 contained in $f(V)$.

Recall from Definition 5.17 that an action of a group G on a topological space X is a homomorphism $\varphi : G \to \mathrm{Aut}_{\mathsf{Top}}(X)$, where $\mathrm{Aut}_{\mathsf{Top}}(X)$ is the group of homeomorphisms $f : X \to X$ (see Definition 5.9 and Proposition 5.10). Since $\mathrm{Aut}_{\mathsf{Top}}(X)$ is a subgroup of $\mathrm{Aut}_{\mathsf{Sets}}(X) = S_X$, we have an action of G on the set X, with the further property that for every $g \in G$, the function $\varphi_g : X \to X$ is a homeomorphism (writing just φ_g for $\varphi(g)$). Note that it is sufficient that φ_g be continuous for every $g \in G$, since this implies $\varphi_g^{-1} = \varphi_{g^{-1}}$ is also continuous.

EXAMPLE 24.21 The group $\mathbb{Z}$ acts on the topological space $\mathbb{R}$ by translation: for $n \in \mathbb{Z}$, define $\varphi_n = \varphi(n) : \mathbb{R} \to \mathbb{R}$ by $\varphi_n(x) = n + x$ for all $n \in \mathbb{Z}$ and $x \in \mathbb{R}$. (Note that $\varphi_0 = \mathrm{id}_{\mathbb{R}}$, $\varphi_m \circ \varphi_n = \varphi_{m+n}$ for all $m, n \in \mathbb{Z}$, and each φ_n is continuous.) In fact $\mathbb{R}$ similarly acts on $\mathbb{R}$ by translation. Note that viewing the resulting action of $\mathbb{R}$ on $\mathbb{R}$ as a function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the function is simply the addition operation. In fact this makes $\mathbb{R}$ a *topological group* (a notion defined and discussed further in the exercises).

EXAMPLE 24.22 The group $\mathbb{R}^\times$ acts via scalar multiplication on $\mathbb{R}^n$ (as a topological space): for $r \in \mathbb{R}^\times$, define $\varphi_r : \mathbb{R}^n \to \mathbb{R}^n$ by $\varphi_r(\mathbf{x}) = r\mathbf{x}$ (so $\varphi_1 = \mathrm{id}_{\mathbb{R}^n}$, $\varphi_r \circ \varphi_s = \varphi_{rs}$ for all $r, s \in \mathbb{R}^\times$, and each φ_r is continuous).

EXAMPLE 24.23 There is an action of the group $S_2 = \{e, \sigma\}$ on $\mathbb{R}^2$ defined by $\varphi_\sigma(x, y) = (y, x)$ for $(x, y) \in \mathbb{R}^2$.

More generally define an action of S_n on $\mathbb{R}^n$ (as a topological space) by the formula

$$\varphi_\sigma(x_1, x_2 \ldots, x_n) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \ldots, x_{\sigma^{-1}(n)}).$$

Then $\varphi_e = \mathrm{id}_{\mathbb{R}^n}$, φ_σ is continuous, and $\varphi_{\sigma\tau} = \varphi_\sigma \circ \varphi_\tau$ since

$$\begin{aligned} \varphi_\sigma(\varphi_\tau(x_1, x_2 \ldots, x_n)) &= \varphi_\sigma(x_{\tau^{-1}(1)}, x_{\tau^{-1}(2)}, \ldots, x_{\tau^{-1}(n)}) \\ &= y_{\sigma^{-1}(1)}, y_{\sigma^{-1}(2)}, \ldots, y_{\sigma^{-1}(n)}), \end{aligned}$$

where $(y_1, y_2, \ldots, y_n) = (x_{\tau^{-1}(1)}, x_{\tau^{-1}(2)}, \ldots, x_{\tau^{-1}(n)})$, i.e., $y_i = x_{\tau^{-1}(i)}$, so

$$\begin{aligned} \varphi_\sigma(\varphi_\tau(x_1, x_2 \ldots, x_n)) &= (x_{\tau^{-1}(\sigma^{-1}(1))}, x_{\tau^{-1}(\sigma^{-1}(2))}, \ldots, x_{\tau^{-1}(\sigma^{-1}(n))}) \\ &= (x_{(\sigma\tau)^{-1}(1)}, x_{(\sigma\tau)^{-1}(2)}, \ldots, x_{(\sigma\tau)^{-1}(n)}) \\ &= \varphi_{\sigma\tau}(x_1, x_2 \ldots, x_n). \end{aligned}$$

EXAMPLE 24.24 Consider the action of $\mathrm{SL}_2(\mathbb{R})$ on the complex upper half-plane $X = \{x + iy \in \mathbb{C} \mid x, y \in \mathbb{R}, y > 0\}$ defined in Example **??**, and view X as a subspace of $\mathbb{C}$ with its usual topology (see Example 24.16). Since the functions $\varphi_\gamma : X \to X$ are continuous for $\gamma \in \mathrm{SL}_2(\mathbb{R})$, this is in fact defines an action of $\mathrm{SL}_2(\mathbb{R})$ on the topological space X.

Recall that if G acts on a set X, then X is partitioned into its orbits under the action (Lemma 4.14). We may thus view the orbits as equivalence classes under the equivalence relation defined by $x \sim y$ if $y = g \cdot x$ for some $g \in G$. The set of orbits is often denoted[61] $G\backslash X$, i.e.,

$$G\backslash X = \{ G \cdot x \mid x \in X \}.$$

Suppose now that X is a topological space with an action of G, let $X' = G\backslash X$ be the set of orbits, and let $\pi : X \to X'$ denote the quotient map (sending x to its orbit $G \cdot x$). Recall from Example 24.6 that we can endow X' with the quotient topology: $V \subset X'$ is open if and only if $\pi^{-1}(V)$ is open. We can thus view $X' = G\backslash X$ (read as "X mod G") as a topological space, called the *quotient space* of X by the action of G.

EXAMPLE 24.25 Consider the action of $\mathbb{Z}$ on $\mathbb{R}$ by translation (Example 24.21). The quotient space $\mathbb{Z}\backslash\mathbb{R}$ is precisely the one already considered at the end of Example 24.6 (there written as $\mathbb{R}/\mathbb{Z}$).

[61] Note that this is consistent with the usual notation for right cosets: if H is a subgroup of G, then H acts on G by left-multiplication, the orbits are the cosets Hg, and the set of such is $H\backslash G$. On the other hand this notation is clearly sometimes bad, for example we wouldn't want to write $G\backslash G$ for the set of conjugacy classes in G (i.e., orbits of G under the conjugation action by G).

We claim that in fact $\mathbb{Z}\backslash\mathbb{R}$ is homeomorphic to unit circle S^1 (viewed as a topological space as in Example 24.5). For $r \in \mathbb{R}$, let $[r]$ denote the orbit of r under the action of $\mathbb{Z}$ (so $[r] = \mathbb{Z} + r$).[62] Note that the function

$$\begin{array}{rccc} f: & \mathbb{Z}\backslash\mathbb{R} & \to & S^1 \\ & [r] & \mapsto & (\cos(2\pi r), \sin(2\pi r)) \end{array}$$

is well-defined (i.e., independent of the choice of r in $[r]$). It is continuous since if V is an open subset of S^1, then $f^{-1}(V)$ is open (since $\pi^{-1}(f^{-1}(V)) = h^{-1}(V)$ is open, where $h = f \circ \pi : \mathbb{R} \to S^1$ is defined by $x \mapsto (\cos(2\pi r), \sin(2\pi r))$). It is also clearly bijective. There are several ways to see that the inverse function $g : S^1 \to \mathbb{Z}/\mathbb{R}$ is continuous. For example, we can write S^1 as a union of open sets, say

$$\begin{aligned} U_1 &= \{(x,y) \in S^1 \mid x > 0\}, \quad U_2 = \{(x,y) \in S^1 \mid y > 0\}, \\ U_3 &= \{(x,y) \in S^1 \mid x < 0\}, \quad U_4 = \{(x,y) \in S^1 \mid y < 0\}, \end{aligned}$$

such that the restriction of g to each U_i is continous. (Letting g_i denote the restriction, this shows that if V is an open subset of $\mathbb{Z}\backslash\mathbb{R}$, then $g^{-1}(V) = \bigcup_{i=1}^4 g_i^{-1}(V)$ is open in S^1. To see for example that g_1 is continuous, write it as the composite of functions $U_1 \to \mathbb{R} \xrightarrow{\pi} \mathbb{Z}\backslash\mathbb{R}$, where the first map is $(x,y) \mapsto (\sin^{-1} y)/2\pi$.)

Recall that quotient maps $X \to X' = X/\sim$ are always continuous (Example 3.8). However they are not always open[63] (as in Definition 24.19). On the other hand if the equivalence classes are orbits under a group action, then the quotient map is in fact open:

LEMMA 24.26 *Suppose that the group G acts on the topological space X, and let $X' = G\backslash X$ be the quotient space. Then the quotient map $\pi : X \to X'$ is open.*

Proof. Let V be an open subset of X. We must prove that $\pi(V)$ is open, which (by the definition of the quotient topology) means that $\pi^{-1}(\pi(V))$ is open. But

$$x \in \pi^{-1}(\pi(V)) \iff \pi(x) \in \pi(V) \iff \pi(x) = \pi(y) \text{ for some } y \in V,$$

and $\pi(x) = \pi(y)$ if and only if $x = g \cdot y = \varphi_g(y)$ for some $g \in G$, so it follows that

$$\pi^{-1}(\pi(V)) = \bigcup_{g \in G} \varphi_g(V).$$

Since each $\varphi_g : X \to X$ is a homeomorphism, we have that each $\varphi_g(V)$ is open, and hence so is their union. □

The lemma implies for example that the function $h : \mathbb{R} \to S^1$ in Example 24.25 is an open map (being the composite of π with a homeomorphism).

In the preceding example we also saw a particular case of the following lemma, whose proof we leave as an exercise:

LEMMA 24.27 *Suppose that the group G acts on the topological space X, let $X' = G\backslash X$ be the quotient space, and $\pi : X \to X'$ the quotient map. Let Y be a topological space, and $f : X' \to Y$ a function. Then f is continuous if and only if $f \circ \pi$ is continuous.*

[62] Our usual notation for the orbit would be $\mathbb{Z} \cdot r$, but this gets confusing if the group operation is addition

[63] For example let $X = \mathbb{R}$, and define the equivalence relation so the equivalence classes are $\{\pm 1\}$ and $\{x\}$ if $x \notin \{\pm 1\}$.

EXAMPLE 24.28 Consider the example of S_2 acting on $\mathbb{R}^2$ as in Example 24.23. We'll show that the quotient space $S_2\backslash\mathbb{R}^2$ is homeomorphic to the half-plane

$$Y = \{(x,y) \in \mathbb{R}^2 \mid x \geq y\}$$

(with its topology as a subspace of $\mathbb{R}^2$). Firstly we have the function $f : Y \to S_2\backslash\mathbb{R}^2$ defined by $f(x,y) = [(x,y)]$ (again using $[(x,y)]$ to denote the orbit of (x,y)), clearly continuous since it's the composite of the inclusion $Y \to \mathbb{R}^2$ with the quotient map $\mathbb{R}^2 \to S_2\backslash\mathbb{R}^2$. On the other hand we have the function

$$\begin{array}{rccc} g: & S_2\backslash\mathbb{R}^2 & \to & Y \\ & [(x,y)] & \mapsto & (\max(x,y), \min(x,y)), \end{array}$$

easily seen to be well-defined. Note that $f \circ g$ is the identity on $S_2\backslash\mathbb{R}^2$ since $[(\max(x,y), \min(x,y))] = [(x,y)]$, and $g \circ f$ is the identity on Y since if $(x,y) \in Y$, then $x = \max(x,y)$ and $y = \min(x,y)$. Finally by the lemma, the continuity of g is equivalent to that of $\pi \circ g$, which is continuous since $(x,y) \mapsto \max(x,y)$ and $(x,y) \mapsto \min(x,y)$ are. (How does this description generalize to $S_n\backslash\mathbb{R}^n$?)

EXAMPLE 24.29 Consider the action of $\mathbb{R}^\times$ on $\mathbb{R}$ as in Example 24.22, with $n = 1$. There are just two orbits: $[0] = \{0\}$ and $[1] = \mathbb{R}^\times$. The open subsets of the quotient space $X' = \{[0],[1]\}$ are precisely $\emptyset$, $\{[1]\}$ and $\{[0],[1]\}$, but not $\{[0]\}$. Therefore X' is *not* Hausdorff: there are no disjoint neighborhoods of $[0]$ and $[1]$.

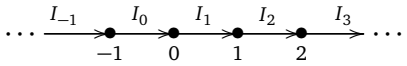
In fact if $n \geq 1$, then the quotient $\mathbb{R}^\times\backslash\mathbb{R}^n$ is not Hausdorff. Note that $\{\mathbf{0}\} = \mathbb{R}^\times \cdot \mathbf{0}$ is again an orbit. Suppose that V is a neighborhood of $[\mathbf{0}]$. Then $\pi^{-1}(V)$ is a neighborhood of $\mathbf{0}$ in $\mathbb{R}^n$, so $B(\mathbf{0},\varepsilon) \subset \pi^{-1}(V)$ for some $\varepsilon > 0$. But every point in $\mathbf{x} \in \mathbb{R}^n$ is in the same orbit as an element of $B(\mathbf{0},\varepsilon)$, since $\mathbf{x} = r(r^{-1}\mathbf{x})$ and $|r^{-1}\mathbf{x}| < \varepsilon$ if $r > \varepsilon^{-1}|\mathbf{x}|$. This proves that $\pi(B(\mathbf{0},\varepsilon)) = \mathbb{R}^\times\backslash\mathbb{R}^n$, so in fact V must be the whole quotient space, so it cannot be disjoint from a neighbourhood of any other point.

EXAMPLE 24.30 Instead of $\mathbb{R}^\times\backslash\mathbb{R}^n$ as in preceding example, let $X = \mathbb{R}^{n+1} \smallsetminus \{\mathbf{0}\}$ and consider $\mathbb{R}^\times\backslash X$. The resulting quotient space is called n-dimensional (real) projective space and usually denoted $\mathbb{P}^n(\mathbb{R})$; it is Hausdorff (and in fact a manifold), as shown in the exercises.

25 Group actions on graphs

Note that within directed graphs (Definition 17.10, both that V and E can be infinite:

EXAMPLE 25.1 Let $V = \mathbb{Z}$ and let E denote the set of closed intervals of the form $I_n = [n-1,n] = \{x \in \mathbb{R} \mid n-1 \leq x \leq n\}$ for $n \in \mathbb{Z}$. Define $\vec{t}$ by $t(I_n) = (n-1,n)$, so the resulting graph is represented by the diagram



To define a (non-directed) graph, we want to forget the order of the endpoints of each edge, so we view the "endpoint" function as taking values in $(V \times V)/\sim$, where $(u,v) \sim (v,u)$ (and $(u,v) \sim (u,v)$). Note that $(V \times V)/\sim = S_2\backslash(V \times V)$ where the action of $S_2 = \{e,\sigma\}$ on $V \times V$ is defined by $\sigma(u,v) = (v,u)$ (and $e(u,v) = (u,v)$).

DEFINITION 25.2 A *graph* Γ is a disjoint pair of sets V and E and a function $t : E \to (V \times V)/\sim$ (the *endpoint* function).

For $u, v \in V$, denote the equivalence class of (u, v) by $[u, v]$. Note that every directed graph $\vec{\Gamma} = (V, E, \vec{t})$ determines a graph $\Gamma = (V, E, t)$, where we let $t(e) = [t_0(e), t_1(e)]$, i.e., t is the composite

$$\begin{array}{ccccc} E & \xrightarrow{\vec{t}} & V \times V & \longrightarrow & (V \times V)/\sim \\ & & (u, v) & \mapsto & [u, v]. \end{array}$$

EXAMPLE 25.3 The diagram at the start of the describes the graph associated to the directed graph in Example **??**. So V and E are the same, and t is defined by

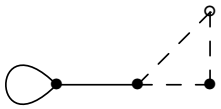
$$a \mapsto [u, u],\ b \mapsto [u, v],\ c \mapsto [v, x],\ d \mapsto [v, w],\ e \mapsto [x, w].$$

EXAMPLE 25.4 The graph associated to the directed graph in Example 25.1 is (omitting the labels) described by

A subgraph of a graph is given by subsets of vertices and edges with the same endpoint function; more precisely:

DEFINITION 25.5 If $\Gamma = (V, E, t)$ is a graph, then $\Gamma' = (V', E', t')$ is a *subgraph* of Γ if $V' \subset V$, $E' \subset E$ and $t'(e) = t(e)$ for all $e. \in E'$ (viewing $(V' \times V')/\sim$ as a subset of $(V \times V)/\sim$.

EXAMPLE 25.6 Let Γ be the graph in Example 25.3. Let $V' = \{u, v, w\}$, $E' = \{a, b\}$, and define t' by $a \mapsto [u, u]$, $b \mapsto [u, v]$. Then $\Gamma' = (V', E', t')$ is the subgraph of Γ pictured by the solid vertices and edges:



A morphism of graphs is given by functions on their vertices and edges which are compatible with endpoints; more precisely:

DEFINITION 25.7 If $\Gamma = (V, E, t)$ and $\Gamma' = (V', E', t')$ are graphs, then a *morphism* from Γ' to Γ is a pair of functions $\alpha : V' \to V$, $\beta : E' \to E$ such that $t \circ \beta = \overline{\alpha} \circ t'$, where $\overline{\alpha} : (V' \times V')/\sim \to (V \times V)/\sim$ is defined by $[v', w'] \mapsto [\alpha(v'), \alpha(w')]$. We say the morphism is *injective* if α and β are both injective.

(Note that $\overline{\alpha}$ is indeed well-defined, i.e., if $[v_1', w_1'] = [v_2', w_2']$, then $[\alpha(v_1'), \alpha(w_1')] = [\alpha(v_2'), \alpha(w_2')]$.)

EXAMPLE 25.8 If $\Gamma' = (V', E', t')$ is a subgraph of $\Gamma = (V, E, t)$, then the inclusions $V' \hookrightarrow V$ and $E' \hookrightarrow E$ define an injective morphism from Γ' to Γ.

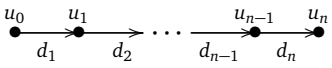
EXAMPLE 25.9 Let Γ be the graph in Example 25.4. Then there is a morphism from Γ to the graph denoted Γ' in Example 25.6 defined by

$$\alpha(n) = \begin{cases} u, & \text{if } n \equiv 0, 2 \bmod 3; \\ v, & \text{if } n \equiv 1 \bmod 3; \end{cases} \qquad \beta(I_n) = \begin{cases} a, & \text{if } n \equiv 0 \bmod 3; \\ b, & \text{if } n \equiv 1, 2 \bmod 3. \end{cases}$$

We leave it as an exercise to show that composites of morphisms are morphisms, and hence that graphs from a category, which we denote by Graphs. Therefore there are notions of isomorphisms, automorphisms, and group actions, which we return to later. As usual we will write $\Gamma \xrightarrow{\gamma} \Gamma'$ to denote an element $\gamma = (\alpha, \beta) \in \mathrm{Hom}_{\mathsf{Graphs}}(\Gamma, \Gamma')$.

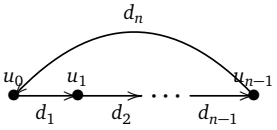
There are similar notions of subgraphs, morphisms, etc for directed graphs, but we will focus mainly on (directionless) graphs. Here are some more examples (of both):

EXAMPLE 25.10 For $n \geq 0$, define $\vec{P}_n$ to be the directed graph with $V = \{u_0, u_1, \ldots, u_n\}$, $E = \{d_1, \ldots, d_n\}$ and $t(d_i) = (u_{i-1}, u_i)$:



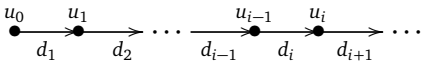
We let P_n denote the associated graph. Note that P_n is a subgraph of P_{n+1}.

EXAMPLE 25.11 For $n \geq 1$, define $\vec{C}_n$ to be the directed graph with $V = \{u_0, u_1, \ldots, u_{n-1}\}$, $E = \{d_1, \ldots, d_n\}$, $t(d_i) = (u_{i-1}, u_i)$ for $i = 1, \ldots, n-1$ and $t(d_n) = (u_{n-1}, u_0)$:



Alternatively, view the set of vertices as $\{u_{\bar{i}} \mid \bar{i} \in \mathbb{Z}/n\mathbb{Z}\}$, edges as $\{d_{\bar{i}} \mid \bar{i} \in \mathbb{Z}/n\mathbb{Z}\}$, and define t by $d_{\bar{i}} \mapsto (u_{\overline{i-1}}, u_{\bar{i}})$. Again let C_n denote the associated graph. Note that there is a morphism $P_n \to C_n$ defined by $\alpha(u_i) = u_{\bar{i}}$ and $\beta(d_i) = d_{\bar{i}}$.

EXAMPLE 25.12 We may define infinite versions of the graphs in Examples 25.10 and 25.11. Let $\vec{P}_\infty$ denote the directed graph with $V = \{u_i \mid i \in \mathbb{Z}, i \geq 0\}$, $E = \{d_i \mid i \in \mathbb{Z}, i \geq 1\}$, and $t(d_i) = (u_{i-1}, u_i)$:



Similarly let $\vec{C}_\infty$ denote the directed graph with $V = \{u_i \mid i \in \mathbb{Z}\}$, $E = \{d_i \mid i \in \mathbb{Z}\}$, and $t(d_i) = (u_{i-1}, u_i)$. Again let P_∞ and C_∞ denote the associated graphs. Note that P_∞ is a subgraph of C_∞, and C_∞ is isomorphic to the graph in Example 25.4 via $u_i \mapsto i$ and $d_i \mapsto I_i$.

DEFINITION 25.13 If Γ is a graph, then a *path* of *length* n in Γ is a morphism $P_n \xrightarrow{\gamma} \Gamma$.

Note that to give a path $\gamma = (\alpha, \beta)$ of length n in $\Gamma = (V, E, t)$ is equivalent to giving the sequences of vertices $(v_0, \ldots, v_n)$ and edges $(e_1, \ldots, e_n)$, where $v_i = \alpha(u_i) \in V$ and $e_i = \beta(d_i) \in E$; for α and β to define a morphism $P_n \to \Gamma$, we need that $t(e_i) = (v_{i-1}, v_i)$ for $i = 1, \ldots, n$. We call γ a path from v_0 to v_n.

EXAMPLE 25.14 Let Γ be the graph in Example 25.3. Then there is a path of length 2 from u to w in Γ, given by the sequences of vertices (u, v, w) and edges (b, d). There is also a path of length 3 from u to w given by the sequences (u, v, x, w) and (b, c, e), and another path of length 3 given by (u, u, v, w) and (a, b, d). The following sequences define paths of length 4 from u to w:

- (u, v, u, v, w) and (b, b, b, d);
- (u, v, w, v, w) and (b, d, d, d);
- (u, u, v, x, w) and (a, b, c, e).

EXAMPLE 25.15 The morphism $P_n \to C_n$ defined in Example 25.11 is a path of length n from $u_{\bar{0}}$ to $u_{\bar{0}}$.

Let $P_n \xrightarrow{\gamma} \Gamma$ be a path, with associated vertices $(v_0, \ldots, v_n)$ and edges $(e_1, \ldots, e_n)$. Note that γ is injective if and only if the sequence of vertices $v_0, v_1, \ldots, v_n$ are distinct (as this implies the edges $e_1, \ldots, e_n$ are also distinct, and hence the functions α and β are injective). So in Example 25.14, the first two paths are injective, and the rest are not.

DEFINITION 25.16 We say a path (with associated vertices $(v_0, \dots, v_n)$ and edges $(e_1, \dots, e_n)$) is *reduced* if $e_i \neq e_{i+1}$ for $i = 1, \dots, n-1$.

Note that an injective path is necessarily reduced, but a reduced path need not be injective (for example, the path in Example 25.15). Which of the paths in Example 25.14 are reduced?

DEFINITION 25.17 We say that a graph $\Gamma = (V, E, t)$ is *connected* if for every pair of vertices $v, w \in V$, there is a path from v to w (or equivalently, an injective path from v to w).

For example, the graph Γ in Example 25.3 is connected, but its subgraph Γ' in Example 25.6 is not. The graphs P_n, C_n, P_∞ and C_∞ are all connected.

DEFINITION 25.18 If Γ is a graph, then a *circuit* of *length* n in Γ is an injective morphism $C_n \xrightarrow{\gamma} \Gamma$.

Note that to give a circuit of length n in $\Gamma = (V, E, t)$ is equivalent to giving a sequence of vertices $(v_0, \dots, v_n)$ and edges $(e_1, \dots, e_n)$ (with each $v_i \in V$, $e_i \in E$) such that

- $v_0, v_1, \dots, v_{n-1}$ are distinct. and $v_n = v_0$;
- $e_1, e_2, \dots, e_n$ are distinct;
- $t(e_i) = [v_{i-1}, v_i]$ for $i = 1, \dots, n$.

EXAMPLE 25.19 Let Γ be the graph in Example 25.3. There is a circuit of length 3 in Γ with associated vertices (x, v, w, x) and edges (c, d, e). There is also a circuit of length 1 with vertices (u, u) and edge (a).

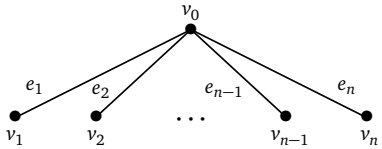
EXAMPLE 25.20 Suppose that $\Gamma = (V, E, t)$ is any graph, $e \in E$ is any edge, and $t(e) = [v, w]$. The morphism $C_2 \to \Gamma$ with associated vertices (v, w, v) and edges (e, e) is *not* a circuit; even though the vertices v, w may be distinct, the edges e, e are not.

DEFINITION 25.21 A graph is a *tree* if it is non-empty, connected and has no circuits.

EXAMPLE 25.22 The graphs P_n (for any $n \geq 0$), P_∞ and C_∞ are trees, but the graph C_n (for any $n \geq 1$) is not (since the identity is a circuit of length n).

EXAMPLE 25.23 The graph Γ in Example 25.3 is not a tree since it has circuits (described in Example 25.19.

EXAMPLE 25.24 The graph $\Gamma = (V, E, t)$ with $V = \{v_0, v_1, \dots, v_n\}$, $E = \{e_1, \dots, e_n\}$ and $t(e_i) = [v_0, v_i]$ for $i = 1, \dots, n$ is a tree:



Let us now return to the notion of group actions on trees. Recall that an action of a group G on a tree $\Gamma = (V, E, t)$ is a homomorphism $G \to \mathrm{Aut}_{\mathsf{Graphs}}(\Gamma)$. This means that there are morphisms $\Gamma \xrightarrow{\varphi_g} \Gamma$ for $g \in G$ such that[64] $\varphi_{e_G} = \mathrm{id}_\Gamma$ and $\varphi_{gh} = \varphi_g \circ \varphi_h$ for all $g, h \in G$. Recall that each φ_g is a pair of functions $\alpha_g : V \to V$, $\beta_g : E \to E$, compatible with endpoints in the sense that $t \circ \beta_g = \overline{\alpha}_g \circ t$ (where $\overline{\alpha}_g$ is as in Definition 25.7). For the φ_g to define an action therefore means that $\alpha_{e_G} = \mathrm{id}_V$, $\beta_{e_G} = \mathrm{id}_E$, and $\alpha_{gh} = \alpha_g \circ \alpha_h$ and

[64] writing e_G for the identity in G to distinguish it from edges $e \in E$

$\beta_{gh} = \beta_g \circ \beta_h$ for all $g, h \in G$. In other words, to give an action of G on Γ is equivalent to giving actions of G on V and on E which preserve endpoints in the sense if $g \in G$, $e \in E$ and $t(e) = [v, w]$, then $t(g \cdot e) = [g \cdot v, g \cdot w]$.

EXAMPLE 25.25 The group S_n acts on the graph Γ in Example 25.24 via $\sigma \cdot v_0 = v_0$, $\sigma \cdot v_i = v_{\sigma(i)}$ and $\sigma \cdot e_i = e_{\sigma(i)}$ for $\sigma \in S_n$ and $i = 1, \ldots, n$. Recall that t is defined by $e_i \mapsto (v_0, v_i)$, so

$$t(\sigma \cdot e_i) = t(e_{\sigma(i)}) = [v_0, v_{\sigma(i)}] = [\sigma \cdot v_0, \sigma \cdot v_i]$$

as required.

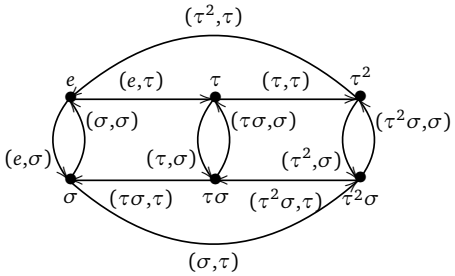
EXAMPLE 25.26 The group $\mathbb{Z}/n\mathbb{Z}$ acts on the graph C_n via $[a] \cdot u_{[i]} = a_{[a+i]}$ and $[a] \cdot d_{[i]} = d_{[a+i]}$ for $[a], [i] \in \mathbb{Z}/n\mathbb{Z}$. We leave it as an exercise to show that in fact D_n acts on C_n.

EXAMPLE 25.27 Similarly $\mathbb{Z}$ acts on C_∞ via $a \cdot u_i = u_{a+i}$ and $a \cdot d_i = d_{a+i}$ for $a, i \in \mathbb{Z}$.

Given a group, we will now define certain graphs on which the group acts.

DEFINITION 25.28 If G is a group and S is a subset of G, then the *Cayley graph of G relative to S* is the graph $\Gamma(G, S)$ with vertices $V = G$, edges $E = G \times S$, and endpoints defined by $t(g, s) = [g, gs]$ for all $g \in G$, $s \in S$. Similary the *directed Cayley graph (of G relative to S)* is the directed graph $\vec{\Gamma}(G, S)$ defined by $V = G$, $E = G \times S$ and $t(g, s) = (g, gs)$.

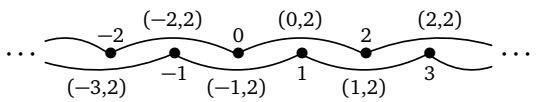
EXAMPLE 25.29 Let $G = S_3$ and $S = \{\sigma\, \tau\}$, where $\sigma = (12)$ and $\tau = (123)$, so $G = \{e, \tau, \tau^2, \sigma, \tau\sigma, \tau^2\sigma\}$. The 6 vertices of $\Gamma(G, S)$ are labelled by elements $g \in G$, and the 12 edges by pairs $(g, s) \in G \times S$. Since $\sigma\tau = \tau^2\sigma$, the directed Cayley graph $\vec{\Gamma}(G, S)$ is



The Cayley graph $\Gamma(G, S)$ is then the associated (directionless) graph. Note that since $\sigma^2 = e$, there are two edges with endpoints $[g, g\sigma]$ for each $g \in G$, namely (g, σ) and $(g\sigma, \sigma)$.

EXAMPLE 25.30 Let $G = \mathbb{Z}/n\mathbb{Z}$ and let $S = \{\overline{1}\}$. Then there are n vertices, $\overline{i}$, and n edges, $(\overline{i}, \overline{1})$, where $i \in \mathbb{Z}/n\mathbb{Z}$. Since the endpoints of $(\overline{i-1}, \overline{1})$ are $\overline{i-1}$ and $\overline{i}$, there is an isomorphism $C_n \to \Gamma(G, S)$ defined by $u_{\overline{i}} \mapsto \overline{i}$ and $d_{\overline{i}} \mapsto (\overline{i-1}, \overline{1})$ (with notation as in Example 25.11).

EXAMPLE 25.31 Similarly to the preceding example, $\Gamma(\mathbb{Z}, \{1\})$ is isomorphic to C_∞. On the other hand $\Gamma(\mathbb{Z}, \{2\})$ is



Note that the Cayley graphs in Examples 25.29 and 25.30 were connected, but $\Gamma(\mathbb{Z}, \{2\})$ in Example 25.31 was not. These are examples of the following, whose proof we leave as an exercise:

PROPOSITION 25.32 *Let G be a group and $S \subset G$. Then the Cayley graph $\Gamma(G,S)$ is connected if and only if $G = \langle S \rangle$.*

For any subset S of a group G, there is an action of G on $\Gamma(G,S)$ defined by:

- $g \cdot h = gh$ for $g \in G$, $h \in V = G$;
- $g \cdot (h,s) = (gh,s)$ for $g \in G$, $(h,s) \in E = G \times S$.

We leave it as an exercise to show that this does indeed define an action of G on $\Gamma(G,S)$. Furthermore the action is free, in a sense[65] to be defined below. First we define what it means for a group action on a set to be free.

DEFINITION 25.33 Suppose that a group G acts on a set X. We say that the action is *free* (or that G acts *freely* on X), if $G_x = \{e_G\}$ for all $x \in X$ (where G_x is the stabiliser of x).

For example, the action of G on G by left multiplication (Example 4.9) is free. On the other hand, the action of G on itself by conjugation (Example 4.11) is not free (unless $G = \{e_G\}$) since $G_{e_G} = G$.

DEFINITION 25.34 Suppose that a group G acts on a graph $\Gamma = (V,E,t)$. We say that the action is *free* (or that G acts *freely* on Γ) if the resulting actions on V and E are both free.

EXAMPLE 25.35 Consider the action of $S_2 = \{e,\sigma\}$ on P_2 defined by $\sigma(u_0) = u_1$, $\sigma(u_0) = u_1$ and $\sigma(d_1) = d_1$. Then S_2 acts freely on V but not on E, so the action is not free. Similarly an action can fail to be free because the action on vertices isn't, even if the action on edges; take for example the action of S_2 on C_3 defined by:

$$\sigma(u_0) = u_2,\ \sigma(u_1) = u_1,\ \sigma(u_2) = u_0,\ \sigma(d_1) = d_2,\ \sigma(d_2) = d_1.$$

EXAMPLE 25.36 The action of G on its Cayley graph $\Gamma(G,S)$ (for any $S \subset G$) is free, since if $g \cdot h = h$ then $g = e_G$, and if $g \cdot (h,s) = (h,s)$, then $gh = h$, so again $g = e_G$.

One application in group theory of actions on graphs stems from the following fact. Recall if S is a set, then F_S denotes the free group on S, and we view S as a subset of F_S by identifying $s \in S$ with the reduced word $(s^+) \in F_S$.

PROPOSITION 25.37 *If S is any set, then the Cayley graph $\Gamma(F_S,S)$ is a tree.*

Proof. Firstly $\Gamma(F_S,S)$ is non-empty since $e_{F_S} \in V = F_S$.

Secondly, since F_S is generated by S (for example by Lemma 6.13 with $f = \mathrm{id}_S$), it follows from Proposition 25.32 that $\Gamma(F_S,S)$ is connected.

It remains to prove that $\Gamma(F_S,S)$ has no circuits. Suppose then that $C_n \xrightarrow{\gamma} \Gamma(F_S,S)$ is a circuit of length n. Let $(g_0,g_1,\dots,g_n)$ and $(e_1,\dots,e_n)$ be the corresponding sequences of vertices and edges, so $g_0,\dots,g_{n-1}$ are distinct elements of F_S, $g_n = g_0$, and the e_i (for $i = 1,\dots,n$) are distinct edges $(h_i,s_i) \in F_S \times S$ such that

$$t(h_i,s_i) = [h_i, h_i s_i] = [g_{i-1}, g_i].$$

The last condition means that either $g_{i-1} = h_i$ and $g_i = h_i s_i$, or $g_i = h_i$ and $g_{i-1} = h_i s_i$. Therefore $g_i = g_{i-1} s_i^{\epsilon_i}$ for some $\epsilon_i = \pm 1$. It follows by induction on i that $g_i = g_0 t_1 t_2 \cdots t_i$ for $i = 1,\dots,n$, where $t_i = s_i^{\epsilon_i}$. In particular

$$g_0 = g_n = g_0 t_1 t_2 \cdots t_n,$$

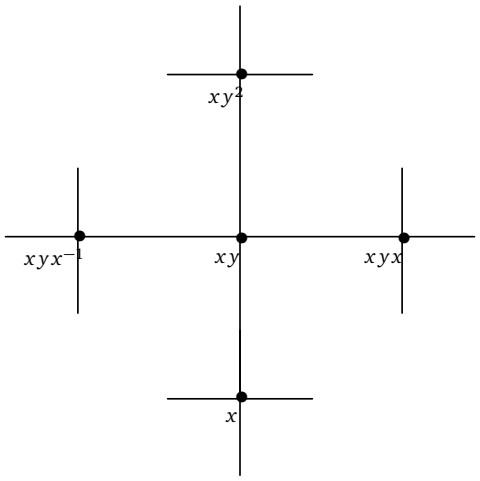
[65] This use of "free" is not directly related to its meaning when describing a group as being free.

and therefore $e_{F_S} = t_1 t_2 \cdots t_n$. Since distinct reduced words are distinct elements of F_S (by definition), it follows that $(t_1, t_2, \ldots, t_n)$ is not a reduced word, i.e., $n \geq 2$, and $t_i = t_{i-1}^{-1}$ for some $i \in \{2, \ldots, n\}$. This implies that $t_{i-1} t_i = e_{F_S}$, and therefore $g_i = g_{i-2} t_{i-1} t_i = g_{i-2}$. This contradicts the injectivity of the circuit γ, unless $n = i = 2$. However in this case, we either have $e_1 = (g_0, s_1) = e_2$ (if $\epsilon_1 = 1$ and $\epsilon_2 = -1$) or $e_1 = (g_1, s_1) = e_2$ (if $\epsilon_1 = -1$ and $\epsilon_2 = 1$), again contradicting the injectivity of γ.

□

EXAMPLE 25.38 Recall that if $\#S = 1$, then F_S is isomorphic to $\mathbb{Z}$, with the element of S corresponding to 1. It follows that the graph $\Gamma(F_S, S)$ is isomorphic to $\Gamma(\mathbb{Z}, \{1\})$, and hence to C_∞, which is a tree.

EXAMPLE 25.39 If $S = \{x, y\}$, then the part of the Cayley graph $\Gamma(F_S, S)$ near the vertex xy is pictured below, with horizontal edges corresponding to ones of the form (g, x) and vertical ones (g, y).



The entire graph extends infinitely in each direction, and looks like this near every vertex, perhaps best visualized as a fractal.

DEFINITION 25.40 We say a group G is *free* if G is isomorphic to F_S for some set S.

If $f : F_S \to G$ is an isomorphism, then the Cayley graphs $\Gamma(F_S, S)$ and $\Gamma(G, f(S))$ are isomorphic, so by Proposition 25.37, $\Gamma(G, f(S))$ is a tree. Since G acts freely on $\Gamma(G, f(S))$, we have the following consequence:

COROLLARY 25.41 *If G is a free group, then G acts freely on a tree.*

It turns out that the converse is also true:

THEOREM 25.42 *If G acts freely on a tree, then G is free.*

We won't give the proof, which is much harder. Note that the tree in the statement is not assumed to be a Cayley graph of G; it can be any tree. As a result we obtain the following consequence, called the Nielsen–Schreier Theorem:

COROLLARY 25.43 *Every subgroup a free group is a free group.*

Proof. Since G is a free group, it has a subset S such that $\Gamma(G,S)$ is a tree. Since G acts freely on $\Gamma(G,S)$, it follows that so does H. Theorem 25.42 therefore implies that H is a free group.

□

26 The Zariski topology

We now describe a construction that underpins a useful perspective in common to number theory and algebraic geometry.

Let R be a ring (commutative as usual). Define $\mathrm{Spec}(R)$ (called the *spectrum* of R) to be the set of prime ideals of R. For each ideal I of R, we define

$$V(I) = \{\, P \in \mathrm{Spec}(R) \,|\, I \subset P \,\}$$

to be the set of prime ideals containg I.

Note that by Proposition 15.14, we have that $V(I)$ is non-empty, unless $I = R$. In particular $V(\{0_R\}) = \mathrm{Spec}(R)$ is non-empty unless $R = \{0_R\}$. Note also that if $I \subset J$, then $V(J) \subset V(I)$.

EXAMPLE 26.1 Recall that all ideals of $\mathbb{Z}$ have the form $(n) = n\mathbb{Z}$, where $n \geq 0$, and that (n) is a prime ideal if and only if $n = 0$ or n is a prime number (see Example 7.51), i.e.,

$$\mathrm{Spec}(\mathbb{Z}) = \{(0)\} \cup \{\, (p) \,|\, p \text{ is prime} \,\}.$$

Therefore if $n > 0$, then $V((n))$ is the finite set

$$\{\, (p) \,|\, p \text{ is a prime dividing } n \,\}$$

(empty if $n = 1$), whereas $V((0)) = \mathrm{Spec}(\mathbb{Z})$.

For any ring R,we define a topology on $X = \mathrm{Spec}(R)$ by designating the closed subsets to be those of the form $V(I)$. Thus the open subsets of X are the complements of such subsets, i.e., $U \subset X$ is open if there exists an ideal $I \subset R$ such that

$$U = X \smallsetminus V(I) = \{\, P \in X \,|\, I \not\subset P \,\}.$$

We claim that this indeed defines a topology (as in Definition 24.1):

LEMMA 26.2 *Let R be a ring, let $X = \mathrm{Spec}(R)$, and let*

$$\mathcal{U} = \{\, X \smallsetminus V(I) \mid I \text{ is an ideal of } R \,\}.$$

(i) $\emptyset, X \in \mathcal{U}$;

(ii) if $U, U' \in \mathcal{U}$, *then* $U \cap U' \in \mathcal{U}$;

(iii) if $\{\, U_\alpha \,|\, \alpha \in \mathcal{A} \,\}$ *is any subset of* $\mathcal{U}$, *then* $\bigcup_{\alpha\in\mathcal{A}} U_\alpha \in \mathcal{U}$.

Proof. We prove the equivalent statements about closed subsets:

(i) We have $\{0_R\} \subset P$ for every $P \in X$, so $V(\{0_R\}) = X$, and therefore $\emptyset \in \mathcal{U}$. On the other hand R itself is not a prime ideal, so $R \not\subset P$ if $P \in X$. Therefore $V(R) = \emptyset$, and $X \in \mathcal{U}$.

(ii) Suppose that $U = X \smallsetminus V(I)$ and $U' = X \setminus V(I')$, where I and I' are ideals of R. We leave it as an exercise to show that

$$V(II') = V(I) \cup V(I'),$$

where II' is the ideal of R generated by $\{aa' \mid a \in I, a' \in I'\}$. Therefore $U \cap U' = X \smallsetminus V(II') \in \mathcal{U}$.

(iii) For each $\alpha \in \mathcal{A}$, we have $U_\alpha = X \smallsetminus V(I_\alpha)$ for some ideal I_α of R, and so $\bigcup_{\alpha\in\mathcal{A}} U_\alpha$ is the complement in X of $\bigcap_{\alpha\in\mathcal{A}} V(I_\alpha)$. Let J denote the ideal generated by $\bigcup_{\alpha\in\mathcal{A}} I_\alpha$. Thus if $P \in X$, then $J \subset P$ if and only if $I_\alpha \subset P$ for all $\alpha \in \mathcal{A}$, i.e., $V(J) = \bigcap_{\alpha\in\mathcal{A}} V(I_\alpha)$.

□

DEFINITION 26.3 If R is a ring, then the topology on $\mathrm{Spec}(R)$ defined above (i.e., with the closed subsets of R being those of the form $V(I) = \{P \in \mathrm{Spec}(R) \mid I \subset P\}$ where I is an ideal of R) is called the *Zariski topology*.

EXAMPLE 26.4 If K is a field, then $\mathrm{Spec}(K) = \{\{0_K\}\}$ has just one element, and the open subsets of $\mathrm{Spec}(K)$ are of course $\emptyset$ and $\mathrm{Spec}(K)$.

EXAMPLE 26.5 Suppose that $R = K_1 \times K_2$, where K_1 and K_2 are both fields. Then the only ideals of R are $\{(0_{K_1}, 0_{K_2})\}$, R,

$$P_1 = K_1 \times \{0_{K_2}\} \quad \text{and} \quad P_2 = \{0_{K_1}\} \times K_2.$$

Only these last two are prime ideals, so $\mathrm{Spec}(R) = \{P_1, P_2\}$. The four ideals I of R give the four subsets of $\mathrm{Spec}(R)$ as closed subsets $V(I)$, so $\mathrm{Spec}(R)$ has the discrete topology (see Example 24.3).

EXAMPLE 26.6 Recall from Example 26.1 that $\mathrm{Spec}\,\mathbb{Z}$ consists of the zero ideal and the ideal $p\mathbb{Z}$ for each prime (number) p, and the only infinite closed subset of $\mathrm{Spec}(\mathbb{Z})$ is the whole set $\mathrm{Spec}(\mathbb{Z})$ itself. On the other hand, for each finite subset

$$S \subset \{p\mathbb{Z} \mid p \text{ is a prime number}\},$$

we can let n be the product of the primes p such that $p\mathbb{Z} \in S$, and we see that $S = V(I)$ is a closed subset of $\mathrm{Spec}(\mathbb{Z})$. Therefore the closed subsets of $\mathrm{Spec}(\mathbb{Z})$ are:

- $\mathrm{Spec}(\mathbb{Z})$ itself;
- finite subsets $S \subset \mathrm{Spec}(\mathbb{Z})$ such that $(0) \notin S$.

Note in particular that the only closed subset of $\mathrm{Spec}(\mathbb{Z})$ containing (0) is the whole space $\mathrm{Spec}(\mathbb{Z})$. Therefore (0) is an element of every non-empty open subset $U \subset \mathrm{Spec}(\mathbb{Z})$. In particular $\mathrm{Spec}(\mathbb{Z})$ is not Hausdorff (since for example there are no open subsets U and U' such that $(0) \in U$, $(2) \in U'$, and $U \cap U' = \emptyset$).

EXAMPLE 26.7 Let $R = \mathbb{C}[X]$. Then R is a PID, so its ideals have the form (f), where either $f = 0$ or f is non-zero polynomial, which we can assume is monic (since $(f) = (\alpha f)$ for $\alpha \in \mathbb{C}^\times$). Since every non-constant polymonial in $\mathbb{C}[X]$ has roots, we may write it in the form

$$f(X) = (X - \alpha_1)(X - \alpha_2)\cdots(X - \alpha_n)$$

where $n = \deg(f)$; furthermore (f) is a prime ideal if and only if $n = 1$ (or $f = 0$). Therefore

$$\mathrm{Spec}(R) = \{(0)\} \cup \{(X - \alpha) \mid \alpha \in \mathbb{C}\};$$

i.e., it contains the zero ideal and an element for each $\alpha \in \mathbb{C}$. Exactly as in Example 26.6, we find that the closed subsets are

- $\mathrm{Spec}(R)$ itself;
- finite subsets $S \subset \mathrm{Spec}(R)$ such that $(0) \notin S$.

Thus exactly as in the preceding example, we conclude that (0) is an element of every non-empty open subset of $\mathrm{Spec}(R)$.

EXAMPLE 26.8 Suppose now that $R = \mathbb{C}[X,Y]$ is the polynomial ring over $\mathbb{C}$ in two variables. Since R is a domain, the ideal (0) is prime, and is therefore an element of $\mathrm{Spec}(R)$.

Similarly to the preceding example, we also have an element of $\mathrm{Spec}(R)$ for each $(\alpha,\beta) \in \mathbb{C}^2$; indeed one can show that the ideal

$$P_{\alpha,\beta} := (X-\alpha, Y-\beta)$$

is prime, and in fact maximal since it is the kernel of the surjective homomorphism $\varphi_{\alpha,\beta} : R \to \mathbb{C}$ defined by $f \mapsto f(\alpha,\beta)$. Conversely Hilbert's Nullstellensatz (which we will not prove) implies that every maximal ideal of R has the form $P_{\alpha,\beta}$ for some $(\alpha,\beta) \in \mathbb{C}^2$.

In this case however there are other prime ideals besides (0) and those of the form $P_{\alpha,\beta}$. For example (X) and $(Y-X^2)$ are prime ideals. It turns out in this case that there are three sorts of prime ideals:

- (0);
- (f), for irreducible $f \in R$;
- $P_{\alpha,\beta}$, for $\alpha,\beta \in \mathbb{C}$.

We will not prove this, but just mention it to give more indication of the relation with algebraic geometry. The closed subsets of $\mathrm{Spec}(R)$ are related to algebraic subsets of $\mathbb{C}^2$, i.e., sets of solutions of systems of polynomial equations. For example if I is the ideal (f) where $f(X,Y) = XY - X^3 = X(Y-X^2)$, then the closed subset $V(I)$ consists of the elements

- (X), $(Y-X^2)$ (since X are $Y-X^2$ are the irreducible factors of f), and
- $P_{\alpha,\beta}$ for $\alpha,\beta \in \mathbb{C}^2$ such that $f(\alpha,\beta) = \alpha(\beta-\alpha^2) = 0$ (i.e., $X(Y-X^2)$ is an element of $P_{\alpha,\beta} = \ker(\varphi_{\alpha,\beta})$), or equivalently either $\alpha = 0$ or $\beta = \alpha^2$.

Thus the closed subset $V(I)$ includes the maximal ideals of R corresponding to points in the algebraic subset[66] of $\mathbb{C}^2$ defined by the equation $XY = X^3$, along with the prime ideals generated by the irreducible factors of $XY - X^3$.

Let us return now to the general setting of a ring R. If $s \in R$, then

$$V((s)) = \{\, P \in \mathrm{Spec}(R) \,|\, s \in P \,\}$$

is a closed subset of $\mathrm{Spec}(R)$, so that its complement, which we denote U_s, is open. Note that if P is a prime ideal of R and $s^n \in P$ for some $n > 0$, then $s \in P$. Furthermore $s^0 = 1_R \notin P$, so

$$\begin{aligned} U_s &= \{\, P \in \mathrm{Spec}(R) \,|\, s \notin P \,\} \\ &= \{\, P \in \mathrm{Spec}(R) \,|\, s^n \notin P \text{ for all } n \geq 0 \,\} \\ &= \{\, P \in \mathrm{Spec}(R) \,|\, P \cap S = \emptyset \,\}, \end{aligned}$$

where S is the multiplicative subset $\{\, s^n \,|\, n \in \mathbb{Z}, n \geq 0 \,\}$ of R. It therefore follows from Corollary 15.6 that we have a bijection

$$\mathrm{Spec}(R_S) \underset{P_S \hookleftarrow P}{\overset{Q \mapsto i^{-1}(Q)}{\rightleftarrows}} U_s \subset \mathrm{Spec}(R)$$

(where $i : R \to R_S$ is the ring homomorphism defined by $i(r) = r/1_R$).

[66] Its intersection with $\mathbb{R}^2$ can be viewed as the union of the line $X = 0$ and the parabola $Y = X^2$, but we chose to work with polynomials over the algebraically closed field $\mathbb{C}$ in order to be able to give a more concise description of $\mathrm{Spec}(R)$.

This is in fact a homeomorphism between $\mathrm{Spec}(R_S)$ and the open subset U_s of $\mathrm{Spec}(R)$ (with its subspace topology, as defined in Example 24.5). We first remark though that for any ring homomorphism $f : R \to T$, one has a continuous function[67] $f^* : \mathrm{Spec}(T) \to \mathrm{Spec}(R)$ defined by $f^*(Q) = f^{-1}(Q)$; we leave the proof as an exercise. Thus if S is *any* multiplicative subset of R, we have the ring homomorphism $i : R \to R_S$, and hence the continuous map $i^* : \mathrm{Spec}(R_S) \to \mathrm{Spec}(R)$. Furthermore it is injective by Corollary 15.4, and we leave it as an exercise to show that i^* defines a homeomorphism to its image (with its topology as a subspace of $\mathrm{Spec}(R)$). Returning to the case when $S = \{ s^n | n \in \mathbb{Z}, n \geq 0 \}$ for some $s \in R$, we saw that this image is the open subset U_s of $\mathrm{Spec}(R)$ defined above.

EXAMPLE 26.9 Let $R = \mathbb{Z}$ and $s = n$ for some $n \neq 0$. Recall from Example 26.1 that the elements of $\mathrm{Spec}(\mathbb{Z})$ are (0) and (p) for prime numbers p. The closed subset $V((n))$ is the finite subset of $\mathrm{Spec}(\mathbb{Z})$ whose elements are (p) for prime divisors p of n. The open subset

$$U_s = \{(0)\} \cup \{ (p) \,|\, p \text{ is a prime not dividing } n \}$$

of $\mathrm{Spec}(\mathbb{Z})$ is therefore homeomorphic to $\mathrm{Spec}(\mathbb{Z}_S)$, where $\mathbb{Z}_S$ is the subring $\mathbb{Z}[1/n] = \{ m/n^k \,|\, m, k \in \mathbb{Z}, k \geq 0 \}$ of $\mathbb{Q}$.

Returning to the setting of an arbitrary multiplicative subset S of R, the image of $i^* : \mathrm{Spec}(R_S) \to \mathrm{Spec}(R)$ need not be open. Suppose for example that $S = R \setminus Q$ for some $Q \in \mathrm{Spec}(R)$.

PROPOSITION 26.10 *Suppose that $S = R \setminus Q$ for some prime ideal Q of R, and let $i : R \to R_S$ be the homomorphism defined by $i(r) = r/1_R$. Then the image of the $i^* : \mathrm{Spec}(R_S) \to \mathrm{Spec}(R)$ is the intersection of the open subsets $U \subset \mathrm{Spec}(R)$ such that $Q \in U$.*

Proof. Suppose first that P is in the image of i^*. By Corollary 15.6, this means that $P \cap S = \emptyset$, i.e., $P \subset Q$. We must show that if U is an open subset of $\mathrm{Spec}(R)$ such that $Q \in U$, then $P \in U$ as well. Since U is open, we have $U = \mathrm{Spec}(R) \setminus V(I)$ for some ideal I of R, and since $Q \in U$, we have $Q \notin V(I)$, i.e., $I \not\subset Q$. Since $P \subset Q$, it follows that $I \not\subset P$, i.e., $P \in U$.

Conversely suppose that $P \in U$ for every open subset $U \subset \mathrm{Spec}(R)$ such that $Q \in U$. In particular, for each $s \in S$, we have $Q \in U_s$, so $P \in U_s$, i.e., $s \notin P$. This proves that $P \cap S = \emptyset$ (i.e., $P \subset Q$), so it follows from Corollary 15.6 that P is in the image of i^*. □

EXAMPLE 26.11 Suppose that $R = \mathbb{Z}$ and $S = \mathbb{Z} \setminus Q$, where $Q = (p)$ for some prime p, so $\mathbb{Z}_S$ is as in Example 15.8. We saw that $\mathbb{Z}_S$ has exactly two prime ideals, namely $\{0\}$ and $p\mathbb{Z}_S$, so the image of $i^* : \mathrm{Spec}(\mathbb{Z}_S) \to \mathrm{Spec}(\mathbb{Z})$ is the subset $\{(0), (p)\}$. Note that this is indeed the intersection of the open subsets of $\mathrm{Spec}(\mathbb{Z})$ containing (p) (since (0) is in *every* non-empty open subset, and for each prime $q \neq p$, we have for example the open subset U_q containing (p) as an element, but not (q)). Furthermore the topology on $\mathrm{Spec}(\mathbb{Z}_S)$ corresponds to the one arising from $\{(0), (p)\}$ as a subspace of $\mathrm{Spec}(\mathbb{Z})$: its open subsets are

$$\emptyset, \quad \{(0)\} \quad \text{and} \quad \{(0), (p)\}$$

(but not (p), which is closed but not open).

[67] Furthermore if $g : T \to U$ is another ring homomorphism, then we have $(g \circ f)^* = f^* \circ g^*$, so we can view Spec as defining a functor from the category of rings to that of topological spaces, but reversing arrows (or *contravariant*), i.e., $\mathsf{Rings} \to \mathsf{Top}^{\mathrm{op}}$. We will later see encounter more contravariant functors.

Finally we remark that Proposition 26.10 partly explains the source of the terminology "localisation." If $S = R \setminus Q$, then $\mathrm{Spec}(R_S)$ is identified with the intersection of arbitrarily small neighborhoods of Q in $\mathrm{Spec}(R)$, so it describes $\mathrm{Spec}(R)$ "locally" at the point Q. Furthermore the type of relation described in Example 26.8 imparts a more geometric perspective to this notion, applicable not only to rings arising in the conventional framework of algebraic geometry (i.e., quotients of polynomial rings over fields), but to arbitary rings.

EXAMPLE 26.12 Prove the claim that if $f : R \to S$ is a homomorphism of rings, then for each prime ideal Q of S, its preimage $f^{-1}(Q)$ is a prime ideal of R, so we obtain a function $\mathrm{Spec}(S) \to \mathrm{Spec}(R)$ defined by $Q \mapsto f^{-1}(Q)$. We obtain a morphism

$$\mathrm{Spec}(f) \in \mathrm{Hom}_{\mathsf{Top}}(\mathrm{Spec}(S), \mathrm{Spec}(R))$$

in the category of topological spaces. Show that this makes Spec a contravariant functor from Rings to Top.

27 Contravariant functors

Now fix N; for each homomorphism $f : M \to M'$ of R-modules, we now instead associate an R-linear homomorphism $\mathrm{Ext}^i_R(M', N) \to \mathrm{Ext}^i_R(M, N)$.

Recall that $\mathrm{Ext}^i_R(M, N) = H_{-i}(C_\bullet)$, where $(C_\bullet, \delta_\bullet)$ is defined by applying $\mathrm{Hom}_R(-, N)$ to $P_\bullet$ for any projective resolution $P_\bullet \xrightarrow{\epsilon} M$, and similarly $\mathrm{Ext}^i_R(M', N) = H_{-i}(C'_\bullet)$, where $(C'_\bullet, \delta'_\bullet)$ is defined by applying $\mathrm{Hom}_R(-, N)$ to $P'_\bullet$ for any projective resolution $P'_\bullet \xrightarrow{\epsilon'} M$. By Lemma 21.8, there is a chain map $\eta_\bullet : P_\bullet \to P'_\bullet$ compatible with f (in the sense that $f \circ \epsilon = \epsilon' \circ \eta_0$), and any two such chain maps are homotopic.

We now apply $\mathrm{Hom}_R(-, N)$ to $\eta_\bullet$ to get a chain map $\theta_\bullet : C'_\bullet \to C_\bullet$. More precisely, we claim that the R-linear homomorphisms

$$\begin{array}{rccccccc} \theta_i : & C'_i & = & \mathrm{Hom}_R(P'_{-i}, N) & \longrightarrow & \mathrm{Hom}_R(P_{-i}, N) & = & C_i \\ & & & f & \longmapsto & f \circ \eta_{-i} & & \end{array}$$

define a chain map. Indeed the fact that $\eta_\bullet$ is a chain map implies that $\eta_{-i} \circ d_{-i+1} = d'_{-i+1} \circ \eta_{-i+1}$ for all $i \in \mathbb{Z}$, and hence

$$\begin{array}{rcl} \delta_i(\theta_i(f)) & = & \delta_i(f \circ \eta_{-i}) = f \circ \eta_{-i} \circ d_{-i+1} \\ & = & f \circ d'_{-i+1} \circ \eta_{-i+1} = \theta_{i-1}(f \circ d'_{-i+1}) = \theta_{i-1}(\delta'_i(f)) \end{array}$$

for all $f \in C'_i = \mathrm{Hom}_R(P'_{-i}, N)$. Therefore $\theta_\bullet$ induces an R-linear homomorphism on the homology of the chain complexes:

$$\theta_{-i,*} : \mathrm{Ext}^i_R(M', N) = H_{-i}(C'_\bullet) \longrightarrow H_{-i}(C_\bullet) = \mathrm{Ext}^i_R(M, N).$$

Again we need to prove that this homomorphism is independent of the choices made, namely the projective resolutions $P_\bullet \xrightarrow{\epsilon} M$ and $P'_\bullet \xrightarrow{\epsilon'} M'$, as well as the chain map $\eta_\bullet$ compatible with f (given by Lemma 21.8). Suppose then that we choose different projective resolutions $Q_\bullet \xrightarrow{\varepsilon} M$ and $Q'_\bullet \xrightarrow{\varepsilon'} M'$, and let $\omega_\bullet : Q_\bullet \to Q'_\bullet$ be a chain map compatible with f. Applying $\mathrm{Hom}_R(-, N)$ to $Q_\bullet$, $Q'_\bullet$ and $\omega_\bullet$ then yields chain complexes $D_\bullet$, $D'_\bullet$, and a chain map $\zeta_\bullet : D'_\bullet \to D_\bullet$. We need to show that the resulting homomorphism $\zeta_{-i,*} : H_{-i}(D'_\bullet) \to H_{-i}(D_\bullet)$ corresponds to $\theta_{-i,*}$ via the canonical isomorphisms

$H_{-i}(D_\bullet) \xrightarrow{\sim} H_{-i}(C_\bullet)$ and $H_{-i}(D'_\bullet) \xrightarrow{\sim} H_{-i}(C'_\bullet)$; i.e. the diagram

$$\begin{array}{ccc} H_{-i}(D'_\bullet) & \xrightarrow{\zeta_{-i,*}} & H_{-i}(D_\bullet) \\ \downarrow\wr & & \downarrow\wr \\ H_{-i}(C'_\bullet) & \xrightarrow{\theta_{-i,*}} & H_{-i}(C_\bullet) \end{array} \tag{41}$$

commutes. Again let $\varphi_\bullet : P_\bullet \to Q_\bullet$ denote the chain map as in (23), but with $Q_\bullet$ instead of $P'_\bullet$, so that applying $\mathrm{Hom}_R(-,N)$ yields the chain map $\psi_\bullet : D_\bullet \to C_\bullet$ inducing the isomorphism $H_{-i}(D_\bullet) \xrightarrow{\sim} H_{-i}(C_\bullet)$. Similarly let $\varphi'_\bullet : P'_\bullet \to Q'_\bullet$ denote chain map giving $\psi'_\bullet : D'_\bullet \to C'_\bullet$ and hence $H_{-i}(D'_\bullet) \xrightarrow{\sim} H_{-i}(C'_\bullet)$. The commutativity of (41) would then follow from a corresponding statement about the chain maps, i.e. the commutativity of

$$\begin{array}{ccc} D'_\bullet & \xrightarrow{\zeta_\bullet} & D_\bullet \\ \downarrow\psi'_\bullet & & \downarrow\psi_\bullet \\ C'_\bullet & \xrightarrow{\theta_\bullet} & C_\bullet. \end{array} \tag{42}$$

However it is not necessarily the case that $\psi_\bullet \circ \zeta_\bullet = \theta_\bullet \circ \psi'_\bullet$; what is true instead is that the two composites are homotopic, and therefore induce the same homomorphisms on homology.

To prove that they are homotopic, note that the two chain maps $P_\bullet \to Q'_\bullet$ defined by the composites $\omega_\bullet \circ \varphi_\bullet$ and $\varphi'_\bullet \circ \eta_\bullet$ are both as in Lemma 21.8 (for $f : M \to M'$), and are therefore homotopic. Furthermore $\psi_\bullet \circ \zeta_\bullet$ is given by applying $\mathrm{Hom}_R(-,N)$ to $\omega_\bullet \circ \varphi_\bullet$, and similarly $\theta_\bullet \circ \psi'_\bullet$ is given by applying $\mathrm{Hom}_R(-,N)$ to $\varphi'_\bullet \circ \eta_\bullet$, so the argument preceding Lemma 21.8 shows that they are indeed homotopic.

We have now shown that $\theta_{-i,*} : \mathrm{Ext}^i_R(M',N) \to \mathrm{Ext}^i_R(M,N)$ is independent of the choices made in its definition. We thus obtain a contravariant functor $G = \mathrm{Ext}^i_R(-,N)$ from the category R-Mod to itself, where we let $G(f) = \theta_{-i,*}$ for $f \in \mathrm{Hom}_R(M,M')$. Indeed if $f = \mathrm{id}_M$, then we may choose $P'_\bullet = P_\bullet$ and $\eta_\bullet = \mathrm{id}_{P_\bullet}$, so that $C'_\bullet = C_\bullet$, $\theta_\bullet = \mathrm{id}_{C_\bullet}$ and hence $G(\mathrm{id}_M)$ is the identity. Furthermore if $f : M \to M'$ and $f' : M' \to M''$ are R-module homomorphisms, and $\eta'_\bullet : P'_\bullet \to P''_\bullet$ is chosen compatibly with f' (where $P''_\bullet \xrightarrow{\epsilon''} M''$ is a projective resolution), then $\eta'_\bullet \circ \eta_\bullet$ is compatible with $f' \circ f$. Applying $\mathrm{Hom}_R(-,N)$ then yields chain maps $\theta_\bullet : C'_\bullet \to C_\bullet$, $\theta'_\bullet : C''_\bullet \to C'_\bullet$ such that $\theta_\bullet$ induces $G(f)$, $\theta'_\bullet$ induces $G(f')$, and $\theta_\bullet \circ \theta'_\bullet$ induces $G(f' \circ f)$, from which it follows that $G(f' \circ f) = G(f) \circ G(f')$.

Contravariant long exact sequence of Ext

Let us now fix an R-module N and consider the interaction between the contravariant functors $G^i = \mathrm{Ext}^i_R(-,N)$ and exact sequences of the form $0 \longrightarrow M \xrightarrow{f} M' \xrightarrow{f'} M'' \longrightarrow 0$. We leave it as an exercise to show that if $P_\bullet \xrightarrow{\epsilon} M$ and $P''_\bullet \xrightarrow{\epsilon''} M''$ are any projective resolutions of M and M'', then there is a projective resolution[68] $P'_\bullet \xrightarrow{\epsilon'} M$ such that each $P'_i = P_i \oplus P''_i$, and furthermore the obvious R-linear homomorphisms

$$\varphi_i : P_i \to P'_i \quad \text{and} \quad \varphi'_i : P'_i \to P''_i,$$

(so $\varphi_i(x) = (x,0)$ and $\varphi'_i(x,y) = y$) define chain maps $\varphi_\bullet : P_\bullet \to P'_\bullet$ and $\varphi'_\bullet : P'_\bullet \to P''_\bullet$ extending compatibly with f and f' to the resolutions (as in Lemma 21.8).

[68] We caution however that the homomorphisms $d'_i : P'_i \to P'_{i-1}$ might not be the "obvious" ones.

Applying $\mathrm{Hom}_R(-,N)$ to $P_\bullet$, $P'_\bullet$ and $P''_\bullet$ therefore yields chain complexes and chain maps

$$\psi_\bullet : C'_\bullet \longrightarrow C_\bullet \quad \text{and} \quad \psi'_\bullet : C''_\bullet \longrightarrow C'_\bullet$$

such that $G^i(f) : \mathrm{Ext}^i_R(M',N) \to \mathrm{Ext}^i_R(M,N)$ is the homomorphism $\psi_{-i,*} : H_{-i}(C'_\bullet) \to H_{-i}(C_\bullet)$, and similarly $G^i(f') = \psi'_{-i,*}$. Furthermore the fact that $P'_i = P_i \oplus P''_i$ (with φ_i and φ'_i being the obvious inclusion and projection) ensures that the sequences

$$0 \longrightarrow C''_i \xrightarrow{\psi'_i} C'_i \xrightarrow{\psi_i} C_i \longrightarrow 0$$

are exact (also an exercise). We can therefore apply Lemma 18.7 again to deduce the following:

COROLLARY 27.1 *Suppose that N is an R-module, and let G^i denote the contravariant functor $\mathrm{Ext}^i_R(-,N)$. If*

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

is an exact sequence of R-modules, then there are R-linear homomorphisms $\partial^i : G^i(M) \to G^{i+1}(M'')$ such that the following sequence is exact:

$$\begin{aligned} 0 \longrightarrow G^0(M'') &\xrightarrow{G^0(f')} G^0(M') \xrightarrow{G^0(f)} G^0(M) \xrightarrow{\partial^0} G^1(M'') \\ &\xrightarrow{G^1(f')} G^1(M') \xrightarrow{G^1(f)} G^1(M) \xrightarrow{\partial^1} G^2(M'') \longrightarrow \cdots . \end{aligned}$$

EXAMPLE 27.2 Consider the same exact sequence of $\mathbb{Z}$-modules as in Example 22.2, but now viewed as $0 \longrightarrow M \xrightarrow{f} M' \xrightarrow{f'} M'' \longrightarrow 0$, and let $N = \mathbb{Z}/2\mathbb{Z}$. Applying Corollary 27.1 now gives a long exact sequence

$$\begin{aligned} 0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) &\xrightarrow{G^0(f')} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{G^0(f)} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \\ &\xrightarrow{\partial_0} \mathrm{Ext}^1(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{G^1(f')} \mathrm{Ext}^1(\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \xrightarrow{G^1(f)} \mathrm{Ext}^1(\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \\ &\xrightarrow{\partial^1} \mathrm{Ext}^2(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \longrightarrow \cdots . \end{aligned}$$

Since $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z},\mathbb{Z}/2\mathbb{Z})$ can be identified with $\mathbb{Z}/2\mathbb{Z}$, $\mathrm{Ext}^i(\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) = 0$ for all $i > 0$ (since, for example, $\mathbb{Z}$ is a projective $\mathbb{Z}$-module), the exact sequence takes the form

$$\begin{aligned} 0 \longrightarrow \mathbb{Z}/2\mathbb{Z} &\xrightarrow{G^0(f')} \mathbb{Z}/2\mathbb{Z} \xrightarrow{G^0(f)} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\partial^0} \mathrm{Ext}^1(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \\ &\xrightarrow{G^1(f')} 0 \xrightarrow{G^1(f)} 0 \xrightarrow{\partial^1} \mathrm{Ext}^2(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \longrightarrow 0 \longrightarrow \cdots \\ &\longrightarrow 0 \longrightarrow \mathrm{Ext}^i(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) \longrightarrow 0 \longrightarrow \cdots . \end{aligned}$$

A similar argument to Example 22.2 produces the same conclusions: $\mathrm{Ext}^1(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z})$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, and $\mathrm{Ext}^i(\mathbb{Z}/2\mathbb{Z},\mathbb{Z}/2\mathbb{Z}) = 0$ for all $i > 1$.

28 Tor: definition and examples

One can view the functors $\mathrm{Ext}^i_R(M,-)$ as measuring the failure of (right)-exactness of $\mathrm{Hom}_R(M,-)$, and similarly $\mathrm{Ext}^i_R(-,N)$ as measuring the failure of (right)-exactness of $\mathrm{Hom}_R(-,N)$. They are examples of *derived functors*; we will not develop the general theory of derived functors, but we will briefly discuss another example, namely Tor-functors, which measure the failure of exactness of tensor products.

Recall (from the beginning of §19) that if M is an R-module and $g : B \to C$ is a surjective homomorphism of R-modules, then $\mathrm{id}_M \otimes g$ is also surjective. In fact something

stronger is true: the functor $M \otimes_R -$, i.e. the functor F of Example 13.8, is right-exact. We will prove this using properties of $\mathrm{Hom}_R(-,N)$.

Recall that the contravariant functor $G = \mathrm{Hom}_R(-,N)$ is left-exact (Example 20.8), i.e. if $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is exact, then so is

$$0 \longrightarrow \mathrm{Hom}_R(C,N) \xrightarrow{G(g)} \mathrm{Hom}_R(B,N) \xrightarrow{G(f)} \mathrm{Hom}_R(A,N). \tag{43}$$

In fact we have the following converse:

LEMMA 28.1 *A sequence $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ of homomorphisms of R-modules is exact if and only if the sequence (43) is exact for all R-modules N.*

Proof. The "only if" direction is the statement that $\mathrm{Hom}_R(-,N)$ is left-exact (for all R-modules N).

Suppose then that (43) is exact for *all* R-modules N. We will prove that $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is exact by making some convenient choices for N.

To prove that g is surjective, let $N = C/\mathrm{im}(g)$. Recall that $G(g) : \mathrm{Hom}_R(C,N) \to \mathrm{Hom}_R(B,N)$ is defined by $\varphi \mapsto \varphi \circ g$. We have a canonical element of $\mathrm{Hom}_R(C,N)$ given by the projection map $\pi\colon C \to N$. Moreover, this maps to zero in $\mathrm{Hom}_R(B,N)$ since $\pi \circ g$ is zero as π vanishes on the image of g. So $\pi \in \ker(G(g))$, but the exactness of (43) implies that $G(g)$ is injective, so $\pi = 0$. On the other hand, π is surjective, so $N = 0$, and therefore $\mathrm{im}(g) = C$, i.e. g is surjective.

To prove that $\mathrm{im}(f) \subset \ker(g)$, let $N = C$ and consider $\mathrm{id}_C \in \mathrm{Hom}_R(C,N)$. The exactness of (43) implies that $G(g \circ f) = G(f) \circ G(g) = 0$, but $(G(g \circ f))(\mathrm{id}_C) = \mathrm{id}_C \circ g \circ f = g \circ f$, so $g \circ f = 0$, i.e. $\mathrm{im}(f) \subset \ker(g)$.

Finally to prove that $\ker(g) \subset \mathrm{im}(f)$, let $N = B/\mathrm{im}(f)$ and consider the quotient map $\pi : B \to N$. Then $\pi \circ f = 0$, so $\pi \in \ker(G(f))$, which coincides with $\mathrm{im}(G(g))$ by the exactness of (43). Therefore $\pi = \varphi \circ g$ for some $\varphi \in \mathrm{Hom}_R(C,N)$. It follows that if $b \in \ker(g)$, then $\pi(b) = \varphi(g(b)) = 0$, i.e. $b \in \mathrm{im}(f)$. □

We will also make use of the following:

LEMMA 28.2 *For any R-modules M, N and A, there is an isomorphism*

$$\mathrm{Hom}_R(M \otimes_R A, N) \xrightarrow{\alpha_A} \mathrm{Hom}_R(A, \mathrm{Hom}_R(M,N)),$$

which is "natural" in the sense[69] *that if $f : A \to B$ is a homomorphism of R-modules, then the resulting diagram*

$$\begin{array}{ccc} \mathrm{Hom}_R(M \otimes_R B, N) & \xrightarrow{G(\mathrm{id}_M \otimes f)} & \mathrm{Hom}_R(M \otimes_R A, N) \\ \downarrow{\scriptstyle \alpha_B} & & \downarrow{\scriptstyle \alpha_A} \\ \mathrm{Hom}_R(B, \mathrm{Hom}_R(M,N)) & \xrightarrow{F(f)} & \mathrm{Hom}_R(A, \mathrm{Hom}_R(M,N)) \end{array}$$

commutes, where G and F are the contravariant functors $\mathrm{Hom}_R(-,N)$ and $\mathrm{Hom}_R(-, \mathrm{Hom}_R(M,N))$.

[69] The isomorphisms α_D define what is called a *natural isomorphism* from the (contravariant) composite functor $F \circ (M \otimes_R -)$ to G. We have already seen other examples of natural isomorphisms of functors, such as in Example 20.2, which defines a natural isomorphism between the identity functor on R-Mod and $\mathrm{Hom}_R(R,-)$. There is also a similar, but more general, notion of a *natural transformation* from one functor to another.

Proof. We just sketch the proof since the details are tedious but straighforward.

If $\varphi : M \otimes_R A \to N$ is an R-linear homomorphism, then let

$$\alpha_A(\varphi) : A \longrightarrow \mathrm{Hom}_R(M,N)$$

be the R-linear homomorphism sending a to the R-linear homomorphism $M \to N$ defined by $m \mapsto \varphi(m \otimes a)$, i.e.

$$((\alpha_A(\varphi))(a))(m) = \varphi(m \otimes a).$$

One then checks the following:

- $(\alpha_A(\varphi))(a)$ is indeed an R-linear homomorphism, i.e. an element of $\mathrm{Hom}_R(M,N)$;
- the resulting function $\alpha_A(\varphi)$ is indeed an R-linear homomorphism, i.e. an element of $\mathrm{Hom}_R(A, \mathrm{Hom}_R(M,N))$;
- the resulting function α_A is indeed an R-linear homomorphism.

To show that α_A is bijective, one constructs the inverse β_A as follows: for $\psi \in \mathrm{Hom}_R(A, \mathrm{Hom}_R(M,N))$, the function $M \times A \longrightarrow N$ defined by $(m,a) \mapsto (\psi(a))(m)$ is R-bilinear, so there is a unique R-linear homomorphism $\beta_A(\psi)$ sending $m \otimes a$ to $(\psi(a))(m)$. The fact that α_A and β_A are each other's inverses is just a matter of unravelling definitions, as is the commutativity of the diagram in the statement of the lemma. □

We are now ready to prove that $M \otimes_R -$ is right-exact:

PROPOSITION 28.3 *The functor $M \otimes_R -$ is right-exact; i.e. if $A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$ is an exact sequence of homomorphisms of R-modules, then so is the sequence*

$$M \otimes_R A \xrightarrow{\mathrm{id}_M \otimes f} M \otimes_R B \xrightarrow{\mathrm{id}_M \otimes g} M \otimes_R C \longrightarrow 0.$$

Proof. Let N be an R-module, and let $Q = \mathrm{Hom}_R(M,N)$. By Lemma 28.1 (more specifically, the "only if" direction applied to Q instead of N), the sequence

$$0 \longrightarrow \mathrm{Hom}_R(C,Q) \xrightarrow{F(g)} \mathrm{Hom}_R(B,Q) \xrightarrow{F(f)} \mathrm{Hom}_R(A,Q)$$

is exact (where $F = \mathrm{Hom}_R(-,Q)$).

It then follows from Lemma 28.2 that the sequence

$$0 \longrightarrow \mathrm{Hom}_R(M \otimes_R C, N) \xrightarrow{G(\mathrm{id}_M \otimes g)} \mathrm{Hom}_R(M \otimes_R B, N) \xrightarrow{G(\mathrm{id}_M \otimes f)} \mathrm{Hom}_R(M \otimes_R A, N)$$

is exact.

Applying Lemma 28.1 again (now the "if" direction) implies that

$$M \otimes_R A \xrightarrow{\mathrm{id}_M \otimes f} M \otimes_R B \xrightarrow{\mathrm{id}_M \otimes g} M \otimes_R C \longrightarrow 0$$

is exact. □

We have now shown that the functor $F = M \otimes_R -$ is right-exact, so it is exact if and only if $M \otimes_R A \xrightarrow{F(f)} M \otimes_R B$ is injective for all injective R-module homomorphisms $A \xrightarrow{f} B$. However we saw in Example 19.1 that this can fail.

There are however R-modules M for which F is exact; such modules are called *flat*. For example it follows from (8) that there is a (natural) isomorphism $R^n \otimes_R A \xrightarrow{\sim} A^n$ for all R-modules A, and it follows that R^n is a flat R-module (if $A \xrightarrow{f} B$ is injective, then so is the function $A^n \to B^n$ defined by f in each component). More generally the same

argument shows that every free R-module is flat. Furthermore since every projective R-module is a direct summand of a free R-module, one easily deduces that projective modules are flat.

One can define functors measuring the failure of (now left-)exactness of $F = M \otimes_R -$ in a manner similar to the construction of Ext. More precisely, let $P_\bullet \xrightarrow{\epsilon} M$ be a free (or projective) resolution, form the chain complex $C_\bullet = G(P_\bullet)$, where $G = - \otimes_R N$, and let $\mathrm{Tor}_i^R(M,N) = H_i(C_\bullet)$. Arguments similar to those for Ext show that this is independent of the choice of projective resolution and functorial in both M and N. Furthermore $\mathrm{Tor}_0^R(M,N)$ can be identified with $M \otimes_R N$. Rather than going into the details, we will just give an example.

EXAMPLE 28.4 Let $R = \mathbb{Z}$, $M = \mathbb{Z}/m\mathbb{Z}$ and $N = \mathbb{Z}/n\mathbb{Z}$ (where m and n are positive integers). We have a free resolution of M with $P_0 = \mathbb{Z}$, $P_1 = \mathbb{Z}$, $P_i = 0$ for $i > 1$, and d_1 defined by $d_1(a) = ma$:

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbb{Z} \xrightarrow{d_1} \mathbb{Z} \xrightarrow{\epsilon} \mathbb{Z}/m\mathbb{Z} \longrightarrow 0.$$

Applying the functor $G = - \otimes_{\mathbb{Z}} N$ gives the chain complex $C_\bullet$ with $C_i = P_i \otimes_{\mathbb{Z}} (\mathbb{Z}/n\mathbb{Z})$. Since $\mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/n\mathbb{Z})$ can be identified with $\mathbb{Z}/n\mathbb{Z}$, we can identify $C_\bullet$ with the chain complex

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbb{Z}/n\mathbb{Z} \xrightarrow{\delta_1} \mathbb{Z}/n\mathbb{Z} \longrightarrow 0,$$

where δ_1 is the homomorphism corresponding to $d_1 \otimes \mathrm{id}_N$. Since $\overline{a} \in \mathbb{Z}/n\mathbb{Z}$ corresponds to $a \otimes \overline{1}$, and $d_1(a) = ma$, we have $\delta_1(\overline{a}) = \overline{ma}$. It follows that $\mathrm{im}(\delta_1) = d\mathbb{Z}/n\mathbb{Z}$ and $\ker(\delta_1) = d^{-1}n\mathbb{Z}/n\mathbb{Z}$, where $d = \gcd(m,n)$, so that

- $\mathrm{Tor}_0^{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) = H_0(C_\bullet) = (\mathbb{Z}/n\mathbb{Z})/\mathrm{im}(\delta_1)$ is isomorphic to $\mathbb{Z}/d\mathbb{Z}$;
- $\mathrm{Tor}_1^{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) = H_1(C_\bullet) = \ker(\delta_1) = d^{-1}n\mathbb{Z}/n\mathbb{Z}$ is also isomorphic to $\mathbb{Z}/d\mathbb{Z}$;
- $\mathrm{Tor}_i^{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) = H_i(C_\bullet) = 0$ for $i > 1$.

Note that the description of $\mathrm{Tor}_0^{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z})$ is consistent (in view of Example 12.4) with the claim that $\mathrm{Tor}_0^R(M,N)$ can be identified with $M \otimes_R N$.

29 Homology of simplicial complexes

Recall that in §17, we computed the homology of certain chain complexes associated to graphs (see Examples 17.11–17.14). We defined the notion of a graph purely combinatorially, but we can also view its homology groups as those of the associated topological space, i.e. the one you draw to represent the graph. One can formalize the intuition behind the construction of this topological space, and apply it in higher dimension to arrive at the notion of a "simplicial complex." The associated topological spaces will be built from k-simplices, where a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangular region in a plane, etc. More precisely:

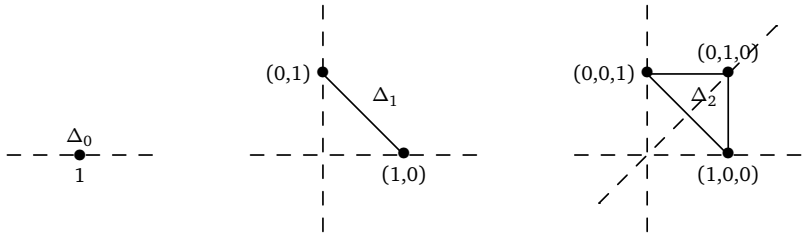
DEFINITION 29.1 Let k be a non-negative integer. The *k-simplex*, denoted Δ_k, is the topological space

$$\{(x_0, x_1, \ldots, x_k) \in \mathbb{R}^{k+1} \mid x_0 + x_1 + \cdots x_k = 1,\ x_0, x_1, \ldots, x_k \geq 0\}$$

(with its usual topology as a subspace of $\mathbb{R}^n$).

So in low dimension we can picture them as follows:

(44)



where Δ_2 is the triangular region bounded by the three line segments.

Rather than consider the general theory of homology of topological spaces, we will restrict our attention to ones built, via combinatorial input, out of k-simplices. One can then define the homology groups directly from the combinatorial data, making their calculation more explicit.

In the same way that the combinatorial description of a graph specifies the endpoints of each edge, the data of a simplicial complex will specify the lower-dimensional simplices that comprise the boundary of each simplex. Note that the boundary of Δ_k consists of the points with at least one coordinate 0. The boundary is therefore comprised of the $k+1$ subspaces, called *facets*, defined by setting each of the coordinates equal to 0.

DEFINITION 29.2 For $i = 0, \ldots, k$, the i^{th} *facet* of Δ_k, denoted $\Delta_k^{(i)}$, is the subspace

$$\{(x_0, x_1, \ldots, x_k) \in \Delta_k \,|\, x_i = 0\}.$$

Note that (assuming[70] $k \geq 1$) each facet $\Delta_k^{(i)}$ is homeomorphic to the $(k-1)$-simplex Δ_{k-1} via

$$\begin{array}{ccc} \alpha_k^{(i)} : \Delta_{k-1} & \longrightarrow & \Delta_k^{(i)} \\ (x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) & \longleftrightarrow & (x_0, x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k). \end{array}$$

Furthermore if $0 \leq i < j \leq k$, then the above homeomorphisms identify the i^{th} facet of $\Delta_k^{(j)}$ with the $(j-1)^{\text{st}}$ facet of $\Delta_k^{(i)}$ (i.e. $\alpha_k^{(i)} \circ \alpha_{k-1}^{(j-1)} = \alpha_k^{(j)} \circ \alpha_{k-1}^{(i)}$).

EXAMPLE 29.3 As can be seen from the third figure in (44), the boundary of Δ_2 consists of the line segments $\Delta_2^{(0)}$ (with endpoints $(0,0,1) \leftrightarrow \Delta_1^{(0)}$ and $(0,1,0) \leftrightarrow \Delta_1^{(1)}$ under the identification with Δ_1), $\Delta_1^{(1)}$ (with endpoints $(0,0,1) \leftrightarrow \Delta_1^{(0)}$ and $(1,0,0) \leftrightarrow \Delta_1^{(1)}$) and $\Delta_1^{(1)}$ (with endpoints $(0,1,0) \leftrightarrow \Delta_1^{(0)}$ and $(1,0,0) \leftrightarrow \Delta_1^{(1)}$).

EXAMPLE 29.4 The 3-simplex Δ_3 is a regular tetrahedron in $\mathbb{R}^4$ whose four faces are the four facets $\Delta_3^{(i)}$, each of which can be identified with the 2-simplex Δ_2.

We now define the type of combinatorial data to which we will associate topological spaces and chain complexes.

DEFINITION 29.5 A *(finite) simplicial (n-)complex* is a sequence of (finite) sets S_0, S_1, $\ldots$, S_n, together with *boundary (or face) functions* $f_k^{(i)} : S_k \to S_{k-1}$ for $k = 1, \ldots, n$ and $i = 0, \ldots, k$ such that

$$f_{k-1}^{(i)} \circ f_k^{(j)} = f_{k-1}^{(j-1)} \circ f_k^{(i)}$$

for all i, j, k such that $0 \leq i < j \leq k$ and $k \geq 2$.

[70] Note that $\Delta_0^{(0)}$ is the empty set; the natural convention would be to define Δ_{-1} to be the empty set as well.

Before we give examples, we will explain how to build a topological space from a simplical complex. We will only consider finite simplicial complexes (i.e. the sets S_k are finite), although the definitions and constructions make sense without this assumption.

Firstly let $X_0 = \Delta_0 \times S_0$ with the discrete[71] topology. Note that since Δ_0 is just a single point, the product can be identified with S_0 itself, as a (finite) set of points with the discrete topology.

Now let $\widetilde{X}_1 = \Delta_1 \times S_1$ endowed with the product topology, where Δ_1 has its usual topology as a line segment, and S_1 has the discrete topology. Thus $\widetilde{X}_1$ is a (finite) disjoint union[72] of copies of Δ_1 indexed by the elements of S_1. Thus if $S_1 = \{e_1, \dots, e_m\}$ and $S_0 = \{v_1, \dots, v_m\}$, then so far we have (writing $\Delta_{k,s}$ for $\Delta_k \times \{s\}$):

$$\widetilde{X}_1 = \left\{ \begin{array}{ll} \text{———} & \Delta_{1,e_1} \\ \text{———} & \Delta_{1,e_2} \\ \vdots & \\ \text{———} & \Delta_{1,e_m} \end{array} \right. \qquad X_0 = \left\{ \begin{array}{ll} \bullet & \Delta_{0,v_1} \\ \bullet & \Delta_{0,v_2} \\ \vdots & \\ \bullet & \Delta_{0,v_n} \end{array} \right.$$

Now we define X_1 by "gluing" the boundary of $\widetilde{X}_1$ to X_0 via identifications dictated by the boundary functions $f_1^{(i)}$ for $i = 0, 1$. More precisely, letting Y_1 denote the boundary of $\widetilde{X}_1$, i.e.

$$Y_1 = \{(0,1),(1,0)\} \times S_1 = (\Delta_1^{(0)} \cup \Delta_1^{(1)}) \times S_1,$$

we have the function $\beta_1 : Y_1 \to X_0$ defined by $((0,1),e) \mapsto (1, f_0^{(0)}(e))$ and $((1,0),e) \mapsto (1, f_1^{(0)}(e))$, i.e. sending the "initial" endpoint of the copy of Δ_1 indexed by e to the copy of Δ_0 indexed by $f_1^{(0)}(e)$, and similarly for "final" endpoints using $f_1^{(1)}$. The resulting topological space is best conveyed by the intuitive description of gluing each $y \in Y_1$ to $\beta_1(y)$, but for a more formal definition, let X_1 be the quotient space (see Example 24.6) of the disjoint union $\widetilde{X}_1 \cup X_0$ by the equivalence relation generated[73] by $y \sim \beta_1(y)$ for $y \in Y_1$.

We can then iterate this process. More precisely, suppose that $1 < k \leq n$ and X_{k-1} has been defined. We then let $\widetilde{X}_k = \Delta_k \times S_k$ (where S_k has the discrete topology, so $\widetilde{X}_k$ is a disjoint union of copies of Δ_k indexed by the elements of S_k), and let $Y_k = \left(\bigcup_{i=0}^k \Delta_k^{(i)}\right) \times S_k$ be the boundary of $\widetilde{X}_k$. We can then define X_k by gluing $\widetilde{X}_k$ along Y_k to X_{k-1} via the function $\beta_k : Y_k \to X_{k-1}$ defined by $\beta_k(\alpha_k^{(i)}(x), s) = [(x, f_k^{(i)}(s))]$ (where $\alpha_k^{(i)}$ was the homeomorphism identifying Δ_{k-1} with $\Delta_k^{(i)}$, and $[(x,t)] \in X_{k-1}$ denotes the equivalence class of $(x,t) \in \widetilde{X}_{k-1}$), so we identify the i^{th} facet of the k-simplex indexed by s with the (image in X_{k-1}) of the $(k-1)$-simplex indexed by $f_k^{(i)}(s)$. The compatibility formula in Definition 29.5 is precisely what is required to ensure that β_k is well-defined:

[71] For consistency with the notation in the rest of the inductive construction, we could also denote X_0 by $\widetilde{X}_0$, and view its topology as being the one on the product $\Delta_0 \times S_0$ (where Δ_0 and S_0 both have the discrete topology).

[72] The disjoint union is given the obvious topology: U is open if its intersection with each copy of Δ_1 open.

[73] This is the "smallest" equivalence relation such that $y \sim \beta_1(y)$ for all $y \in Y_1$. So $x \sim x'$ if and only if any of the following hold: 1) $x = x'$, 2) $x \in Y_1$ and $x' = \beta_1(x) \in X_0$, 3) $x' \in Y_1$ and $x = \beta_1(x') \in X_0$, or 4) $x, x' \in Y_1$ and $\beta_1(x) = \beta_1(x')$.

if $y = (\alpha_k^{(i)}(x), s) = (\alpha_k^{(j)}(x'), s)$ is in the intersection of two facets (with $i < j$) on the same simplex $\Delta_{k,s}$, then $x = \alpha_{k-1}^{(j-1)}(x'')$ and $x' = \alpha_{k-1}^{(i)}(x'')$ for some $x'' \in \Delta_{k-2}$, so

$$\beta_{k-1}(x, f_k^{(i)}(s)) = (x'', f_{k-1}^{(j-1)}(f_k^{(i)}(s))) = (x'', f_{k-1}^{(i)}(f_k^{(j)}(s))) = \beta_{k-1}(x', f_k^{(j)}(s)),$$

and therefore $(x, f_k^{(i)}(s)) \sim (x', f_k^{(j)}(s))$.

Finally the process terminates with X_n.

EXAMPLE 29.6 A simplicial 1-complex is the same as a directed graph, with vertices $V = S_0$, edges $E = S_1$, and endpoints functions $t_0 = f_1^{(0)}$ and $t_1 = f_1^{(1)}$. The associated topological space X_1, obtained by gluing the endpoints of edges to the designated vertices, is the one described by the diagram representing the graph.

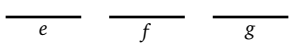
EXAMPLE 29.7 Let $S_0 = \{u, v, w\}$, $S_1 = \{e, f, g\}$, $S_2 = \{a, b\}$, and define the functions $f_k^{(i)}$ as follows:

- $f_1^{(0)}(e) = f_1^{(0)}(f) = u$, $f_1^{(0)}(g) = v$;
- $f_1^{(1)}(e) = v$, $f_1^{(1)}(f) = f_1^{(1)}(g) = w$;
- $f_2^{(0)}(a) = f_2^{(0)}(b) = e$;
- $f_2^{(1)}(a) = f_2^{(1)}(b) = f$;
- $f_2^{(2)}(a) = f_2^{(2)}(b) = g$.

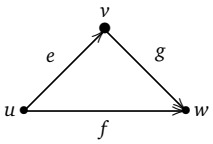
We leave it as an exercise to verify the formulas $f_1^{(i)} \circ f_2^{(j)} = f_1^{(j-1)} \circ f_2^{(i)}$ for $0 \le i < j \le 2$.

Now $X_0 (= \widetilde{X}_0)$ is just a discrete set of three points, which we label by the corresponding element of S_0:

Similarly $\widetilde{X}_1$ is a disjoint union of three 1-simplices (i.e. line segments) labeled by the elements of S_1:

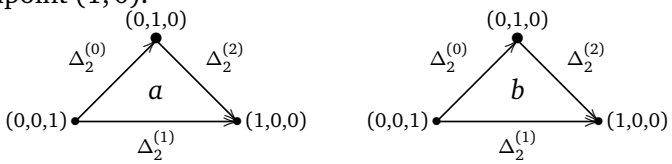


We then form X_1 by gluing the endpoint $(0,1) \in \Delta_1^{(0)}$ on (the 1-simplex indexed by) e to (the 0-simplex indexed by) $f_1^{(0)}(e) = u$, $(1,0) \in \Delta_1^{(1)}$ on e to $f_1^{(1)}(e) = v$, so e runs from u to v, and similarly f runs from u to w, and g from v to w. Systematically using arrows to indicate the direction on each 1-simplex from $(0,1)$ to $(1,0)$, we can therefore picture X_1 as follows:



Similarly $\widetilde{X}_2$ is a union of two 2-simplices; again for consistency, let us add arrows to the three 1-simplices $\Delta_2^{(i)}$ on their boundaries to indicate the direction from the endpoint[74] $(0,1)$ to the endpoint $(1,0)$:

(45)



[74] More precisely, the one to which it corresponds under $\alpha_2^{(i)}$.

Finally we form X_2 by gluing the boundary of the 1-simplex labeled a to X_1, sending $\Delta_2^{(0)}$ to $f_2^{(0)}(a) = e$, $\Delta_2^{(1)}$ to $f_2^{(1)}(a) = f$ and $\Delta_2^{(2)}$ to $f_2^{(2)}(a) = g$, and similarly for b. Thus X_2 is homeomorphic to the sphere $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$, with $\Delta_{2,a}$ corresponding to one hemisphere, $\Delta_{2,b}$ to the other, and X_1 to the equator.

Let us now explain how to associate a chain complex to a simplicial n-complex $S_0, S_1, \ldots, S_n$, with boundary functions $f_k^{(i)} : S_k \to S_{k-1}$. For $k = 0, \ldots, n$, we let C_k denote the free $\mathbb{Z}$-module on S_k, and we define $d_k : C_k \to C_{k-1}$ (for $k \geq 1$) by[75]

$$d_k(s) = \sum_{i=0}^{k} (-1)^i f_k^{(i)}(s).$$

We leave it as an exercise to show that the compatibility condition on the $f_k^{(i)}$ in Definition 29.5 ensures that $(C_\bullet, d_\bullet)$ is indeed a chain complex (with $C_k = 0$ unless $0 \leq k \leq n$ and $d_k = 0$ unless $1 \leq k \leq n$). Since $C_\bullet$ is a chain complex, we can define its homology $H_i(C_\bullet)$. If one first sets up the theory of (singular) homology of general topological spaces, then it turns out that the homology of the space associated to a simplicial complex coincides with that of the chain complex we have just constructed. We could *define* the homology of the topological space associated to the simplicial complex to be $H_i(C_\bullet)$, but this would sweep an important fact under the rug, namely that if two topological spaces are homeomorphic, then the associated homology groups (in each degree) are isomorphic.

EXAMPLE 29.8 If we view a simplicial 1-complex as a graph, then the chain complex we just constructed is the same as the one considered in §17, almost. The $\mathbb{Z}$-modules C_i are the same as the M_i already defined, but the sign convention in the definition of d_1 is different. In §17, we defined $d_1(e)$ to be $t_1(e) - t_0(e)$ for $e \in V = S_1$, but recall that we related simplicial 1-complexes to graphs by identifying $t_i(e)$ with $f_1^{(i)}(e)$, so now we have $d_1(e) = t_0(e) - t_1(e)$. This does not however affect the homology, since $\ker(d_1) = \ker(-d_1)$ and $\mathrm{im}(d_1) = \mathrm{im}(-d_1)$, so the conclusions in Examples 17.11–17.14 are also valid with the sign conventions we are using in the more general setting.

EXAMPLE 29.9 Consider the simplicial 2-complex of Example 29.7. The associated chain complex has the form

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow C_2 \xrightarrow{d_2} C_1 \xrightarrow{d_1} C_0 \longrightarrow 0 \longrightarrow \cdots,$$

where

- $C_2 = \mathbb{Z} \cdot a \oplus \mathbb{Z} \cdot b$;
- $C_1 = \mathbb{Z} \cdot e \oplus \mathbb{Z} \cdot f \oplus \mathbb{Z} \cdot g$;
- $C_0 = \mathbb{Z} \cdot u \oplus \mathbb{Z} \cdot v \oplus \mathbb{Z} \cdot w$;
- $d_1(e) = u - v$, $d_1(f) = u - w$ and $d_1(g) = v - w$;
- $d_2(a) = d_2(b) = e - f + g$.

We leave it as an exercise to show that

- $B_0 = \{ku + \ell v + mw \mid k + \ell + m = 0\}$,
- $Z_1 = B_1 = \mathbb{Z} \cdot (e - f + g)$,
- $Z_2 = \mathbb{Z} \cdot (a - b)$,

[75] Again we write simply s (instead of e_s) for the basis element of C_k corresponding to the element $s \in S_k$.

and to deduce that

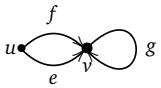
$$H_i(C_\bullet) \cong \begin{cases} \mathbb{Z}, & \text{if } i = 0 \text{ or } 2; \\ 0, & \text{otherwise.} \end{cases}$$

EXAMPLE 29.10 Let $S_0 = \{u, v\}$, $S_1 = \{e, f, g\}$, $S_2 = \{a, b\}$, and define the functions $f_k^{(i)}$ as follows:

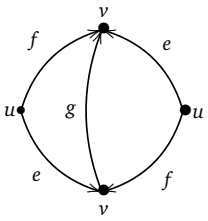
- $f_1^{(0)}(e) = f_1^{(0)}(f) = u$, $f_1^{(0)}(g) = v$;
- $f_1^{(1)}(e) = f_1^{(1)}(f) = f_1^{(1)}(g) = v$;
- $f_2^{(0)}(a) = e$, $f_2^{(0)}(b) = f$;
- $f_2^{(1)}(a) = f$, $f_2^{(1)}(b) = e$;
- $f_2^{(2)}(a) = f_2^{(2)}(b) = g$.

We leave it as an exercise to show that this does indeed define a simplicial 2-complex.

The space X_0 consists of two points, and X_1 can now be pictured as follows:



Again $\widetilde{X}_2$ is a union of two 2-simplices (exactly as in Example 29.7; see (45)), but to visualize how its boundary is glued to X_1, it may be easier to depict X_1 as the quotient of the space represented by the graph:



where we glue the two edges labeled e to each other and similarly for f (with orientations preserved so the two vertices labeled u are identified, and similarly for v).

Now gluing the boundary of the 2-simplex labeled a (in 44) to the triangle on the left, and similarly the one labeled b to the triangle on the right (now with $(0,0,1)$ as u, $(0,1,0)$ as the v at the bottom, and $(1,0,0)$ at the top), we obtain the space X_2. This can be visualized as a hemisphere, but with the identifications made along its boundary (i.e. the equator) as in the description of X_1; more precisely, the resulting space X_2 is homeomorphic to the quotient $H^2/\sim$, where H^2 is (for example) the upper hemisphere $\{(x,y,z) \in S^2 \,|\, z \geq 0\}$ and the equivalence relation is defined by $(x,y,z) \sim (x',y',z')$ if either 1) $(x,y,z) = (x',y',z')$ or 2) $x = -x'$, $y = -y'$ and $z = z' = 0$.

This can also be viewed as the quotient of the sphere S^2 under the obvious action of the group $\{\pm 1\}$ (i.e. $\epsilon \cdot (x,y,z) = (\epsilon x, \epsilon y, \epsilon z)$). Indeed the composite $\varphi : H^2 \to S^2 \to \{\pm 1\}\backslash S^2$ (of the inclusion with the quotient map) is continuous, and $\varphi(\mathbf{x}) = \varphi(\mathbf{x}')$ if and only if $\mathbf{x} \sim \mathbf{x}'$, from which it follows that there is a continuous bijection $X_2 \to S^2$. One can check directly that this bijection also sends open sets to open sets, or use the general fact from topology that a continuous bijection of compact Hausdorff spaces is a homeomorphism. Finally recall that we have seen this space before; it was (the case $n = 2$ of) an exercise to prove that $\{\pm 1\}\backslash S^2$ is homeomorphic to the projective plane $\mathbb{P}^2(\mathbb{R})$ considered in Example 24.30.

We can then proceed as in Example 29.10 to compute the homology $H_i(C_\bullet)$. We now have

- $C_2 = \mathbb{Z} \cdot a \oplus \mathbb{Z} \cdot b$;
- $C_1 = \mathbb{Z} \cdot e \oplus \mathbb{Z} \cdot f \oplus \mathbb{Z} \cdot g$;
- $C_0 = \mathbb{Z} \cdot u \oplus \mathbb{Z} \cdot v$;
- $d_1(e) = d_1(f) = u - v$ and $d_1(g) = 0$;
- $d_2(a) = e - f + g$ and $d_2(b) = f - e + g$.

We leave it as an exercise to show that

- $B_0 = \mathbb{Z} \cdot (u - v)$,
- $Z_1 = \mathbb{Z} \cdot (e - f) \oplus \mathbb{Z} \cdot g$,
- $B_1 = \mathbb{Z} \cdot (e - f + g) \oplus \mathbb{Z} \cdot (f - e + g)$,
- $Z_2 = 0$,

and to deduce that

$$H_i(C_\bullet) \cong \begin{cases} \mathbb{Z}, & \text{if } i = 0; \\ \mathbb{Z}/2\mathbb{Z}, & \text{if } i = 1; \\ 0, & \text{otherwise.} \end{cases}$$