Chapter 3

Modules

3.1 Modules, submodules and homomorphisms

The problem of classifying all rings is much too general to ever hope for an answer. But one of the most important tools available – for general non-commutative rings – is really to focus not on the ring itself, but on the structure of its *module category*. Just as groups act on sets, rings act on modules...

Let R be a ring. A *left R-module* means an Abelian group M together with a multiplication $R \times M \to M$ denoted $(r, m) \mapsto rm$ such that

(M1) $r(m_1 + m_2) = rm_1 + rm_2$;

(M2) $(r_1 + r_2)m = r_1m + r_2m$;

(M3) $(r_1r_2)m = r_1(r_2m)$;

(M4) $1_Rm = m$

for all $r, r_1, r_2 \in R, m, m_1, m_2 \in M$. Strictly speaking, what I am calling a left R-module should be called a *unital* left R-module because I always include axiom (M4). By the way, the axioms imply that $(-r)m = -(rm)$, $0_Rm = 0_M$ and $r0_M = 0_M$ for all $r \in R, m \in M$.

There is another notion called a right R-module. As you can probably guess, this is exactly the same idea, but the operation of $r \in R$ on $m \in M$ is written on the *right*, i.e. the operation is a map $M \times R \to M$ denoted $(m, r) \mapsto mr$. The axioms become

(M1′) $(m_1 + m_2)r = m_1r + m_2r$;

(M2′) $m(r_1 + r_2) = mr_1 + mr_2$;

(M3′) $m(r_1r_2) = (mr_1)r_2$;

(M4′) $m1_R = m$

for all $r, r_1, r_2 \in R, m, m_1, m_2 \in M$.

You need to be somewhat ambidextrous when working with modules. I will try usually to work with left modules – and all the results we prove for left modules have right module analogues. You really cannot avoid the need for right modules from time to time, however. If necessary, we will write ${}_RM$ to emphasize that M is a left R-module, or M_R to emphasize that M is a right R-module.

Now, given any ring R, let R^{op} denote the same Abelian group but with new multiplication $\cdot$ defined by $r \cdot s := sr$, the right hand side being the old multiplication in R. If M is a left

R-module, then we can view M as a right R^{op}-module, by defining the right action of R^{op} on M by $mr := rm$, where the right hand side of this equation is the old left action of R on M. Similarly, any right R-module can be viewed as a left R^{op}-module. This "*op*" trick will occasionally be useful for technical reasons.

In the special case that R is commutative, $R^{op} = R$. So we obtain from the previous paragraph *in the commutative case* the standard way to view any left R-module as a right R-module (or vice versa): if M is a left R-module, define a right action of R on M by $mr := rm$. In view of this, when working with commutative rings, I allow myself to be especially careless and usually just talk about *R-modules* without making the "left" or "right" clear.

Before giving the many examples you already know, let me define *R-submodules*. Given a left R-module M, an R-submodule $N \leq M$ means a sub-Abelian group of M such that $rn \in N$ for all $r \in R, n \in N$. I leave you to formulate the definition for right modules!

As with rings, I use the convention that XY denotes the sub-Abelian group of M generated by $\{xy \mid x \in X, y \in Y\}$, for any subsets $X \subseteq R, Y \subseteq M$. Then, saying that N is an R-submodule of M means simply that $N = RN$. In that case, N is itself a left R-module with the operation being the restriction of the operation on M.

Given any subset $X \subseteq M$ (a left R-module), RX is the *submodule of M generated by X*. So, elements of RX look like $r_1x_1 + \cdots + r_nx_n$ for $n \geq 0, r_i \in R, x_i \in X$. In particular, we say that X *generates* M if $M = RX$. Then, M is a *finitely generated* R-module if M is generated by some finite subset X of M, and is a *cyclic* R-module if M is generated by a single element $x \in M$. In this last case, we have that $M = Rx$ so every element of M looks like rx for some $r \in R$.

Now for first examples:

3.1.1. Any ring R itself is a left R-module, denoted ${}_RR$, the left action just being the multiplication; similarly, R is itself a right R-module, denoted R_R, the right action being just the multiplication. These are called the *left regular* and *right regular* modules, respectively.

Observe that ${}_RR$ is actually a *cyclic* left R-module, because $R = R1_R$.

The left R-submodules of ${}_RR$ are precisely the sub-Abelian groups I of R such that $RI = I$. These were called *left ideals* of R in section 2.1. Similarly, the right R-submodules of R_R are the *right ideals* of R. If R is commutative, left ideals, right ideals and (two-sided) ideals coincide, i.e. the left submodules of ${}_RR$ are the right submodules of R_R. So, in the commutative case we just talk of submodules of R, a.k.a. ideals. For instance, if R is a PID, then all submodules of the regular module R are cyclic.

3.1.2. Now take $R = \mathbb{Z}$. Any Abelian group M is a left (but we henceforth omit "left" since $\mathbb{Z}$ is commutative) $\mathbb{Z}$-module, defining $nm = m + m + \cdots + m$ (n *times*) for $n \in \mathbb{N}, m \in M$. Conversely, given a $\mathbb{Z}$-module, it is in the first place an Abelian group and the $\mathbb{Z}$-module structure is determined uniquely by the Abelian group structure. So: *Abelian groups = $\mathbb{Z}$-modules*. Thus, you can think of the notion of R-module for general R as a generalization of Abelian groups!. The case of Abelian groups is always the *most important case* in the general module theory we will be developing.

3.1.3. Let R be a field F. Then, an F-module (left but we omit it) is exactly the same as a *vector space* over F; F-submodules are the same as vector subspaces. So the notion of R-module also captures the notion of vector spaces over a field.

3.1.4. Let R be any ring. Let $M_n(R)$ be the set of all $n \times n$ matrices over R, itself a ring under matrix addition and multiplication. Let A be any left R-module and consider the space $C_n(A)$ of column vectors of height n with entries in A, viewed as an Abelian group under vector addition. Then, $C_n(A)$ is a left $M_n(R)$-module via multiplication of a matrix by a vector. Similarly, $R_n(A)$, the space of row vectors of width n with entries in A is a right $M_n(R)$-module.

As you might expect, the next job is to introduce homomorphisms of R-modules and discuss the isomorphism theorems. Let M, N be left (or right) R-modules. A *homomorphism* $f : M \to N$ means a morphism of Abelian groups such that $f(rm) = rf(m)$ for all $r \in R, m \in M$.

Remark. Ring theorists tend adopt the convention of writing homomorphisms between left R-modules on the right. So they would write mf instead of $f(m)$, for instance. I'm not going to do this – which occasionally later on we will need to use the "*op*" trick mentioned earlier as a result. You need to be flexible on this issue.

We then obtain categories R-$\mathfrak{mod}$ and $\mathfrak{mod}$-R of all *left R-modules* and all *right R-modules* respectively, morphisms being the R-module homomorphisms as just defined. If $f : M \to N$ is an R-module homomorphism, its kernel and image, defined in the same way as for Abelian groups, are automatically R-submodules of M and N respectively.

Given any R-submodule K of M, the quotient Abelian group M/K becomes an R-module if we define $r(m+K) = rm+K$ for all $r \in R, m \in M$, and this gives us the *quotient R-module* of M by the R-submodule K. The map $\pi : M \to M/K, m \mapsto m+K$ is an R-module homomorphism, the *canonical quotient map*. We have the all important:

Universal property of quotients. *Let $N \leq M$ be an R-submodule of a (left or right) R-module M, $\pi : M \to M/N$ be the canonical quotient map. Given any R-module homomorphism $f : M \to M'$ with $N \subseteq \ker f$, there exists a unique homomorphism $\bar{f} : M/N \to M'$ such that $f = \bar{f} \circ \pi$.*

Now the following results all follow as in the case of Abelian groups:

First isomorphism theorem. *Let $f : M \to M'$ be an R-module homomorphism and $N = \ker f$. Then, f factors through the quotient M/N to induce an isomorphism $\bar{f} : M/N \to \operatorname{im} f$.*

Second isomorphism theorem. *Let $K, L \leq M$ be R-submodules of an R-module M. Then, $K/(K \cap L) \cong (K+L)/L$.*

Third isomorphism theorem. *Let $K \leq L \leq M$ be R-submodules of an R-module M. Then, L/K is an R-submodule of M/K and $M/L \cong (M/K)/(L/K)$.*

We have the lattice isomorphism theorem for submodules. We should first observe that the set of all R-submodules of a fixed left R-module M form a complete lattice: meet is given by taking intersections and join is given by taking sums. Then:

Lattice isomorphism theorem. *Let $f : M \to M'$ be an epimorphism with kernel K. Then, the map $N \mapsto f(N)$ gives an isomorphism between the lattice of R-submodules of M containing K the lattice of R-submodules of M'.*

In general, the structure of R-modules can be very varied. The best possible case is when R is a field (or more generally when R is a *simple ring* discussed later on), in which case the fundamental theorem of vector spaces classifies R-modules up to isomorphism by their dimension.

By definition, a *simple* (or *irreducible*) R-module means a non-zero R-module M having no R-submodules other than M itself and (0). We say that an R-module M *has a composition series* if there is a chain of R-submodules of M

$$M = M_0 > M_1 > \cdots > M_n = (0)$$

such that each consecutive factor M_i/M_{i-1} for $i = 1, \ldots, n$ is a simple R-module. You should compare the definitions just made with the analogous definitions we made when studying groups.

Of course, a general R-module M may or may not have a composition series (see later when we discuss *Artinian modules*). But if it does, we have the analogue of the *Jordan-Hölder theorem* for modules (the proof is exactly the same as the proof we gave for groups). This asserts that two different composition series of a given R-module M have the same length and that the composition factors appearing in the two series are isomorphic (after reordering). Thus the length of a composition series of M, and the set of isomorphism types of the composition factors appearing in the composition series, give important invariants of the module M up to isomorphism.

But even if you are lucky and you can prove that M has a composition series and can in some sense determine the simple composition factors appearing in any such composition series, the precise way the composition factors fit together to form the module M – for instance the precise structure of the lattice of submodules of M – can be very difficult to understand.

3.2 Direct products and direct sums

Let R be a ring. Throughout the section, we will only discuss *left R-modules*, though of course all the definitions and results have right module analogues. We first want to explain that the category R-$\mathfrak{mod}$ is an *additive category* in the sense of section 0.5.

Recall this means first of all that given two left R-modules M, N, the set $\mathrm{Hom}_R(M, N)$ of all R-module homomorphisms from M to N actually has the additional structure of an Abelian group. Indeed, given homomorphisms $f, g : M \to N$, we define their sum $f + g : M \to N$ by $(f + g)(m) = f(m) + g(m)$ for all $m \in M$. This gives the operation on $\mathrm{Hom}_R(M, N)$ making it into an Abelian group. For instance, the zero element of $\mathrm{Hom}_R(M, N)$ is the homomorphism 0 with $0(m) = 0_N$ for all $m \in M$. Composition of homomorphisms distributes over addition.

The category R-$\mathfrak{mod}$ clearly has a *zero object*, namely, the zero module. All that is left for R-$\mathfrak{mod}$ to be an additive category is that every pair of R-modules has both a product and a coproduct. These are defined in exactly the same way as products and coproducts of Abelian groups:

(P) The *product* of two R-modules M_1, M_2 is the Cartesian product $M_1 \times M_2$ as an Abelian group with action of $r \in R$ defined by $r(m_1, m_2) = (rm_1, rm_2)$. The R-module homomorphisms $\pi_i : M_1 \times M_2 \to M_i$ satisfying the universal property of products are just the projections, $\pi_i(m_1, m_2) = m_i$. I always refer to products of R-modules as *direct products*.

(C) The *coproduct* of two R-modules M_1, M_2 is the same R-module $M_1 \times M_2$ as the product (bear in mind the corollary in section 0.5). But for some perverse reason, we denote it by $M_1 \oplus M_2$ in this case and write the element (m_1, m_2) instead as the sum $m_1 + m_2$. The maps $\iota_i : M_i \to M_1 \oplus M_2$ making $M_1 \oplus M_2$ into the coproduct are then just the natural inclusions. Of course, as the notation suggests, we always call $M_1 \oplus M_2$ the *direct sum* instead of the coproduct.

Actually, the category R-$\mathfrak{mod}$ is much richer than being just an additive category, as we shall see. For instance, it actually possesses *arbitrary* products and coproducts (i.e. not just of finite families of objects). So now let M_i $(i \in I)$ be a possibly infinite family of left R-modules. Then their product $\prod_{i\in I} M_i$ is just their Cartesian product, with the action of R being "coordinatewise", together with the natural projections $\pi_i : \prod_{i\in I} M_i \to M_i$. Note I always try to visualize an element of $\prod_{i\in I} M_i$ as an "infinite tuple" $m = (m_i)_{i\in I}$.

Turn now to coproducts for our family M_i $(i \in I)$, which turns out to be the more useful notion in module theory. Then, their coproduct $\coprod_{i\in I} M_i$ is defined to be the R-submodule of $\prod_{i\in I} M_i$ consisting of all tuples $m = (m_i)_{i\in I}$ such that $m_i = 0$ for all but finitely many $i \in I$ (the same as for coproducts of Abelian groups in (0.3.5)). We will write an element of $\coprod_{i\in I} M_i$ not as an infinite tuple, but as a sum $\sum_{i\in I} m_i$, since all but finitely many m_i are zero. In this notation, the obvious inclusions $M_i \hookrightarrow \coprod_{i\in I} M_i$ are precisely the maps appearing in the abstract definition of coproduct. You should of course see that for non-zero modules M_i, $\coprod_{i\in I} M_i = \prod_{i\in I} M_i$ if and only if the index set I is finite. Henceforth, I denote $\coprod_{i\in I} M_i$ as $\bigoplus_{i\in I} M_i$ and call it *direct sum*.

We have now defined *direct sum* of a family of modules M_i $(i \in I)$. It gives a way of building a new module out of a collection of old modules. You can call the direct sum $\bigoplus_{i\in I} M_i$ the *external* direct sum. We now want to know how to recognize when a given module M is actually $\bigoplus_{i\in I} M_i$ for some submodules M_i of M, i.e. when is M an *internal* direct sum? (The distinction is similar to the difference between external and internal semidirect products made in section 1.6.)

So now suppose we are given some R-module M and a collection of R-submodules M_i $(i \in I)$ of M. The R-submodule of M generated by the M_i is just their sum $\sum_{i\in I} M_i$, meaning the set of all elements of M which can be written as $\sum_{i\in I} m_i$ for $m_i \in M_i$ with all but finitely many being zero (so that the possibly infinite sum has meaning). Now we have the notions of *span* and *linear independence*. If

$$M = \sum_{i\in I} M_i$$

then we say that the M_i *span* M. If the property

$$\sum_{i\in I} m_i = 0 \quad \Rightarrow \quad m_i = 0 \ \forall i \in I$$

holds whenever we are given elements $m_i \in M_i$ with all but finitely many being zero, we say that $\sum_{i\in I} M_i$ is *direct*, and the submodules M_i are called *linearly independent*.

The following properties are equivalent:

(1) $\sum_{i\in I} M_i$ is direct;

(2) $M_i \cap \left(\sum_{j\in I-\{i\}} M_j\right) = (0)$ for all $i \in I$;

(3) any $m \in \sum_{i\in I} M_i$ can be written as $\sum_{i\in I} m_i$ for *unique* elements $m_i \in M_i$, all but finitely many being zero.

If the M_i span M, so $\sum_{i\in I} M_i = M$, and they are linearly independent, so $\sum_{i\in I} M_i$ is direct, then we write

$$M = \bigoplus_{i\in I} M_i$$

as say that M is the *internal direct sum* of the submodules M_i. You can check that if M is the internal direct sum of the submodules M_i, then the unique map $\bigoplus_{i\in I} M_i \to M$ induced by the inclusions $M_i \hookrightarrow M$ according to the universal property of coproducts is in fact an isomorphism, so that M is *isomorphic* to the external direct sum of the M_i. This should explain the language.

Here are two basic definitions used when discussing direct sums. An R-module M is called *decomposable* if M is the internal direct sum of two non-zero proper submodules M_1, M_2 of M. Otherwise, M is *indecomposable*. An R-submodule N of an R-module M is called a *summand* of M if there exists another R-submodule C of M such that $M = N \oplus C$. This submodule C is then called a *complement* to N in M.

Example. Consider the $\mathbb{Z}$-module $M = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ (otherwise known as the Klein 4 group!) It has *three* $\mathbb{Z}$-submodules (otherwise known as subgroups!) of order 2, namely $A = \{(0,0),(1,0)\}$, B $= \{(0,0),(0,1)\}$ and $C = \{(0,0),(1,1)\}$. We therefore have that

$$M = A \oplus B = A \oplus C = B \oplus C.$$

Thus the module M is *decomposable*, but there are *many* ways of decomposing it as a direct sum of two non-zero proper submodules. The submodule A of M is a *summand* of M, while both B and C are *complements* to A in M. Thus a summand can in general have *many different* complements. Corresponding to the three subgroups of M of order 2, we obtain three different *composition series* of M, namely

$$M \supset A \supset (0), \qquad M \supset B \supset (0), \qquad M \supset C \supset (0).$$

These are composition series because all factors are isomorphic to $\mathbb{Z}_2$ which is simple. Thus M in this case has exactly *three different composition series*.

Another example. This time consider the $\mathbb{Z}$-module $M = \mathbb{Z}_4 = \{0,1,2,3\}$ (a.k.a. the cyclic group of order 4). It contains a *unique* $\mathbb{Z}$-submodule $A = \{0,2\}$ of order 2 (hence $\cong \mathbb{Z}_2$), and the quotient M/A is also of order 2 (hence $\cong \mathbb{Z}_2$). In this case, A is *not* a summand of M, for it cannot possibly possess a complement. The chain

$$M \supset A \supset (0)$$

is a *composition series* of M in this case, and this is the *unique* composition series of M.

3.3 Short exact sequences

An *exact sequence* of R-modules means a (finte or infinite) sequence

$$\ldots \longrightarrow M_{i-1} \xrightarrow{f_i} M_i \xrightarrow{f_{i+1}} M_{i+1} \longrightarrow \ldots$$

such that $\ker f_{i+1} = \operatorname{im} f_i$ for each i. If one just has that $\ker f_{i+1} \supseteq \operatorname{im} f_i$, i.e. that $f_{i+1} \circ f_i = 0$, then it is called a *complex* of R-modules. A *short exact sequence* is an exact sequence of the form

$$0 \longrightarrow K \xrightarrow{i} M \xrightarrow{\pi} Q \longrightarrow 0.$$

So exactness in this special case means that i is injective, π is surjective and $\operatorname{im} i = \ker \pi$.

Observe that if we have a short exact sequence as above, then M contains the isomorphic copy $i(K)$ of K as a submodule, and the factor module $M/i(K)$ is isomorphic to Q. Conversely, given an R-module M and a submodule K, we define $i : K \hookrightarrow M$ to be the inclusion and let $\pi : M \to M/K = Q$ be the quotient map. Then we obtain a short exact sequence of R-modules as above. In other words, short exact sequences simply give a convenient notation for writing down *extensions* of a module Q by a module K. You should compare with section 1.6 where we discussed extensions of groups!

Split short exact sequences. *Let*

$$0 \longrightarrow K \xrightarrow{i} M \xrightarrow{\pi} Q \longrightarrow 0.$$

be a short exact sequence of R-modules. Then, the following properties are equivalent:

(1) *there exists an R-module homomorphism $\tau : Q \to M$ such that $\pi \circ \tau = \mathrm{id}_Q$;*

(2) *there exists an R-module homomorphism $j : M \to K$ such that $j \circ i = \mathrm{id}_K$;*

(3) *$i(K)$ has a complement in M, i.e. there exists a submodule $Q' \leq M$ with $M = i(K) \oplus Q'$ (note Q' is obviously isomorphic to Q).*

Proof. (1)⇒(3). Let $Q' = \tau(Q)$. Since $\pi \circ \tau = \mathrm{id}_Q$, τ is injective, hence an isomorphism between Q and its image Q'. We now claim that $M = i(K) \oplus Q'$. Take $m \in M$. Then, $m - \tau(\pi(m))$ is in the kernel of π hence the image of i. So, $m - \tau(\pi(m)) = i(k)$ for some $k \in K$, i.e. $m = i(k) + \tau(\pi(m))$ showing $M = i(K) + Q'$. Moreover, if $m \in i(K) \cap Q'$ then $\pi(m) = 0$ and $m = \tau(q)$ some $q \in Q$. Then, $0 = \pi(\tau(q)) = q$ so $m = \tau(q) = 0$. This shows $i(K) + Q'$ is direct.

(3)⇒(2). We have that $M = i(K) \oplus Q'$. Define $\bar{j} : M \to i(K)$ to be the projection along this direct sum. Set $j = i^{-1} \circ \bar{j}$, where $i^{-1} : i(K) \to K$ is the isomorphism. Then it is immediate that $j \circ i = \mathrm{id}_K$.

(2)⇒(1). Set $Q' = \ker j$. Then, $M = i(K) \oplus Q'$, which is proved in an entirely similar way to the proof of (1)⇒(3) above. Then, the restriction of π to Q' is an isomorphism between Q' and Q; let τ be its inverse, giving an isomorphism $\tau : Q \to Q'$. This does the job. □

We call a short exact sequence in which any of the equivalent conditions of the lemma hold a *split short exact sequence*. The maps τ and j in the lemma are called *splittings*, of π and i respectively. Note if we have a split short exact sequence, then $M \cong Q \oplus K$, and the extension of Q by K encoded by the sequence is called a *split extension* of Q by K.

So, the split extensions of R-modules are exactly *direct sums*. You should compare this with section 1.6: there, the split extensions of groups were *semidirect products*. Of course, for *Abelian groups*, all semidirect products are just ordinary direct products so the two theories coincide for Abelian groups (= $\mathbb{Z}$-modules).

3.4 Semisimple modules

We have already mentioned briefly that an R-module M is called *simple* or *irreducible* if it is non-zero and has no submodules other that M itself and the zero submodule (0). We call M *semisimple*

if M can be decomposed as a direct sum of simple R-modules. Note the zero module counts as a semisimple modules (being the direct sum of zero simple R-modules).

3.4.1. **Lemma.** *Let M be an R-module with $M = \sum_{i\in I} S_i$ where the S_i are simple submodules of M. If N is any submodule of M, there exists a subset $J \subseteq I$ such that*

$$M = N \oplus \bigoplus_{j\in J} S_j.$$

In particular (taking $N = (0)$) $M = \bigoplus_{k\in K} S_k$ for some subset $K \subseteq I$.

Proof. Let $\mathscr{F}$ be the collection of all subsets J of I such that $N + \sum_{j\in J} S_j$ is direct. To apply Zorn's lemma, we check that every chain $(J_\omega)_{\omega\in\Omega}$ in $\mathscr{F}$ has an upper bound. Indeed, consider $J = \bigcup_{\omega\in\Omega} J_\omega$; we need to verify that $J \in \mathscr{F}$. Well, if $N + \sum_{j\in J} S_j$ is not direct, then there exists $n \in N$ and non-zero $s_{j_k} \in S_{j_k}$ for $j_1, \ldots, j_r \in J$ such that $n + s_{j_1} + \cdots + s_{j_r} = 0$. But then for sufficiently large $\omega \in \Omega$, all j_k lie in J_ω so that $N + \sum_{j\in J_\omega} S_j$ is not direct either, a contradiction.

So by Zorn's lemma, we can pick a maximal element $J \in \mathscr{F}$. For this J, $N + \sum_{j\in J} S_j$ is direct. It just remains to show that $P = N + \sum_{j\in J} S_j$ is actually equal to M. Well, if not, we can find $i \in I$ such that $P + S_i$ is strictly larger than P. So, $P \cap S_i \neq S_i$, hence $P \cap S_i = (0)$ as S_i is simple. So $P + S_i$ is direct, which contradicts the maximality of J. □

Recall that a submodule N of an R-module M has a *complement* if there is an R-submodule C of M such that $M = N \oplus C$.

Characterization of semisimple modules. *For an R-module M, the following are equivalent:*

(1) *M is semisimple;*
(2) *M is a sum of simple modules;*
(3) *every R-submodule of M is complemented;*
(4) *every short exact sequence*

$$0 \longrightarrow K \longrightarrow M \longrightarrow Q \longrightarrow 0$$

with M in the middle is split.

Proof. (1)⇒(2). Obvious.

(2)⇒(3). Follows by Lemma 3.4.1.

(3)⇒(1). Let S be the sum of all the simple submodules of M. We claim that $S = M$. Well, if not, then by assumption, $M = S \oplus T$ for some non-zero submodule T. Let N be a non-zero cyclic submodule of T, and choose (by last week's homework problem) a maximal submodule N' of N. Then, N' has a complement in M by assumption, say $M = N' \oplus U$. Then,

$$N = N' \oplus (U \cap N)$$

so N' has a complement in N. Now, $U \cap N \cong N/N'$ is simple so lies in S. But by construction, $U \cap N$ lies in T so cannot lie in S. This gives the contradiction needed.

(3)⇔(4). This is immediate from the definition (3) of split short exact sequence in section 3.3. □

There is a basic but very important warning to be made at this point. Suppose $M = \bigoplus_{i\in I} S_i$ is a semisimple R-module and $N \leq M$ is a submodule. By Lemma 3.4.1, $M = N \oplus \bigoplus_{j\in J} S_j$ for some subset $J \subseteq I$. Hence, $N \cong M/\bigoplus_{j\in J} S_j$ so

$$N \cong \bigoplus_{i\in I-J} S_i.$$

But you will *not* in general be able to write

$$N = \bigoplus_{i\in I-J} S_i$$

here. You should appreciate the subtle but fundamental difference between the preceeding two formulae! Note the argument just given showed in particular that any submodule of a semisimple module is itself semisimple. It is also obvious from (2) of the characterization that every quotient module of a semisimple module is semisimple. So:

3.4.2. **Corollary.** *Every submodule and every quotient module of a semisimple module is itself semisimple.*

We conclude this discussion of semisimple modules with an almost trivial, but nevertheless critical, example of semisimple modules: Suppose that D is a division ring. Then, ${}_DD$ is a simple left D-module (because obviously D has no non-zero proper left ideals). It follows immediately that all non-zero cyclic D-modules are simple. Now let M be any left D-module. Then, for each $0 \neq m \in M$, the cyclic submodule Dm is simple. Since $M = \sum_{0 \neq m \in M} Dm$, we get from the characterization that M is semisimple. Thus:

All modules over a division ring are semisimple.

Of course the fundamental property of vector spaces – that every vector space has a basis – is a special case of this argument.

3.5 Free modules

Let R be a ring, and continue to work only with *left* R-modules. We next introduce the notion of *free R-module*. You should compare the definitions with that of *free groups* (section 1.10), but also keep in mind that (aside from semisimple modules) *free modules are the next best thing to vector spaces*!

Let F be a (left) R-module and X be a set. Like in any category, a free R-module on X is an R-module F together with a map $i : X \to F$ (which we'll see from the construction below is necessarily injective, hence often just think of X as a subset of R) such that

(FM) given any other R-module M and a set map $f : X \to M$, there exists a unique R-module homomorphism $\bar{f} : F \to M$ such that $\bar{f}(x) = f(x)$ for all $x \in X$.

If there exists an R-module F that is free on the set X, then certainly by the usual argument F is unique up to a canonical isomorphism. So we just talk about *the* free module on X. Such a module always exists because:

Existence of free modules. *For any set X, there is a left R-module that is free on X.*

Proof. For $x \in X$, let R_x be a copy of the left regular R-module ${}_RR$, denoting the element in R_x corresponding to the 1_R by x. Consider

$$F = \bigoplus_{x \in X} R_x$$

and view X as a subset of F in the obvious way. We claim that F is free on X. Take a set map $f : X \to M$ to an R-module M. Now, every element of $R_x \cong R$ can written as rx for a unique $r \in R$. Hence, every element of F looks like $\sum_{x \in X} r_x x$ for uniquely determined coefficients $r_x \in R$, all but finitely many r_x's being zero. So if we are trying to extend f to an R-module homomorphism, there is no option but to define

$$\bar{f}\left(\sum_{x \in X} r_x x\right) = \sum_{x \in X} r_x f(x).$$

Moreover, as you easily check, this equation really does define an R-module homomorphism. □

As I said before, in view of this result, we will call F simply *the* free R-module on X, and identify X with a subset of R.

By definition, a subset X of an R-module M is called a *basis* of M if X is *linearly independent*, meaning

$$\sum_{x\in X} r_x x = 0 \quad \Rightarrow \quad r_x = 0 \ \forall x \in X$$

whenever $r_x \in R$ are coefficients with all but finitely many being zero, and moreover X *spans* or *generates* M, meaning that $M = RX$. Observe X is a basis for the free R-module F on X.

You should be able to show conversely (copying the proof of the existence theorem) that if F is any R-module such that $X \subseteq F$ is a basis, then F is free on X. In other words, an R-module M is *the free R-module on the subset $X \subseteq M$* if and only if *X is a basis for M*. Thus, the free R-modules are exactly the R-modules possessing a basis. In that case, the module is isomorphic to a direct sum of copies of the regular R-module ${}_RR$, the number of such copies that suffices being the cardinality of that basis of M. But be careful: it is perfectly possible for am R-module M to have two different bases X and Y having different cardinality. Consider "invariant basis number" property

(IBN) for every free left (or equivalently right – but this is not obvious) R-module F, every basis of F has the same cardinality.

For rings satisfying IBN, you can define the *rank* of a free R-module F the cardinality of any basis of F. For example, all fields, or more generally all division algebras, have IBN. Of course the rank of a vector space is usually called dimension... There are plenty of other rings with IBN, but also rings in which it does not hold.

In view of the theorem, if R is a commutative ring, we can define the notion of *rank* of a free R-module, namely, the cardinality of any basis. Thus, if R is a field, rank is exactly what is more usually called dimension. For more general rings than commutative rings, there may or may not be a well-defined notion of rank of free modules, depending on which ring you are talking about. About the only thing we can say about the cardinality of a basis of a free module in full generality is:

3.5.1. **Lemma.** *Let F be a free R-module. The following are equivalent:*
(1) *F is a finitely generated R-module;*
(2) *F has a finite basis;*
(3) *every basis of F is finite.*

Proof. Clearly if F has a finite basis, it is finitely generated. Conversely, suppose F is generated by finitely many elements $f_1, \dots, f_n$ and let X be any basis for F. We can write $f_i = \sum_{x\in X} a_{i,x}x$ for each $i = 1, \dots, n$, where all but finitely many of the coefficients $a_{i,x} \in R$ are zero for each fixed i. But then $X' = \{x \in X \mid a_{i,x} \neq 0 \text{ for some } i = 1, \dots, n\}$ is finite and every generator $f_1, \dots, f_n$ lies in RX'. Hence, $M = RX'$, so since X is linearly independent we must actually have that $X = X'$. This shows that X is finite. □

We now want to prove that all commutative rings have IBN.

3.5.2. **Lemma.** *Let R and S be two rings and $\pi : R \twoheadrightarrow S$ be an epimorphism. If S has IBN then so does R.*

Proof. Recall that if M is an R-module, M/IM is an R/I-module in a canonical way. Let X be a basis for M. So

$$M = \bigoplus_{x\in X} Rx$$

hence

$$M/IM \cong \bigoplus_{x\in X} Rx/Ix \cong \bigoplus_{x\in X} S\bar{x}.$$

Since S has IBN, it makes sense to talk about the rank of M/IM. We've just shown it is $|X|$ for any basis X for M. Hence any two bases of M have the same cardinality. □

3.5.3. **Corollary.** *Let R be a commutative ring and M be a free R-module. Then, any two bases of M have the same cardinality.*

Proof. Let I be a maximal ideal of R, so that $F = R/I$ is a field. By linear algebra, R/I has IBN. Hence so does R by the lemma. □

Free R-modules are extremely important, mainly because:

3.5.4. **Theorem.** *Every (finitely generated) R-module M is the quotient of a (finitely generated) free module.*

Proof. Let F be the free module on the set M. The set map $M \to M$ extends to a unique R-module homomorphism $F \to M$ which is clearly surjective. Thus, M is a quotient of F. In case M is finitely generated, the argument is the same, but one takes F to be the free module on some finite generating set of M instead. □

It is worth pointing out a special case of this argument. Recall an R-module M is *cyclic* if $M = Rx$ for some $x \in M$. In that case, observing that ${}_RR$ is free on 1_R, the set map $1_R \mapsto x$ extends to a unique R-module homomorphism $f : R \to M, r \mapsto rx$. It is surjective since $M = Rx$. Hence, $M \cong R/\ker f$. This shows: *any cyclic R-module is a quotient of the regular module ${}_RR$ by a left ideal.*

Finally in this section, we record an important property of free modules:

Projective property of free modules. *Every short exact sequence $0 \to K \to M \to F \to 0$ with F free is split.*

Proof. Let X be a basis for F. For each $x \in X$, we can find a pre-image $m_x \in M$ such that $\pi(m_x) = x$, since π is surjective. We obtain a set map $h : X \to M, x \mapsto m_x$. By the universal property of free modules, this extends to a unique R-module homomorphism $\sigma : F \to M$ such that $\sigma(x) = m_x$ for each $x \in X$. So, $\pi \circ \sigma : F \to F$ is an R-module homomorphism such that $(\pi \circ \sigma)(x) = x$ for all $x \in X$. But X generates F, to the only such map is the identity map, hence $\pi \circ \sigma = \mathrm{id}_F$. □

3.6 The Krull-Schmidt theorem

Let R and S be rings. An (R, S)-*bimodule* M means an abelian group M with the structure both of a left R-module and of a right S-module, such that

$$(rm)s = r(ms)$$

for all $r \in R, s \in S$ and $m \in M$, i.e. the left action of R commutes with the right action of S. Often we write ${}_RM_S$ to remind us that M is an (R, S)-bimodule. For example, we have the regular (R, R)-bimodule, namely, the ring R itself, denoted ${}_RR_R$ if we're thinking of it as a bimodule.

By the way, as much as possible, I try to make this the only place I ever need to think about right modules... recall usually (except when discussing bimodules) that I can always eliminate right modules, since the category $\mathfrak{mod}$-R of right modules over a ring R is isomorphic to the category of R^{op}-$\mathfrak{mod}$ of left modules over the opposite ring R^{op}.

There is an obvious notion of sub-(R, S)-bimodule: a sub-abelian group that is both an R-submodule and an S-submodule. There is an obvious notion of morphism of (R, S)-bimodules: a map that is both a left R-module homomorphism and a right S-module homomorphism. We write R-$\mathfrak{mod}$-S for the category of all (R, S)-bimodules. I'm not going to pursue any other generalities

about bimodules yet, but we will meet them again in a more essential way when we discuss tensor products...

Here is how bimodules arise naturally for us right now. Let M be a left R-module. Consider the abelian group $hom_R(M, M)$. We usually denote it by $\mathrm{End}_R(M)$ for short. Of course it is actually a ring, since it makes sense to compose two such endomorphisms. Note I always write maps on the left. So M is itself a left $\mathrm{End}_R(M)$-module. Hence M is a right $\mathrm{End}_R(M)^{op}$-module. Clearly this action commutes with the left action of R, so we've just made M into an $(R, \mathrm{End}_R(M)^{op})$-bimodule.

Now we need to think about the endomorphism ring of a finitely generated free left R-module. Say

$$M = \bigoplus_{i=1}^{n} Rx_i,$$

i.e. $x_1, \ldots, x_n$ is a basis for M. Any endomorphism $f : M \to M$ is completely determined by the images of $x_1, \ldots, x_n$, and since M is free, these images can be any element of M. Hence there is a bijection between endomorphsms of M and $n \times n$ matrices with entries in R, the endomorphism f corresponding to the matrix $(a_{i,j})_{1 \leq i,j \leq n}$ defined from

$$f(x_i) = \sum_{j=1}^{n} a_{i,j} x_j.$$

Recall that $M_n(R)$ denotes the ring of all $n \times n$ matrices over R.

3.6.1. **Lemma.** *The map just described sending $f : M \to M$ to the matrix $(a_{i,j})_{1 \leq i,j \leq n}$ is a ring isomorphism between $\mathrm{End}_R(M)^{op}$ and $M_n(R)$.*

Proof. We've already established that the map is a bijection. We need to check that it is a ring homomorphism. Take f, g corresponding to matrices $(a_{i,j})_{1 \leq i,j \leq n}$ and $(b_{i,j})_{1 \leq i,j \leq n}$. Consider the product fg in the ring $\mathrm{End}_R(M)^{op}$. This is the endomorphism $g \circ f$ (because of the definition of the opposite ring...). So it corresponds to the matrix $(c_{i,j})_{1 \leq i,j \leq n}$ defined from

$$g(f(x_i)) = \sum_{j=1}^{n} c_{i,j} x_j.$$

But

$$g(f(x_i)) = g(\sum_{k=1}^{n} a_{i,k} x_k) = \sum_{k=1}^{n} a_{i,k} g(x_k) = \sum_{k,j=1}^{n} a_{i,k} b_{k,j} x_j.$$

Hence,

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

This is the matrix product... □

This might look a bit weird. Its the price I pay for working with left modules and writing maps on the left too. But you should think of elements of a free left module as row vectors, then the opposite of endomorphism ring is the ring of $n \times n$ matrices acting on the right on row vectors by matrix multiplication... It is instructive to work out for your self what happens if you work with right modules here. Then you don't need any op when you consider the endomorphism ring of a free right module. Indeed, you should think of elements of a free right module as column vectors, then the endomorphism ring is the ring of $n \times n$ matrices acting on the left on column vectors by matrix multiplication. This is what you are used to in linear algebra (when since fields are commutative left and right mean the same thing...)

By the way, the case that M is free of rank one is already interesting here. In that case, the lemma says that $\mathrm{End}_R({}_R R)^{op}$ is isomorphic to R. The isomorphism explicitly maps an endomorphism $f : R \to R$ to the element $f(1_R)$ of R.

Now we are going to spend a while discussing idempotents. Recall an idempotent $e \in R$ means an element with $e^2 = e$. A family $(e_i)_{i\in I}$ of idempotents is called *orthogonal* if $e_i e_j = e_j e_i = 0$ for all $i \neq j$. An idempotent e is called *primitive* if it is non-zero and it cannot be written as $e = e_1 + e_2$ with e_1, e_2 non-zero orthogonal idempotents. Note if e is an idempotent, then so is $(1-e)$, and e is orthogonal to $(1-e)$. Idempotents arise naturally when we consider endomorphism rings...

3.6.2. **Lemma.** *Let $M_1, \ldots, M_n$ be submodules of a left R-module M. Let $S = \mathrm{End}_R(M)^{op}$. Then,*

$$M = M_1 \oplus \cdots \oplus M_n$$

if and only if there exist orthogonal idempotents $e_1, \ldots, e_n \in S$ with $1_S = e_1 + \cdots + e_n$ and $M_i = Me_i$.

Proof. If $M = M_1 \oplus \cdots \oplus M_n$, let e_i be the composite of the projection of M onto M_i along the direct sum followed by the inclusion of M_i into M. Obviously $e_i^2 = e_i$, the e_i's are orthogonal idempotents with $M_i = Me_i$, and $1_S = e_1 + \cdots + e_n$.

Conversely, given such idempotents, we show that $M = M_1 \oplus \cdots \oplus M_n$. Well, take $m \in M$. Then

$$m = m(e_1 + \cdots + e_n) = me_1 + \cdots + me_n \in M_1 + \cdots + M_n,$$

hence the M_i span M. To show that they are linearly independent, suppose that

$$m_1 + \cdots + m_n = 0$$

for $m_i \in M_i$. Applying e_i, which kills all m_j for $j \neq i$, we get that $m_i e_i = m_i = 0$ for each i. Done. □

3.6.3. **Corollary.** *A submodule N of M is a direct summand with complement C if and only if there exists an idempotent $e \in S = \mathrm{End}_R(M)^{op}$ with $N = Me$ and $C = M(1-e)$.*

3.6.4. **Corollary.** *A non-zero R-module M is indecomposable if and only if 1 is a primitive idempotent in $S = \mathrm{End}_R(M)^{op}$.*

We can generalize this a little bit if we use the following handy theorem:

3.6.5. **Lemma.** *Let $0 \neq e \in S = \mathrm{End}_R(M)^{op}$ be an idempotent. Then, $\mathrm{End}_R(Me)^{op} = eSe$.*

Proof. (Recall that eSe is a subring of S but it is not a unital subring: its identity element is e.) Right multiplication by elements of eSe certainly give endomorphisms of Me. So we have an injective homomorphism $eSe \hookrightarrow \mathrm{End}_R(Me)^{op}$. Conversely, suppose f is an endomorphism of Me. Extend it to an endomorphism $\bar{f}$ of all of M by letting $\bar{f}$ be zero on $M(1-e)$. Then $\bar{f} = e\bar{f}e$ so it is an element of eSe. □

3.6.6. **Corollary.** *Me is indecomposable if and only if e is a primitive idempotent in S.*

Proof. It follows from above that Me is indecomposable if and only if e is a primitive idempotent in eSe. So we need to show e is primitive in eSe if and only if e is primitive in S. One direction is obvious. So suppose that e is not primitive in S. Then we can write $e = e_1 + e_2$ for e_1, e_2 non-zero orthogonal idempotents. Note that $e(1-e) = e_1(1-e) + e_2(1-e) = 0$ in S. Multiplying on the left by e_1 you get that $e_1(1-e) = 0$. Hence $e_1 e = e_1$. Similarly $ee_1 = e_1$. Hence $e_1 \in eSe$. Similarly $e_2 \in eSe$. Hence e is not primitive in eSe. □

Here is the most important example to keep in mind: suppose that R is the ring of $n \times n$ matrices over a field F. Let e_i be the matrix with a 1 in its ii-entry, zeros elsewhere. Then $1 = e_1 + \cdots + e_n$ and the e_i's are orthogonal idempotents. I claim that they are in fact primitive idempotents. This follows by the corollary if we can show that Re_i is indecomposable. But its clearly isomorphic to

the left R-module F^n of column vectors, on which R acts by matrix multiplication. That is an irreducible R-module, because you can map any non-zero vector to any other by multiplying by a matrix. So it is certainly indecomposable.

Now we are nearly done with the preparations for the Krull-Schmidt theorem. This is concerned only with modules of finite length, i.e. modules having a composition series. First a cool lemma.

Fitting's Lemma. *Let ${}_RM$ be an R-module of finite length n. Let $f \in \mathrm{End}_R(M)$ be an endomorphism. Then,*

$$M = \mathrm{im}\, f^n \oplus \ker f^n.$$

Proof. Clearly $\mathrm{im}\, f \geq \mathrm{im}\, f^2 \geq cdots \geq \mathrm{im}\, f^n$. Also as soon as $\mathrm{im}\, f^i = \mathrm{im}\, f^{i+1}$, its equal from then on. So since M has finite length n, we certainly have that $\mathrm{im}\, f^n = \mathrm{im}\, f^{n+1} = \cdots$. Hence $\mathrm{im}\, f^n = \mathrm{im}\, f^{2n}$. Now let $x \in M$. Then, $f^n(x) = f^{2n}(y)$ for some y. Hence,

$$x = f^n(y) + (x - f^n(y)).$$

But $f^n(x - f^n(y)) = f^n(x) - f^{2n}(y) = 0$. So we've written x as a sum of something in $\mathrm{im}\, f^n$ and something in $\ker f^n$. So $M = \mathrm{im}\, f^n + \ker f^n$.

Instead, we have that $\ker f \leq \ker f^2 \leq \cdots \leq \ker f^n$. Since M has finite length n, we get that $\ker f^n = \ker f^{n+1} = \cdots = \ker f^{2n}$ for sure. Now take $x \in \mathrm{im}\, f^n \cap \ker f^n$. Then $x = f^n(y)$ for some y, and $f^n(x) = f^{2n}(y) = 0$. So $y \in \ker f^{2n} = \ker f^n$, so $f^n(y) = x = 0$ already. Hence $\mathrm{im}\, f^n \cap \ker f^n = (0)$. □

3.6.7. **Corollary.** *Let M be indecomposable of finite length and $f \in \mathrm{End}_R(M)$. The following are equivalent:*

(i) f is a monomorphism;

(ii) f is an epimorphism;

(iii) f is an automorphism;

(iv) f is not nilpotent (i.e. no power of f is zero).

You probably recall from the homework exercises that a commutative ring is *local* if it has a unique maximal ideal. Equivalently, the set of non-units in R is closed under addition. This is the right way to generalize the notion of local ring to the non-commutative case. So, define a (not necessarily commutative) ring R to be *local* if the sum of any two non-units in R is again a non-unit in R. These arise naturally as endomorphism rings of indecomposable modules of finite length:

3.6.8. **Corollary.** *Let M be indecomposable of finite length. Then $\mathrm{End}_R(M)$ is a local ring.*

Proof. Take $f, g \in \mathrm{End}_R(M)$ with $f + g$ a unit and g not a unit. We need to show that f is a unit. Well, $(f + g) \circ h = 1$ for some h, so $f \circ h = 1 - g \circ h$. Since g is not invertible, it is not onto, so $g \circ h$ is not onto. Hence by the previous corollory, $g \circ h$ is nilpotent, so $(g \circ h)^n = 0$ for some n. But then

$$f \circ h(1 + (g \circ h) + \cdots + (g \circ h)^{n-1}) = (1 - g \circ h)(1 + (g \circ h) + \cdots + (g \circ h)^{n-1}) = 1.$$

Hence f is onto, having a right inverse. Hence f is an automorphism by the preceeding corollary, i.e. its a unit. □

At last we can prove the Krull-Schmidt theorem: modules of finite length have a unique up to isomorphism decomposition as a direct sum of finitely many indecomposables. This should be compared with the Jordan-H older theorem which is about unique decompositions into irreducibles...

The Krull-Schmidt theorem. *Let M be a module of finite length. Then M has a decomposition as a finite direct sum of indecomposable submodules. Moreover, given two such decompositions*

$$M = M_1 \oplus \cdots \oplus M_m = N_1 \oplus \cdots \oplus N_n,$$

we have that $n = m$ and there exists a permutation $\sigma \in S_n$ such that

(i) $M_{\sigma(k)} \cong N_k$;

(ii) $M = M_{\sigma(1)} \oplus \cdots \oplus M_{\sigma(k)} \oplus N_{k+1} \oplus \cdots \oplus N_n$

for every $k = 1, \ldots, n$.

Proof. The existence of a decomposition into finitely many indecomposables is easy: if $M = M_1 \oplus \cdots \oplus M_n$ for non-zero submodules $M_1, \ldots, M_n$, then n is bounded above by the composition length of M. It follows that a decomposition of M as a direct sum of submodules of maximal length exists. The submodules are then necessarily indecomposable.

So we need to prove the uniqueness. Suppose that $M = M_1 \oplus \cdots \oplus M_m = N_1 \oplus \cdots \oplus N_n$. Let $S = \mathrm{End}_R(M)^{op}$. Corresponding to these decompostions into indecomposables, we have decompositions $1 = e_1 + \cdots + e_m = f_1 + \cdots + f_n$ of $1 \in S$ as sums of orthogonal primitive idempotents, with $M_i = Me_i$ and $N_j = Mf_j$. The ring $f_1 S f_1 = \mathrm{End}_R(N_1)^{op}$ is a local ring with identity f_1. Since

$$f_1 = f_1(e_1 + \cdots + e_m)f_1 = f_1 e_1 f_1 + \cdots + f_1 e_m f_1$$

it follows that there exists $1 \leq \sigma(1) \leq m$ such that $f_1 e_{\sigma(1)} f_1$ is a unit in $f_1 S f_1$.

Consider the composite

$$Mf_1 \overset{e_{\sigma(1)}}{\to} Me_{\sigma(1)} \overset{f_1}{\to} Mf_1$$

where the first map is right multiplication by $e_{\sigma(1)}$ and the second map is right multiplication by f_1. The composite is right multiplication by $f_1 e_{\sigma(1)} f_1$, so it is an isomorphism. Hence the first map is injective and the second one is surjective. I claim moreover that the second map is a split surjection: a splitting is given by right multiplication by $(f_1 e_{\sigma(1)} f_1)^{-1} f_1 e_{\sigma(1)}$. Hence since $Me_{\sigma(1)}$ is indecomposable, both maps are actually isomorphisms. Given this, it is easy to show that

$$M = Me_{\sigma(1)} \oplus Mf_2 \oplus \cdots \oplus Mf_n.$$

Now pass to the quotient module $M/Me_{\sigma(1)}$ and induct... □

Note a special case of the theorem is that any two bases for a finitely generated module over a division ring have the same number of elements, i.e. division rings have IBN. In that case the proof we've just given is exactly the *exchane lemma* which maybe you saw in linear algebra once upon a time!

3.7 Finitely generated modules over PIDs

We have discussed free modules and also the rank of a free module over a commutative ring. Basing your intuition on vector spaces, it is natural to ask questions like

- Are all submodules of a free module F free?
- For commutative rings, is the rank of a free submodule of F less than or equal to the rank of F?

Unfortunately, the answer to both questions is in general *no* (so the analogy with vector spaces is not that good).

For the remainder of the section, R will denote a PID. In this case only, we can develop a reasonable theory of submodules of free modules.

3.7.1. **Theorem.** *Let F be a finitely generated free module over a PID R, and $N \leq F$ be an R-submodule. Then, F is also free and* $\operatorname{rank} N \leq \operatorname{rank} F$.
(Actually, this theorem is true even if F is not finitely generated.)

Proof. Let $x_1, \ldots, x_n$ be a basis for F and $N \leq F$. Set $F_0 = (0)$, $F_i = Rx_1 + \cdots + Rx_i$ and $N_i = N \cap F_i$. We prove by induction on $i = 0, 1, \ldots, n$ that $N_i \leq F_i$ is free of rank $\leq i$, the base case $i = 0$ being trivial.

Now let $i \geq 1$ and consider the induction step. If $N_i = N_{i-1}$, there is nothing to prove. Otherwise, let

$$A = \{a \in R \mid \text{there exists } x \in F_{i-1} \text{ such that } x + ax_i \in N_i\}.$$

Then, A is an ideal of R, so $A = (b)$ for some $b \in R$ as R is a PID. Moreover, $b \neq 0$ as $N_i \neq N_{i-1}$. As $b \in A$, there is some $y \in F_{i-1}$ such that $y + bx_i \in N_i$. Let $z = y + bx_i$. We claim $N_i = N_{i-1} \oplus Rz$, so that N_i is free of rank one more than N_{i-1}.

Well, take any $f \in N_i$. Then, $f = x + cx_i$ for some $x \in F_{i-1}, c \in R$. So $c \in A = (b)$, so $c = bd$ for some $d \in R$ and $f = x + dbx_i$. But then, $f - dz = x - dy \in F_{i-1} \cap N_i = N_{i-1}$. This shows that $N_i = N_{i-1} + Rz$. Finally, to show that $N_{i-1} \cap Rz = (0)$, suppose $rz \in N_{i-1}$. Then, $rz = ry + rbx_i \in F_{i-1}$, hence $rbx_i \in F_{i-1} \cap Rx_i = (0)$, so $rb = 0$ whence $r = 0$ as R is an integral domain. □

In order to obtain more precise information about submodules of free modules over PIDs, we first need to discuss a rather different topic. Let $M_{s,t}(R)$ denote the set of all $s \times t$ matrices with entries in the PID R. Call two matrices $A, B \in M_{s,t}(R)$ *equivalent* if there exist invertible square matrices P and Q such that $B = PAQ$. Note that "equivalence" is indeed an equivalence relation on $M_{s,t}(R)$. I will assume you are familiar with basic notions of matrices (over an arbitrary commutative ring). For instance, a square matrix with entries in R is invertible if and only if its determinant (defined by "Laplace expansion" along some row or column) is a unit in R.

The main job now is to prove:

Canonical form for matrices over PIDs. *If R is a PID then any matrix $A \in M_{s,t}(R)$ is equivalent to a matrix of the form* $\operatorname{diag}(d_1, \ldots, d_u)$ *(where $u = \min(s, t)$) with $d_1 | d_2 | \ldots | d_u$ in R. Moreover, the diagonal entries $d_1, \ldots, d_u$ are unique up to associates.*

Proof. For the proof, we need the *elementary row and column operations*. Let me remind you of these, for a matrix $A \in M_{s,t}(R)$:

(O1) Swap two rows (or columns) of A;

(O2) Scale any row (or column) of A by a unit in R;

(O3) Add a multiple of one row (or column) to another row (or column).

The point is that all of the elementary row and column operations can be performed on the matrix A by pre- or post-multiplying by an invertible matrix. Thus, they do not change the equivalence class of the matrix we are considering.

Now let me explain the *algorithm* to reduce an arbitrary matrix $A \in M_{r,s}(R)$ to the desired canonical form. To guarantee that the algorithm eventually terminates, we introduce the notion of *length* of the matrix A. This is defined to be the the number of primes appearing in the prime factorization of the leading entry $a_{1,1}$ of a, or 0 in case $a_{1,1}$ is zero or a unit.

Now, if $A = 0$, there is nothing to do. Else, some entry of the matrix A is non-zero, and swapping rows and columns, we can assume that $a_{1,1}$ is a non-zero entry of A of minimal length. Now there are three cases:

Case one. $a_{1,1} \nmid a_{1,j}$ for some $j > 1$. Without loss of generality, assume $a_{1,1} \nmid a_{1,2}$. Let d be a GCD of $a_{1,1}$ and $a_{1,2}$, so $a_{1,1} = dy_1, a_{1,2} = dy_2$ for y_1, y_2 coprime. Since $a_{1,1} \nmid a_{1,2}$, y_1 is not a

unit, so $\lambda(y_1) \geq 1$, so $\lambda(d) < \lambda(a_{1,1})$. Write $1 = x_1y_1 + x_2y_2$ for $x_1, x_2 \in R$. Then

$$Q = \left[\begin{array}{cc|ccc} x_1 & -y_2 & 0 & \dots & 0 \\ x_2 & y_1 & 0 & \dots & 0 \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & I & \\ 0 & 0 & & & \end{array}\right]$$

is invertible and AQ has leading term d. Since $\lambda(d) < \lambda(a_{1,1})$, the length of A has gone down, and we can now repeat the algorithm from the beginning.

Case two. $a_{1,1} \neq 0$ and $a_{1,1} \nmid a_{i,1}$ for some $i > 1$. This is proceeds in the same way as case one, working with columns not rows.

Case three. $a_{1,1}$ divides every entry in both the first row and the first column of A. Then, we can use elementary row and column operations to reduce the matrix A to the form

$$\left[\begin{array}{c|ccc} a_{1,1} & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right]$$

Now apply the algorithm recursively to put the matrix B into the canonical diagonal form. Then, if $a_{1,1}$ divides every entry of the (now diagonal) matrix B, we are done. Otherwise, $a_{1,1} \nmid b_{i,i}$ for some i. In this case, add the ith column of B to the first column of A and apply step one to reduce the length of $a_{1,1}$ and continue.

This algorithm gives existence: every $A \in M_{s,t}(R)$ is equivalent to a diagonal matrix in canonical form. It just remains to prove uniqueness. So now suppose $\mathrm{diag}(d_1, \dots, d_u)$ and $\mathrm{diag}(d'_1, \dots, d'_u)$ are equivalent matrices in $M_{s,t}(R)$, where $d_1|d_2|\dots|d_u$ and $d'_1|\dots|d'_u$. We need to prove $d_i \operatorname{ass} d'_i$ for each i. To do this, define $J_i(A)$ to be the ideal of R generated by the determinants of all $i \times i$ minors ("sub-matrices") of a matrix A. The point is that $J_i(A)$ depends only on the *equivalence class* of the matrix A. (Why? Hopefully I'll remember to explain later on when we talk about exterior powers... You should at least be able to check this statement in the case of J_1 right away.) Clearly, $J_i(\mathrm{diag}(d_1, \dots, d_u)) = (d_1 d_2 \dots d_i)$ if $d_1|d_2|\dots|d_u$. Hence,

$$(d_1) = (d'_1), (d_1d_2) = (d'_1d'_2), \quad \dots, \quad (d_1 \dots d_u) = (d'_1 \dots d'_u).$$

This implies $d_i \operatorname{ass} d'_i$ for each i. □

The result just proved shows that to a matrix $A \in M_{s,t}(R)$ you can associate a sequence $d_1|d_2|\dots|d_u$ of R, unique up to associates, called the *invariant factor sequence* of the matrix A. Notice in the special case R is a field, each d_i is either 0 or 1 and the invariant factor sequence takes the form $1|\dots|1|0|\dots|0$; the number of 1's simply records the *rank* of the original matrix, which should be familiar to you from linear algebra. Thus the invariant factor sequence is a generalization to matrices over an arbitrary PID of the notion of rank of a matrix over a field.

Now we go back to studying submodules of free modules:

Structure of submodules of free modules over PIDs. *Let R be a PID, F be a free R-module of rank s and N be an R-submodule of F of rank $t \leq s$. Then, there exists a basis $f_1, \dots, f_s$ for F and elements $d_1, \dots, d_t \in R$ such that $d_1|d_2|\dots|d_t$ and $d_1f_1, d_2f_2, \dots, d_tf_t$ is a basis for N. Moreover, the elements $d_1, \dots, d_t$ are unique in the sense that if we have another basis $f'_1, \dots, f'_s$ for F and elements $d'_1, \dots, d'_t \in R$ such that $d'_1|d'_2|\dots|d'_t$ and $d'_1f'_1, d'_2f'_2, \dots, d'_tf'_t$ is a basis for N, then $d_i \operatorname{ass} d'_i$ for each i.*

Proof. We may as well assume that $N \neq (0)$ since that case is trivial. Let $\{f_1, \ldots, f_s\}$ be a basis for F and $\{n_1, \ldots, n_t\}$ $(t \leq s)$ be a basis for N, applying Theorem 3.7.1. Write

$$n_j = \sum_{i=1}^{s} a_{i,j} f_i$$

for $j = 1, \ldots, t$. So, $A = (a_{i,j})$ is an $s \times t$ matrix. Therefore, applying the canonical form for matrices over PIDs, we can find an invertible $s \times s$ matrix P and an invertible $t \times t$ matrix Q such that $P^{-1}AQ = D = \operatorname{diag}(d_1, \ldots, d_t)$ for elements $d_1 | d_2 | \ldots | d_t$ in R.

Now let $n'_j = \sum_{i=1}^{t} q_{i,j} n_i$, $f'_j = \sum_{i=1}^{s} p_{i,j} f_i$. Since the matrices P and Q are invertible, these also give bases for N and F respectively. Moreover,

$$n'_j = \sum_{i=1}^{t} q_{i,j} n_i = \sum_{i=1}^{t} \sum_{k=1}^{s} a_{k,i} q_{i,j} f_k = \sum_{i=1}^{s} d_j p_{i,j} f_i = d_j f'_j$$

since $AP = QD$. This gives us the required bases for F and N.

It remains to prove uniqueness. So take the two bases $f_1, \ldots, f_s$ and $f'_1, \ldots, f'_s$ as in the statement of the second part of the theorem. Let P be the change of basis matrix from $f_1, \ldots, f_s$ to $f'_1, \ldots, f'_s$ and Q be the change of basis matrix from $d_1 f_1, \ldots, d_t f_t$ to $d'_1 f'_1, \ldots, d'_t f'_t$. Then $P \operatorname{diag}(d_1, \ldots, d_t) = \operatorname{diag}(d'_1, \ldots, d'_t) Q$, so the two $s \times t$ diagonal matrices are equivalent, so $d_i \operatorname{ass} d'_i$ by the uniqueness of the invariant factors of a matrix proved above. □

We now apply the results of the previous section to prove the structure theorems for finitely generated modules over PIDs.

Recall that a cyclic R-module is an R-module M generated by a single element m; in that case, the R-module homomorphism $\pi : R \to M, r \mapsto rm$ is surjective so $M \cong R/\ker \pi$. We call $\ker \pi$ the *order ideal* of the cyclic module M, since it determines M uniquely up to isomorphism. Note $\ker \pi$ is the *annihilator in R of M*, that is, $\{r \in R \mid rM = (0)\}$. If R is even a PID, then $\ker \pi = (f)$ for some $f \in R$, unique up to associates, and then we call f the *order* of M.

For example, if $R = \mathbb{Z}$ then the cyclic $\mathbb{Z}$-modules are exactly the Abelian groups $\mathbb{Z}$, of order 0, or $\mathbb{Z}_n$ of order $\pm n$ (note the order of a cyclic module is only defined up to associates). In other words, the order of a cyclic $\mathbb{Z}$-module is simply its order as an Abelian group, or 0 if it is infinite. We will see another important example, when $R = F[X]$ for F a field, in section **??** when we study normal forms for linear transformations.

The main result is the following:

Structure theorem for finitely generated modules over PIDs. *Let M be a finitely generated R-module, where R is a PID. Then, M can be decomposed as an internal direct sum as*

$$M = M_1 \oplus \cdots \oplus M_s$$

where M_i is a non-zero cyclic submodule of order d_i and $d_1 | \ldots | d_s$ in R. Moreover, s and the orders $d_1, \ldots, d_s$ are uniquely determined up to associates. In other words, if

$$M = M'_1 \oplus \cdots \oplus M'_t$$

for non-zero cyclic submodules M'_i of order d'_i such that $d'_1 | \ldots | d'_t$, then $t = s$ and $d_i \operatorname{ass} d'_i$.

Proof. Existence. Since M is finitely generated, we can find generators $m_1, \ldots, m_s$ for M, where s is taken to be *minimal*. Let F be the free R-module on $\{x_1, \ldots, x_s\}$. Then, M is a quotient of F under the unique R-module homomorphism π sending $x_i \mapsto m_i$. Let $K = \ker \pi$. Applying the structure theorem for submodules of free modules over PIDs, we can find a basis $f_1, \ldots, f_s$ for F and elements $d_1 | d_2 | \ldots | d_s$ in R (where some of the d_i are possibly zero) so that the non-zero elements from $d_1 f_1, \ldots, d_s f_s$ form a basis for K. If any d_i is a unit, then $f_i \in \ker \pi$ so that just

$\pi(f_1), \ldots, \pi(f_{i-1}), \pi(f_{i+1}), \ldots, \pi(f_s)$ generate M, contradicting the minimality of s. So no d_i is a unit.

Now, let $M_i = \pi(Rf_i)$. Observe

$$M_i \cong Rf_i/(K \cap Rf_i) \cong Rf_i/Rd_if_i \cong R/(d_i).$$

Hence, M_i is cyclic of order d_i, so non-zero since d_i is not a unit. Moreover, since $f_1, \ldots, f_s$ generate F, $\pi(f_1), \ldots, \pi(f_s)$ generate M hence $M = M_1 + \cdots + M_s$. The sum is direct because if

$$m_1 + \cdots + m_s = 0$$

for $m_i = \pi(r_if_i) \in M_i$, then $r_1f_1 + \cdots + r_sf_s \in K$. Since $d_1f_1, \ldots, d_sf_s$ is a basis for K, we deduce that $d_i|r_i$ for each i so that each $m_i = 0$.

Uniqueness. Follows from the theorem below. □

The sequence $d_1|d_2|\ldots|d_s$ appearing in the theorem is called the *invariant factor sequence* of the module M. It determines M uniquely up to isomorphism. Thus, the structure theorem is a *classification* of the finitely generated modules over a PID by their invariant factor sequences.

Now let R be any commutative ring and M be an R-module. We call an element $m \in M$ a *torsion element* if $rm = 0$ for some $r \in R^*$. Then, M is called *torsion* if all its elements are torsion elements, and M is *torsion-free* if it has no non-zero torsion elements. For example a torsion $\mathbb{Z}$-module means an Abelian group all of whose elements have finite order; all vector spaces over a field are torsion free. In general, for an arbitrary R-module M, we let

$$M_t = \{m \in M \mid m \text{ is a torsion element}\}.$$

Then:

3.7.2. **Lemma.** *Let R be an integral domain, M an R-module. Then, M_t is an R-submodule of M, and M/M_t is torsion-free.*

Proof. Take $m_1, m_2 \in M_t$. Then, there are $r_1, r_2 \in R^*$ such that $r_im_i = 0$. Set $r = r_1r_2$, non-zero since R is an integral domain. Then, $r(m_1 + m_2) = 0$ so $m_1 + m_2 \in M_t$. The rest of the proof is similar... □

It is obvious that if R is an integral domain, then all free R-modules are torsion-free. In the case of finitely generated modules over PIDs, the converse is true, so that the local property "torsion-free" is equivalent to the global propety "free".

Torsion free ⇒ free for f.g. modules over PIDs. *Let R be a PID and M be a finitely generated R-module. Then, there is a (not necessarily unique!)* free *R-submodule M' of M such that $M = M_t \oplus M'$. In particular, M is free if and only if M is torsion-free.*

Proof. Apply the structure theorem to write $M = M_1 \oplus \cdots \oplus M_s$ where M_i is cyclic of order d_i and $d_1|\ldots|d_s$. Let

$$M' = \bigoplus_{i \text{ s.t. } d_i = 0} M_i.$$

Then, M' is free since it is a direct sum of copies of R. In particular, $M_t \cap M' = \{0\}$. On the other hand, each M_i with $d_i \neq 0$ is contained in M_t. This shows that $M = M_t + M'$ as required. The second statement in the theorem follows immediately since M is torsion-free if and only if $M_t = (0)$ which is if and only if $M = M'$. □

Note: for more general integral domains, the torsion submodule M_t of M will *not* in general have a complement in M. For an example of a torsion-free $\mathbb{Z}$-module that is *not* free, take the Abelian group $\mathbb{Q}$ (which is *not* finitely generated!).

The number of summands s in the decomposition of M in the statement of the structure theorem is the *smallest possible* number such that M can be written as a direct sum of s cyclic submodules. We turn to discussing the other extreme, when M is written as a direct sum of as many submodules as possible: the *primary decomposition* of M.

3.7.3. **Lemma.** *Let R be a PID and M be a cyclic R-module of order p^r where p is prime. Then, the* only *R-submodules of M are the following:*

$$(0) = p^r M \subset p^{r-1}M \subset \cdots \subset pM \subset M$$

and $p^i M/p^{i+1}M \cong R/(p)$. In particular, M is indecomposable.

Proof. Since $M \cong R/(p^r)$, the lattice of submodules of M is isomorphic to the lattice of ideals of R containing p^r, which in turn (since R is a PID) is isomorphic to the lattice of divisors of p^r. The result follows since p is prime. □

Primary decomposition theorem. *Let R be a PID and M be a finitely generated* torsion *R-module. Then, M can be written as*

$$M = M_1 \oplus \cdots \oplus M_t$$

for cyclic R-modules M_i of order $p_i^{n_i}$, where the p_i are (not necessarily distinct) primes. Moreover, given another such decomposition

$$M = M_1' \oplus \cdots \oplus M_{t'}'$$

with M_i' an indecomposable (even uniserial) cyclic of prime power order $q_i^{m_i}$, we have that $t' = t$ and (after reordering) $q_i^{m_i} = p_i^{n_i}$ for each i.

Proof. Existence. Follows from the preceeding theorem by the Chinese Remainder Theorem.
Uniqueness. This is just the Krull-Schmidt theorem. □

The primary decomposition theorem and Lemma 3.7.3 essentially give complete information about the submodule structure of a finitely generated module over a PID, which is remarkable. In particular, you should now be able to determine the *composition factors* of such an M. Note also that the modules M_i appearing in the primary decomposition are *indecomposable*. So the structure theorem is best remembered as saying: every finitely generated R-module is a finite direct sum of copies of finitely generated indecomposables, and the finitely generated indecomposables are the regular module (which is torsion-free) and then the cyclic modules of prime power order (which are torsion).

Special case:

Classification of finitely generated abelian groups. *Every finitely generated abelian group is a finite direct sum of finitely generated indecomposable abelian groups, and the finitely generated indecomposables are cyclic groups of orders infinity or some prime power.*

For example, here are the abelian groups of order 12:

$$\mathbb{Z}_4 \oplus \mathbb{Z}_3, \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3.$$

And so on...

Special case:

Classification of finitely generated $F[x]$-modules. *Let F be a field and M be a finitely generated $F[x]$-module. Then M is a finite direct sum of finitely generated indecomposable $F[x]$-modules, and the finitely generated indecomposables are either the regular module $F[x]$ or quotients $F[x]/(f(x)^n)$ for irreducible polynomials $f(x) \in F[x]$ and $n \geq 1$.*

For example if the field is $\mathbb{C}$ (or any algebraically closed field) the finitely generated torsion indecomposable modules are $\mathbb{C}[x]/(x-\lambda)^n$ for $\lambda \in \mathbb{C}$ and $n \geq 1$.

If the field is $\mathbb{R}$ they are $\mathbb{R}[x]/(x-\lambda)^n$ for $\lambda \in \mathbb{R}$ and $\mathbb{R}[x]/(x^2+bx+c)^n$ for $a, b \in \mathbb{R}$ with $b^2 - 4c < 0$.

If the field is $\mathbb{Q}$ we understand finitely generated $\mathbb{Q}[x]$-modules modulo the hard problem of understanding the irreducible $\mathbb{Q}[x]$-polynomials.

3.8 Applications to linear algebra

I want to remind you of another of the basic ideas in linear algebra: change of basis. We've already used this a bunch of times... Let F be a field, V and W be finite dimensional vector spaces over F and $f : V \to W$ be a linear transformation. Fix bases $v_1, \ldots, v_n$ for V and $w_1, \ldots, w_m$ for W. Then, the linear transformation f is uniquely determined by knowledge of the vectors $f(v_1), \ldots, f(v_n)$. So if we write

$$f(v_j) = \sum_{i=1}^{m} A_{i,j} w_i$$

for an $m \times n$ matrix $A = (A_{i,j})$, the original linear transformation f is determined uniquely by the matrix A (providing of course you know what the bases v_i and w_j are to start with!). We call this matrix A the *matrix of f with respect to the bases* $(v_j), (w_i)$. The above formula is the

Golden rule of linear algebra: The jth column of the matrix A of f in the given bases is $f(v_j)$ expanded in terms of the w_i's.

Then one needs to consider what happens to this matrix A of f in one set of bases if we switch to another set of bases. So first, let $v'_1, \ldots, v'_n$ be another basis for V. Then, we obtain the *change of basis matrix* $P = (P_{i,j})$ by writing the v'_j in terms of the v_i:

$$v'_j = \sum_{i=1}^{n} P_{i,j} v_i.$$

On the other hand, we can write the v_j in terms of the v'_i giving us the *inverse change of basis matrix* $P' = (P'_{i,j})$:

$$v_j = \sum_{i=1}^{n} P'_{i,j} v'_i.$$

The relationship between P and P' comes from the equation:

$$v'_j = \sum_{i=1}^{n} P_{i,j} v_i = \sum_{i,h} P'_{h,i} P_{i,j} v'_h.$$

This tells us that

$$\sum_{i=1}^{n} P'_{h,i} P_{i,j} = \delta_{h,j},$$

where $\delta_{h,j} = 1$ if $h = j$ or 0 otherwise. In other words, $P'P = I_n$. Similarly one gets that $PP' = I_n$. Hence, the change of basis matrix P is invertible, and the inverse change of basis matrix P' really is its inverse!

Now also let $w'_1, \ldots, w'_m$ be another basis for W and let Q be the change of basis matrix in this case, defined from

$$w'_j = \sum_{i=1}^{m} Q_{i,j} w_i.$$

Now let B be the matrix of our original linear transformation f with respect to the bases (v'_j) and (w'_i), so

$$f(v'_j) = \sum_{i=1}^{m} B_{i,j} w'_i = \sum_{i,h} Q_{h,i} B_{i,j} w_h.$$

Also,

$$f(v'_j) = f(\sum_{i=1}^{n} P_{j,i} v_j) = \sum_{i,h} A_{h,j} P_{j,i} w_h.$$

Equating w_h coefficients gives us that $QB = AP$. We have shown:

Change of basis theorem. *Given bases $(v_j), (v'_j)$ for V with change of basis matrix P and bases $(w_i), (w'_i)$ for W with change of basis matrix Q, the relationship between the matrix A of a linear transformation $f : V \to W$ in the unprimed bases to its matrix B in primed bases is given by*

$$B = Q^{-1}AP, \qquad A = QBP^{-1}.$$

Note that the matrices A and B in the theorem are thus *equivalent* matrices in the sense defined in section 3.7. The result proved there shows (since we are working now over a field) that bases for V and W can be chosen with respect to which the matrix of f looks like an $r \times r$ identity matrix in the top left corner with zeros elsewhere. Of course, this integer r is the *rank* of the linear transformation f, and is simply equal to the dimension of the image of f. I'm sure you remember the basic equation in linear algebra which is nothing more than the first isomorphism theorem for vector spaces:

The rank-nullity theorem. $\dim V = \dim \ker f + \dim \operatorname{im} f$.

Now we want to consider the other fundamental question: when $f : V \to V$ is an *endomorphism* of a vector space V. Then it makes more sense to record the matrix of the linear transformation f with respect to just one basis $v_1, \ldots, v_n$ for V (i.e. take $w_i = v_i$ in the above notation). We obtain the $n \times n$ matrix A of f with respect to the basis $v_1, \ldots, v_n$ so that

$$f(v_j) = \sum_{i=1}^{n} A_{i,j} v_i.$$

Now take another basis $v'_1, \ldots, v'_n$ and let P be the change of basis matrix as before. Then the change of basis theorem tells us in the special case that:

Change of basis theorem′. *If (v_i) and (v'_i) are two bases for V related by change of basis matrix P, and A is the matrix of an endomorphism $f : V \to V$ with respect to the first basis, B is its matrix with respect to the second basis, then*

$$B = P^{-1}AP, \qquad A = PAP^{-1}.$$

So: the matrix of f in the (v'_i) basis is obtained from the matrix of f in the (v_j) basis by *conjugating by the change of basis matrix*. Now define two $n \times n$ matrices to be *similar*, written $A \sim B$ if there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. One easily checks that similarity is an equivalence relation on $n \times n$ matrices. The equivalence classes are the *similarity classes* (or *conjugacy classes*) of matrices over F.

We will be interested in finding a set of representatives for the similarity classes of matrices having a particularly nice form: a *normal form*. Equivalently, given a linear transformation $f : V \to V$, we are interested in finding a nice basis for V with respect to which the matrix of f looks especially nice.

Suppose in this section that the field F is *algebraically closed*. This means that the irreducible polynomials in the polynomial ring $F[x]$ are of the form $(x - \lambda)$ for $\lambda \in F$. Fix once and for all a

linear transformation $f : V \to V$ of some finite dimensional vector space V over F. The goal is to pick a nice basis for V...

We make V into an $F[x]$-module, so that x acts on V by the linear transformation f. In other words, a polynomial $a_n x^n + \cdots + a_1 x + a_0 \in F[x]$ acts on V by the linear transformation $a_n f^n + \cdots + a_1 f + a_0$. Note the way V is viewed as an $F[x]$-module encodes the linear transformation f into the module structure: we can recover f simply as the linear transformation of V determined by multiplication x.

Now, V is a finitely generated torsion $F[x]$-module – indeed, it is even finitely generated as an F-module as its finite dimensional. Also, $F[x]$ is a PID. So we can apply the primary decomposition theorem for modules over PIDs to get that V decomposes as

$$V = V_1 \oplus \cdots \oplus V_N$$

where each V_i is a cyclic $F[x]$-module of (non-zero) prime power order. So since we're working over an algebraically closed field, we get scalars $\lambda_i \in F$ and $d_i \in \mathbb{N}$ such that

$$V_i \cong F[x]/((x - \lambda_i)^{d_i})$$

as $F[x]$-modules.

Now to find a nice basis for V, we focus on a particular summand V_i and want a nice basis for that V_i. In other words, using the above isomorphism, we should study the $F[x]$-module $F[x]/((x - \lambda)^d)$ and find a basis in which the matrix of the linear transformation determined by multiplication by x has a nice shape.

3.8.1. **Lemma.** *The images of the elements $(x-\lambda)^{d-1}, (x-\lambda)^{d-2}, \ldots, (x-\lambda), 1$ in the $F[x]$-module $F[x]/((x-\lambda)^d)$ form an F-basis, Moreover, the linear transformation determined by multiplication by x has the following matrix in this basis:*

$$\begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & \ldots & \lambda & 1 \\ 0 & & \ldots & & \lambda \end{bmatrix}$$

(We call the above matrix the Jordan block *of type $(x - \lambda)^d$.)*

Proof. We're considering $F[x]/((x - \lambda)^d)$. As an F-vector space, $F[x]$ has basis given by all the monomials $1, x, x^2, \ldots$ (its infinite dimensional of course). But $(x - \lambda)^d$ is a monic polynomial of degree d so in the quotient $F[x]/((x-\lambda)^d)$, we can rewrite x^d as a linear combination of lower degree monomials. In other words, we just need the monomials $1, x, x^2, \ldots, x^{d-1}$ to span the quotient $F[x]/((x - \lambda)^d)$. Moreover, they're linearly independent, else there'd be an element in the ideal $(x - \lambda)^d$ of degree less than d. Hence, the images of $1, x, \ldots, x^{d-1}$ give a basis for the quotient. It follows easily that the images of $(x - \lambda)^{d-1}, \ldots, (x - \lambda), 1$ also form a basis, since they're related to the preceeding basis by a "unitriangular" change of basis matrix.

It just remains to write down the linear transformation "multiply by x" in this basis. We have that

$$x(x - \lambda)^i = (x - \lambda)^{i+1} + \lambda(x - \lambda)^i.$$

Now just write down the matrix, remembering: the jth column of the matrix of the linear transformation is given by expanding its effect on the jth basis vector in terms of the basis. □

Now recall that we've decomposed $V = V_1 \oplus \cdots \oplus V_N$ with each $V_i \cong F[x]/((x - \lambda_i)^{d_i})$. The lemma gives us a choice of a nice basis for each $F[x]/((x-\lambda_i)^{d_i})$. Lifting it through the isomorphism we get a nice basis for V_i, hence putting them all together, we get a nice basis for V. We've proved the first half of:

Jordan normal form. *There exists a basis for V in which the matrix of the linear transformation f has the* Jordan normal form*:*

$$\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & B_N \end{bmatrix}$$

where B_i is the Jordan block of type $(x-\lambda_i)^{d_i}$. Moreover, this normal form for the linear transformation f is unique *up to permuting the Jordan blocks.*

Proof. It just remains to prove the uniqueness statement. This will follow from uniqueness of the orders of the indecomposable modules appearing in the primary decomposition of the $F[x]$-module V. We just need show that these orders are exactly the $(x-\lambda_i)^{d_i}$: in other words, the types of the Jordan blocks are determined exactly by the orders of the primary components of the module. Since these orders are uniquely determined up to reordering, that gives that the Jordan blocks are uniquely determined up to reordering.

Everything now reduces to considering the special case that f has just one Jordan block of type $(x-\lambda)^d$, and we need to prove that this is also the order of V viewed as an $F[x]$-module. Let the basis in which f has the Jordan block form be $v_1, \dots, v_d$. Then, V is a cyclic $F[x]$-module generated by v_d. So we just need to calculate the order of v_d. We are given that $(x-\lambda)v_i = v_{i-1}$ (where $v_0 = 0$ by convention). Hence, $(x-\lambda)^d v_d = 0$, while the $(x-\lambda)^i v_d$ for $i = 0, \dots, d-1$ are linearly independent. This shows that no non-zero polynomial in x of degree less than d can annihilate the generator v_d, but $(x-\lambda)^d$ does. Hence, the order of v_d is $(x-\lambda)^d$, as required. □

I now wish to give a rather silly consequence of the Jordan normal form. First, recall that the *characteristic polynomial* of the linear transformation f is the polynomial in x obtained by calculating

$$\det(A - xI)$$

where A is the matrix representing f in any basis of V. Write $\chi_f(x) \in F[x]$ for the characteristic polynomial of f. Note its definition is independent of the choice of basis used to calculate it! In other words, we may as well pick the basis in which A has Jordan normal form – then in the above notation,

$$\chi_f(x) = \prod_{i=1}^{N} (\lambda_i - x)^{d_i}.$$

So the characteristic polynomial tells us the *eigenvalues* – ie the diagonal entries in the Jordan normal form – together with their multiplicities. But it doesn't give enough information (unless all eigenvalues turn out to be distinct) to work out the precise sizes of the Jordan blocks in the JNF.

Cayley-Hamilton theorem. $\chi_f(f) = 0$.

Proof. Explicit calculation writing f as a matrix in Jordan normal form and using the above formula for $\chi_f(x)$. □

Remark. The Cayley-Hamilton theorem is true more generally. Let R be an arbitrary *integral domain* and A be an $n \times n$ matrix with entries in R. Then, we can define its *characteristic polynomial* in exactly the same way as above, namely, $\chi_A(x) = \det(A - xI_n) \in R[x]$. Then, we always have that the matrix $\chi_A(A)$ is the zero matrix. Proof: Embed R into its field of fractions (section 2.5) and its field of fractions into its algebraic closure (section 5.3). Then the conclusion is immediate from the Cayley-Hamilton theorem over an algebraically closed field that we just proved.

There is one other useful polynomial associated to the linear transformation f which can help in computing the Jordan normal form: the *minimal polynomial* of the linear transformation f. This is defined as the unique monic polynomial $m_f(x) \in F[x]$ of minimal degree such that $m_f(f) = 0$. In other words (this is how you see uniqueness), $m_f(x)$ is the monic polynomial that generates the ideal of $F[x]$ consisting of all polynomials that act as zero on the $F[x]$-module V.

3.8.2. **Lemma.** *In the standing notation for the Jordan normal form of f, we have that*

$$m_f(x) = \text{LCM}\{(x-\lambda_i)^{d_i} \mid i = 1, \ldots, N\}.$$

Proof. We showed in the proof of the Jordan normal form that the minimal polynomial of a single Jordan block of type $(x-\lambda)^d$ was exactly $(x-\lambda)^d$. This polynomial also gives zero when any Jordan block of type $(x-\lambda)^e$ for $e < d$ is substituted for x, but gives an invertible matrix on any other Jordan block of type $(x-\mu)^c$ for $\mu \neq \lambda$. Hence, $m_f(x)$ must be exactly divisible by $(x-\lambda)^d$ for each eigenvalue λ with d being the maximum of all the d_i with $\lambda_i = \lambda$. □

Of course, the lemma shows in particular that $m_f(x)$ divides $\chi_f(x)$, giving the Cayley-Hamilton theorem again! Knowing the minimal polynomial gives a little extra information: it tells you the biggest Jordan block associated to each eigenvalue λ. For small matrices, knowing both the characteristic and the minimal polynomial is often enough to write down the Jordan normal form right away.

I wish to briefly mention one normal form which (unlike the JNF) makes sense over an arbitrary field F: the rational normal form. So now let F be any field, V a finite dimensional vector space over F and $f : V \to V$ be a linear transformation. Always, we view V as an $F[x]$-module so that x acts on V as the linear transformation f.

We can decompose V using the structure theorem for finitely generated modules over PIDs, giving that V decomposes as an $F[x]$-modules as

$$V = V_1 \oplus \cdots \oplus V_N$$

where each V_i is cyclic of order g_i, where $g_i \in F[x]$ is monic and $g_1 | g_2 | \ldots | g_N$. Hence,

$$V_i \cong F[x]/(g_i)$$

as an $F[x]$-module. So if we're looking for a nice basis for V, we should consider finding a nice basis for the cyclic $F[x]$-module of order g.

3.8.3. **Lemma.** *Let $g(x) \in F[x]$ be the monic polynomial*

$$g(x) = x^m - a_{m-1}x^{m-1} - \cdots - a_1 x - a_0.$$

Then, the $F[x]$-module $F[x]/(g)$ has F-basis the images of $1, x, \ldots, x^{m-1}$. Moreover, in this basis, the matrix of the linear transformation given by multiplication by x has the form:

$$\begin{bmatrix} 0 & & \ldots & & a_0 \\ 1 & 0 & \ldots & & a_1 \\ \vdots & \ddots & \ddots & & \vdots \\ & \ldots & 1 & 0 & a_{m-2} \\ 0 & \ldots & & 1 & a_{m-1} \end{bmatrix}$$

(This matrix is called the companion matrix *of the monic polynomial $g(x)$.)*

Proof. That the given elements form a basis follows in exactly the same way as in the proof of Lemma 3.8.1. You just have to calculate the matrix of the linear transformation given by multiplication by x: its jth column is what x does to the jth basis vector! □

Now take the decomposition of V above. So each V_i is isomorphic to $F[x]/(g_i)$ for a monic polyomial $g_i(x)$. Applying the lemma, we get a basis for V_i and putting all the bases together, we get a basis for V. We deduce from the lemma that:

Rational normal form. *There exists a basis for V in which the matrix of the linear transformation f has the following form:*

$$\begin{bmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_2 & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & C_N \end{bmatrix}$$

where C_i is the companion matrix of a monic polynomial $g_i \in F[x]$ and $g_1|g_2|\dots g_N$.

Remark. One can also show using the uniqueness in the structure theorem for modules over PIDs that the rational normal form for f is *unique*. In other words, if the matrices of the endomorphisms f and g (with respect to some choice of basis) are similar then they have the same rational normal form.

Now consider the minimal polynomial of the linear transformation f again. Recall $m_f(x)$ is the monic polynomial that generates the ideal of $F[x]$ consisting of all polynomials that act as zero on the $F[x]$-module V. Let $V = V_1 \oplus \dots \oplus V_N$ be its decomposition according to the structure theorem, so V_i is of degree g_i and $g_1|\dots|g_N$. Then, it is obvious that $m_f(x) = g_N(x)$. So the minimal polynomial precisely tells you the *largest block* in the rational normal form of f.

For example, if the minimal polynomial has the same degree as the dimension of V (equivalently, if V is already a cyclic $F[x]$-module) you get lucky and the rational normal form is simply the companion matrix of the minimal polynomial, which coincides with the characteristic polynomial by the Cayley-Hamilton theorem.

Finally let me discuss diagonalizability. Let f be an endomorphism of a finite dimensional vector space V. You should recall that for $\lambda \in F$, the λ-eigenspace of V is the subspace

$$V_\lambda = \ker(f - \lambda \,\mathrm{id}),$$

i.e. the span of all *eigenvectors* with eigenvalue λ. We record:

3.8.4. **Lemma.** $\sum_{\lambda \in F} V_\lambda = \bigoplus_{\lambda \in F} V_\lambda$.

Proof. Suppose not. Then we can find distinct eigenvalues $\lambda_1, \dots, \lambda_n$ and $0 \neq v_i \in V_{\lambda_i}$ such that $v_1 + \dots + v_n = 0$. Take such vectors $v_1, \dots, v_n$ with n *minimal*. Applying $f - \lambda_1 \,\mathrm{id}$, which annihilates v_1 and maps v_i to $(\lambda_i - \lambda_1)v_i$ for $i > 1$, we get that $(\lambda_2 - \lambda_1)v_2 + \dots + (\lambda_n - \lambda_1)v_n = 0$. This contradicts the minimality of the choice of n. □

We call the linear transformation f *diagonalizable* over F if there exists a basis for V with respect to which the matrix of f is a diagonalizable matrix.

3.8.5. **Lemma.** *The following are equivalent:*
(1) *f is diagonalizable over F;*
(2) *$V = \sum_{\lambda \in F} V_\lambda$ (i.e. V is spanned by eigenvectors);*
(3) *V has a basis of eigenvectors.*

Proof. The equivalence of (1) and (3) is immediate from the definition. Clearly, (3) implies (2), while (2) implies (3) by Lemma 3.8.4. □

Now we prove the important:

Criterion for diagonalizibility. *The linear transformation f is diagonalizable over F if and only if the minimal polynomial $m_f(x)$ splits as a product of distinct linear factors in $F[x]$.*

Proof. View V as an $F[x]$-module so x acts on V via f. Let

$$V = V_1 \oplus \dots \oplus V_N$$

be the structure theorem decomposition, so V_i is cyclic of order $g_i \in F[x]$ and $g_1|\dots|g_N$. We have observed before that $m_f(x) = g_N(x)$.

Now by Lemma 3.8.5, f is diagonalizable over F if and only if V has a basis $v_1, \dots, v_n$ of eigenvectors which is if and only if V decomposes according to the primary decomposition theorem as

$$V = V_1' \oplus \cdots \oplus V_n'$$

with each V_i' being a one dimensional cyclic $F[x]$-module. This last statement is equivalent to $g_N(x)$, the largest order polynomial appearing in the structure theorem decomposition, being a product of linear factors. □

Criterion for simultaneous diagonalizability. *Let $f_i\,(i \in I)$ be a family of linear transformations of V. Then, they are* simultaneously diagonalizable *(i.e. there exists a basis for V in which the matrices of* all *f_i are diagonal) if and only if each f_i is diagonalizable and $f_i \circ f_j = f_j \circ f_i$ for all $i, j \in I$ (i.e. the f_i commute pairwise).*

Proof. Obviously, if the f_i are simultaneously diagonalizable, they are each diagonalizable and they commute pairwise since diagonal matrices commute.

Conversely, suppose each f_i is diagonalizable and that they commute pairwise. We proceed by induction on $\dim V$, there being nothing to prove if $\dim V = 0$. So suppose $\dim V > 0$. If all f_i act as scalars on V, then any basis for V simultaneously diagonalizes all f_i. Else, choose some f_i which does not act as a scalar on V and consider the f_i-eigenspace decomposition

$$V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_m}$$

of V where $1 \leq \dim V_{\lambda_k} < \dim V$ for each $k = 1, \dots, m$.

We claim that for each $k = 1, \dots, m$ and each $j \in I$, $f_j(V_{\lambda_k}) \subseteq V_{\lambda_k}$ and moreover that the restriction of f_j to the subspace V_{λ_k} is diagonalizable. Since clearly the restrictions of the f_j to V_{λ_k} pairwise commute, the claim will then give by the induction hypothesis that the f_j are simultaneously diagonalizable on each V_{λ_k}, hence on all of V proving the theorem.

For the claim, note that for $v \in V_{\lambda_k}$,

$$f_i f_j(v) = f_j f_i(v) = f_j \lambda_k v = \lambda_k f_j v$$

hence $f_j(v) \in V_{\lambda_k}$. Now f_j is diagonalizable on V, so by the criterion for diagonalizability, the minimal polynomial $m_{f_j}(x)$ of f_j splits as a product of distinct linear factors. But clearly the minimal polynomial of the restriction of f_j to V_{λ_k} divides $m_{f_j}(x)$, hence also splits as a product of distinct linear factors. Hence by the criterion for diagonalizability again, the restriction of f_j to V_{λ_k} is also diagonalizable. □

3.9 Tensor products

Suppose initially that F is a *field*, and let V, W be vector spaces over F. Fix bases $(v_i)_{i \in I}$ for V and $(w_j)_{j \in J}$ for W. Define $V \otimes W$ to be the F-vector space with basis

$$\{v_i \otimes w_j \mid i \in I, j \in J\}.$$

So if V and W are both finite dimensional,

$$\dim V \otimes W = \dim V \dim W.$$

Now we define a map $\iota : V \times W \to V \otimes W$ by

$$\iota\left(\sum_{i \in I} a_i v_i, \sum_{j \in J} b_j w_j\right) = \sum_{i \in I} \sum_{j \in I} a_i b_j v_i \otimes w_j$$

for arbitrary coefficients a_i and b_j in F (all but finitely many of each being zero).

The map $\iota : V \times W \to V \otimes W$ is *bilinear*, meaning that

$$\iota(cv + c'v', w) = c\iota(v, w) + c'\iota(v', w), \quad \iota(v, cw + c'w') = c\iota(v, w) + c'\iota(v, w')$$

for all $v, v' \in V, w, w' \in W, c, c' \in F$.

For $v \in V, w \in W$, we denote $\iota(v, w) \in V \otimes W$ instead by $v \otimes w$, and call such an element of the vector space $V \otimes W$ a *pure tensor*. It is crucial to bear in mind that *not every vector in $V \otimes W$ can be represented as a pure tensor* – for instance, consider $v_1 \otimes w_1 + v_2 \otimes w_2$!. But certainly, every vector in $V \otimes W$ can be written as a sum of finitely many pure tensors, that is, the pure tensors *generate $V \otimes W$ as an Abelian group.*

The vector space $V \otimes W$, together with the bilinear map $\iota : V \times W \to V \otimes W$, has the crucial property:

> *given an F-vector space U and a bilinear map $f : V \times W \to U$, there exists a unique linear map $\bar{f} : V \otimes W \to U$ such that $f = \bar{f} \circ \iota$.*

Indeed, there is no choice but to define $\bar{f}(v_i \otimes w_j) = f(v_i, w_j)$ on the basis elements $v_i \otimes w_j$ of $V \otimes W$ and then extend to all of $V \otimes W$ linearly. The resulting linear map $\bar{f}$ clearly satisfies the property $f = \bar{f} \circ \iota$. This universal property is the key to finding the right generalization of our tensor product $V \otimes W$ of two vector spaces to arbitrary modules over arbitrary rings...

Recall the notion of a *bimodule*. Let R, S, T be rings and suppose ${}_RM_S$ and ${}_SN_T$ are bimodules over the rings as indicated. If U is any R, T-bimodule, we call a map

$$f : {}_RM_S \times {}_SN_T \to {}_RU_T$$

a *balanced map* if the following properties hold:

(1) $f(m + m', n) = f(m, n) + f(m', n)$;

(2) $f(m, n + n') = f(m, n) + f(m, n')$;

(3) $f(ms, n) = f(m, sn)$;

(4) $f(rm, n) = rf(m, n)$;

(5) $f(m, nt) = f(m, n)t$;

for all $r \in R, s \in S, t \in T, m, m' \in M, n, n' \in N$. You should think of this definition of balanced as a generalization of the notion of "bilinear".

Now we define a *tensor product* of ${}_RM_S$ and ${}_SN_T$ to be an R, T-bimodule

$$M \otimes_S N = {}_RM_S \otimes_S {}_SN_T$$

together with a balanced map

$$\iota : M \times N \to M \otimes_S N$$

with the property that for any other R, T-bimodule ${}_RU_T$ and any other balanced map $f : M \times N \to U$, there exists a unique R, T-bimodule homomorphism $\bar{f} : M \otimes_S N \to U$ such that $f = \bar{f} \circ \iota$. As usual with universal properties, if such a bimodule $M \otimes_S N$ and map ι exists, it is unique up to canonical isomorphism. So we will always just call it *the* tensor product of M and N over S. But still, we need to prove existence:

Existence of tensor products. *For any R, S-bimodule ${}_RM_S$ and any S, T-bimodule ${}_SN_T$, the tensor product $\iota : M \times N \to M \otimes_S N$ exists.*

Proof. Let F be the free Abelian group on the *set* $M \times N$ (this could be absolutely huge!). Before doing anything, we need to make F into an R,T-bimodule. Given $r \in R$ define a map $l_r : M \times N \to F$ by $l_r(m,n) = (rm,n) \in F$. The universal property of F gives that this extends uniquely to a homomorphism $\bar{l}_r : F \to F$ of Abelian groups. This map $\bar{l}_r$ determines left multiplication by r on F, making F into a left R-module. Similarly, we make F into a right T-module, hence an R,T-bimodule.

Now let K be the Abelian subgroup of F generated by all elements
(1) $(m+m',n)-(m,n)-(m',n)$;
(2) $(m,n+n')-(m,n)-(m,n')$;
(3) $(ms,n)-(m,sn)$
for all $m,m' \in M, n,n' \in N, s \in S$. Consider the quotient Abelian group $M \otimes_S N := F/K$. Note that K is a left R-submodule of F, as well as a right S-submodule (it is a sub-R,S-bimodule!). Hence the R,S-bimodule structure on F induces a well-defined R,S-bimodule structure on $M \otimes_S N = F/K$. So $M \otimes_S N$ is an R,T-bimodule. Writing $m \otimes n$ for the image of the basis element (m,n) of F in $M \otimes_S N$ under the quotient map, we obtain a map

$$\iota : M \times N \to M \otimes_S N, \quad (m,n) \mapsto m \otimes n.$$

This map is balanced!

Now we check that ι and the R,T-bimodule $M \otimes_S N$ we have constructed really do satisfy the universal property. Let U be an R,T-bimodule and $f : M \times N \to U$ be a balanced map. By the universal property of free module, f extends uniquely to an Abelian group homomorphism $\bar{f} : F \to U$, which is automatically an R,T-bimodule map. Since f is balanced, all generators of K are annihilated by $\bar{f}$. So $\bar{f}$ factors through the quotient F/K to induce the unique $f' : M \otimes N \to U$ with $f = f' \circ \iota$ as required. □

Remarks. (i) Our construction of ${}_RM_S \otimes_S {}_SN_T$ really did not involve the left R- or the right T-module structures. In other words, if we viewed M just as a $\mathbb{Z},S$-bimodule and N just as an $S,\mathbb{Z}$-bimodule, the resulting tensor product ${}_{\mathbb{Z}}M_S \otimes_S {}_SN_{\mathbb{Z}}$ would have been the *same* as ${}_RM_S \otimes_S {}_SN_T$. The only point of including the left and right module structures was to check that they are preserved throughout the construction.

(ii) If R is a commutative ring and M,N are (left, say) R-modules, we have remarked earlier that there is a standard way to view M and N as R,R-bimodules. Then, the tensor product $M \otimes_R N$ makes sense and is an R,R-bimodule. For instance, let F be a field and V,W be F-vector spaces. Then $V \otimes_F W$ is an F-vector space in this way. Moreover, it is canonically isomorphic to the basis-dependent construction of $V \otimes W$ in terms of bases given at the beginning of the section, since that was shown to also satisfy the universal property of tensor product.

We now give some examples of how to apply the universal property to prove things about tensor products.

Examples. (1) Let R be a commutative ring and M,N be R-modules viewed as R,R-bimodules in the standard way. Then, $M \otimes_R N \cong N \otimes_R M$. Proof. Define a map $M \times N \to N \otimes_R M$ by $(m,n) \mapsto n \otimes m$. This is balanced, hence by the universal property it induces a unique R-module map $M \otimes_R N \to N \otimes_R M$ with $m \otimes n \mapsto n \otimes m$. Similarly, there is a unique R-module map $N \otimes_R M \to M \otimes_R N$ with $n \otimes m \mapsto m \otimes n$. The two are inverse to each other, hence $M \otimes_R N \cong N \otimes_R M$.

(2) If F is a field and V,W are finite dimensional F-vector spaces, then

$$\mathrm{Hom}_F(V,W) \cong V^* \otimes W.$$

Proof. Define a map $\theta : V^* \times W \to \mathrm{Hom}_F(V,W)$ by $(f,w) \mapsto \theta_{f,w}$ where $\theta_{f,w}(v) = f(v)w$ for all $v \in V$. This is bilinear so induces a unique map $\bar{\theta} : V^* \otimes W \to \mathrm{Hom}_F(V,W)$.

I claim this map $\bar{\theta}$ is bijective. Well, pick bases $v_1,\ldots,v_n$ for V and $w_1,\ldots,w_m$ for W. Define $f_{i,j} : V \to W$ by letting $f_{i,j}(\sum_{s=1}^n a_s v_s) = a_i w_j$. Then, the $f_{i,j}$ form a basis for $\mathrm{Hom}_F(V,W)$

(matrices!). Also, if $f_1, \ldots, f_n$ denotes the basis for V^* dual to $v_1, \ldots, v_n$, the $f_i \otimes w_j$ form a basis for $V^* \otimes W$. Now our map θ satisfies $\theta_{f_i, w_j} = f_{i,j}$. Hence, $\bar{\theta}$ maps one basis bijectively to the other. So $\bar{\theta}$ is an isomorphism.

(3) If ${}_RM_S$, ${}_SN_T$ and ${}_TP_U$ are three bimodules, then

$$({}_RM_S \otimes_S {}_SN_T) \otimes_T {}_TP_U \cong_R M_S \otimes_S ({}_SN_T \otimes_T {}_TP_U)$$

as R, U-modules. Proof. We would like just to define the isomorphism between them on generators by $(m \otimes n) \otimes p \mapsto m \otimes (n \otimes p)$. The whole problem is to show that there really is such an R, U-bimodule homomorphism! Of course, we need to use the universal property to prove this!

So first, fix $p \in P$ and define a map $f_p : M \times N \to M \otimes (N \otimes P)$ by $f_p(m, n) = m \otimes (n \otimes p)$. This is balanced, so induces a unique $\bar{f}_p : M \otimes N \to M \otimes (N \otimes P)$. Now define a map $(M \otimes N) \times P \to M \otimes (N \otimes P)$ by $(u, p) \mapsto \bar{f}_p(u)$. Again, this is balanced, so induces a unique map $(M \otimes N) \otimes P \to M \otimes (N \otimes P)$, which satisfies $(m \otimes n) \otimes p = f_p(m, n) = m \otimes (n \otimes p)$ on generators, as we wanted originally!

(4) For any ring R and any left R-module M,

$$R \otimes_R M \cong M.$$

Proof. Define a map $R \times M \to M$ by $(r, m) \mapsto rm$. This is balanced, so induces a unique R-module map $R \otimes_R M \to M$ such that $r \otimes m \mapsto rm$ on generators. The inverse is the map $m \in M$ to $1_R \otimes m \in R \otimes_R M$.

(5) $\otimes$ commutes with arbitrary direct sums. This means that given R, S-bimodules M_i $(i \in I)$ and S, T-bimodules N_j $(j \in J)$,

$$\left(\bigoplus_{i \in I} M_i \right) \otimes_S \left(\bigoplus_{j \in J} N_j \right) \cong \bigoplus_{i \in I, j \in J} M_i \otimes_S N_j$$

as R, T-bimodules. Indeed, one defines a map

$$\left(\bigoplus_{i \in I} M_i \right) \times \left(\bigoplus_{j \in J} N_j \right) \cong \bigoplus_{i \in I, j \in J} M_i \otimes_S N_j$$

by

$$\left(\sum_{i \in I} m_i, \sum_{j \in J} n_j \right) \mapsto \sum_{i \in I, j \in J} m_i \otimes n_j.$$

This is balanced, and so the universal property induces the required isomorphism. (How do you construct the inverse to prove this?)

We want to explain that the isomorphisms just constructed in these examples are *natural*! Roughly speaking, this means that they were defined without resorting to anything specific like fixed bases or generators. (In example (2) we did pick bases but only *after* we had constructed the map in a choice-independent way.) So we now need to understand how $\otimes$ is a *functor*, and then will be able to reinterpret isomorphisms like in the examples above as *natural isomorphisms* between functors. The word *natural* in mathematics should always indicate that there is some underlying natural isomorphism between functors – you should never use the too-often-overused word "natural" for anything else!

To start with, fix an R, S-bimodule ${}_RB_S$. Then, given any right R-module A_R, we have explained that

$$A_R \otimes_R {}_RB_S$$

is a right S-module. Now suppose that we are given a morphism $f : A_R \to A'_R$ between two right R-modules. We want to define a morphism

$$f \otimes \mathrm{id}_B : A_R \otimes_R {}_R B_S \to A'_R \otimes_R {}_R B_S$$

of right S-modules. To do this, start with the map $A \times B \to A' \otimes_R B$ defined by $(a, b) \mapsto f(a) \otimes b$. This is balanced, since f is a right R-module homomorphism. Hence by the universal property it induces a unique S-module homomorphism $f \otimes \mathrm{id}_B : A \otimes_R B \to A' \otimes_R B$ as required, satisfying $(f \otimes \mathrm{id}_B)(a \otimes b) = f(a) \otimes b$ for all $a \in A, b \in B$.

In this way, we obtain an (additive) functor

$$? \otimes_R {}_R B_S : \mathfrak{mod}\text{-}R \to \mathfrak{mod}\text{-}S.$$

You could of course do all the same arguments on the other side, to construct instead an additive functor

$${}_R B_S \otimes_S ? : S\text{-}\mathfrak{mod} \to R\text{-}\mathfrak{mod},$$

sending an object ${}_S C$ to the left R-module $B \otimes_S C$, and a morphism $f :_S C \to_S C'$ to the morphism

$$\mathrm{id}_B \otimes f :_R B_S \otimes_S {}_S C \to_R B_S \otimes_S {}_S C'.$$

Indeed, more generally still, you can think of

$$? \otimes_R ?$$

as a functor from $\mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod}$ to $\mathfrak{ab}$. Here, $\mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod}$ denotes the *product* of the categories $\mathfrak{mod}\text{-}R$ and $R\text{-}\mathfrak{mod}$. This is a fairly obvious notion: the objects in a product of two categories are simply pairs (A, B) of objects in each of the categories, while morphisms are pairs of morphisms in each of the categories. So $? \otimes_R ?$ maps an object

$$(A_R, {}_R B) \in \mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod}$$

to their tensor product

$$A_R \otimes_R {}_R B$$

which is an Abelian group. Given morphisms $f : A_R \to A'_R$ and $g :_R B \to_R B'$ (i.e. a morphism (f, g) in the category $\mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod}$), the functor $? \otimes ?$ maps (f, g) to the morphism

$$f \otimes g : A_R \otimes_R {}_R B \to A'_R \otimes_R {}_R B'$$

of Abelian groups. This is defined to be the map with $(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$ (of course, to check that there really is such a well-defined map you need to appeal to the universal property).

Okay, so now we understand that $\otimes$ is a functor, it is now possible to make precise what we meant when we said that the isomorphisms constructed in the examples (1)–(5) above are *natural* isomorphisms. For instance:

Associativity of tensor product. Let R, S be rings. We have the tensor functors

$$? \otimes_R ? : \mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod} \to \mathfrak{ab}$$

and

$$? \otimes_S ? : \mathfrak{mod}\text{-}S \times S\text{-}\mathfrak{mod} \to \mathfrak{ab}.$$

Composing them in two different ways gives two different functors $F = (? \otimes_R ?) \otimes_S ?$ and $G = ? \otimes_R (? \otimes_S ?)$ from $\mathfrak{mod}\text{-}R \times R\text{-}\mathfrak{mod}\text{-}S \times S\text{-}\mathfrak{mod}$ to $\mathfrak{ab}$. The correct statement now is that these two functors are isomorphic.

3.10 Projectives and injectives

Let R be a ring and consider the category R-$\mathtt{mod}$ again. We now introduce the notion of a *projective R-module*, which should be viewed as a generalization of free modules.

For the definition, an R-module P is called *projective* if for every exact sequence

$$B \xrightarrow{\pi} C \longrightarrow 0$$

and every homomorphism $\gamma : P \to C$, there exists a homomorphism $\beta : P \to B$ such that $\gamma = \pi \circ \beta$. In words, P is projective if every homomorphism from P to a quotient of an R-module B *lifts* to a homomorphism from P to B itself.

3.10.1. **Lemma.** *Let $P_i\,(i \in I)$ be R-modules. Then, $P = \bigoplus_{i\in I} P_i$ is projective if and only if each P_i is projective.*

Proof. Let $\pi : B \twoheadrightarrow C$ be an epimorphism and $\iota_i : P_i \to P$ be the canonical inclusions.

Suppose each P_i is projective. Take a map $\gamma : P \to C$. Writing $\iota_i : P_i \to P$ for the canonical inclusion, we get for each i a map $\gamma_i = \gamma \circ \iota_i : P_i \to C$. Since P_i is projective, we can lift γ_i to a map $\beta_i : P_i \to B$ with $\gamma_i = \pi \circ \beta_i$ for each i. Then the universal property of coproducts gives a unique map $\beta : P \to B$ such that $\beta_i = \beta \circ \iota_i$. Then, $\pi \circ \beta \circ \iota_i = \pi \circ \beta_i = \gamma_i = \gamma \circ \iota_i$ for each i, hence $\pi \circ \beta = \gamma$ as required.

Conversely, if P is projective and we have a map $\gamma_i : P_i \to C$, the universal property of coproduct gives us a unique map $\gamma : P \to C$ such that $\gamma_i = \gamma \circ \iota_i$ and $\gamma \circ \iota_j = 0$ for all $j \neq i$. Then, there is a lift $\beta : P \to B$ with $\gamma = \pi \circ \beta$. Define $\beta_i = \beta \circ \iota_i$ to obtain the required lift of γ_i showing that P_i is projective. □

3.10.2. **Corollary.** *Free R-modules are projective.*

Proof. Let F be a free R-module. Then, F is a direct sum of copies of the regular R module ${}_RR$. So by the lemma, we just need to show that ${}_RR$ is projective. Suppose $\pi : B \twoheadrightarrow C$ and we have a map $\gamma :{}_R R \to C$. Choose $b \in B$ such that $\pi(b) = \gamma(1_R)$ and define an R-module homomorphism $\beta :{}_R R \to B$ by $r \mapsto rb$ for $r \in R$. This is a lift of γ. □

Characterization of projectives. *Let P be an R-module. The following properties are equivalent.*

(1) *P is projective.*

(2) *Every short exact sequence of the form*

$$0 \longrightarrow K \longrightarrow M \longrightarrow P \longrightarrow 0$$

ending in P is split.

(3) *P is isomorphic to a summand of a free R-module.*

Proof. (1)⇒(2). Take a short exact sequence

$$0 \to K \to M \xrightarrow{\pi} P \to 0.$$

If P is projective, the idenity map $P \to P$ lifts to a map $\tau : P \to M$ such that $\pi \circ \tau = \mathrm{id}_P$. This is a splitting, so the short exact sequence is split.

(2)⇒(3). By Theorem 3.5.4, there is an epimorphism $\pi : F \to P$ where F is free. Letting $K = \ker \pi$, we obtain a short exact sequence

$$0 \to K \to F \xrightarrow{\pi} P \to 0.$$

By assumption it splits. So, $F \cong K \oplus P'$ with $P' \cong P$. Hence, P is isomorphic to a summand of a free module.

(3)⇒(1). If $P \oplus K$ is free, it is projective by Corollary 3.10.2 hence P is projective by Lemma 3.10.1. □

Examples. Using the characterization, we can give some examples of projectives.

(1) The regular $\mathbb{Z}_6$-module $\mathbb{Z}_6$ is free hence projective. By the Chinese remainder theorem, $\mathbb{Z}_6 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_3$ as a $\mathbb{Z}_6$-module. Hence, both $\mathbb{Z}_2$ and $\mathbb{Z}_3$ are projective $\mathbb{Z}_6$-modules. They are *not* free!!

(2) Let R is a PID and M be a finitely generated R-module. We showed that M is free if and only if it is torsion-free. Now, if M is free then it is projective. Conversely, if it is projective, then it is a summand of a free module, so is torsion-free, so is free. Hence: *finitely generated modules over a PID are free if and only if they are projective.*

(3) On the other hand, consider $\mathbb{Q}$ as a $\mathbb{Z}$-module. We have observed before that although $\mathbb{Q}$ is torsion-free, it is not free (allowed since it is not finitely generated). It is not projective either. Proof: suppose $\mathbb{Q} \oplus M$ is free for some $\mathbb{Z}$-module M. Then, it has a basis $\{f_i\}_{i \in I}$. Consider $1 \in \mathbb{Q} \subseteq \mathbb{Q} \oplus M$. Then, $1 = \sum_{i \in I} a_i f_i$ for integers a_i all but finitely many of which are zero. Let p be coprime to all the non-zero a_i. Write $1/p = \sum_{i \in I} b_i f_i$. Then, $1 = p(1/p) = \sum_{i \in I} p b_i f_i = \sum_{i \in I} a_i f_i$, hence since the f_i form a basis, we get that $pb_i = a_i$ for all i so $p | a_i$ for all i, a contradiction.

We turn now to discussing the dual notion of *injective modules*. An R-module I is called *injective* if for every exact sequence

$$0 \longrightarrow A \xrightarrow{i} B$$

and every R-module homomorphism $\alpha : A \to I$, there exists a homomorphism $\beta : B \to I$ such that $\alpha = \beta \circ i$. In words, I is injective if every homomorphism from a submodule of B to I extends to a homomorphism from B to I. In the category R-$\mathfrak{mod}$, the notion of injective is less useful that that of projective; but in many other subjects it is the notion of injective that is the dominant one.

3.10.3. **Lemma.** *Let I_j $(j \in J)$ be R-modules. Then, $I = \prod_{j \in J} I_j$ is injective if and only if each I_j is injective.*

Proof. This is exactly the same proof as Lemma 3.10.1 but carried out in the *opposite category* $(R\text{-}\mathfrak{mod})^{op}$! If you're not happy with this, you can easily dualize the proof of Lemma 3.10.1 yourself. □

Unfortunately at this point the development of injectives diverges from the development above of projectives. In short, there is no nice characterization of injectives analogous to "P is projective if and only if it is a summand of a free module". However, observe that Theorem 3.5.4 and Corollary 3.10.2 show that every R-module is a quotient of a projective module. There is a dual statement for injectives: every R-module is a submodule of an injective module. The proof takes a little longer.

Criterion for injectivity. *A left R-module I is injective if and only if for every left ideal L of R and every R-module homomorphism $f : L \to I$, there exists an extension of f to an R-module homomorphism $\bar{f} : R \to I$.*

Proof. Suppose I has the given property. Let $i : A \to B$ be any monomorphism and $\alpha : A \to I$. Let

$$\mathscr{H} = \{h : C \to I \mid i(A) \subseteq C \subseteq B, h \circ i = \alpha\}.$$

Note $\mathscr{H}$ is non-empty; partially order $\mathscr{H}$ so that $(h : C \to I) \leq (h' : C' \to I)$ if $C \subseteq C'$ and $h'|_C = h$.

To apply Zorn's lemma, take a chain $\{h_\omega : C_\omega \to I\}_{\omega \in \Omega}$. Set $C = \bigcup_{\omega \in \Omega} C_\omega$ and define $h : C \to I$ by $h(c) = h_\omega(c)$ for any $\omega \in \Omega$ for which $c \in C_\omega$. Then, $h : C \to I$ is an upper bound for the chain. Hence $\mathscr{H}$ has a maximal element, call it $h : C \to I$.

Now we claim that $C = B$, which will complete the proof. Well, suppose $C \neq B$ and pick $b \in B - C$. Let $L = \{r \in R \mid rb \in C\}$, a left ideal of R. Define a map $g : L \to I$ by $r \mapsto h(rb)$; this is an R-module homomorphism. So by assumption, there exists an R-module homomorphism $k : R \to I$ extending g. Now define $\bar{h} : C + Rb \to I$ so that $a + rb \mapsto h(a) + k(r)$. Then, this is a well-defined R-module homomorphism extending h, which contradicts the maximality of h. □

Divisible Abelian groups. Now we can describe the injective $\mathbb{Z}$-modules. Call an Abelian group *divisible* if for every $a \in A$ and every $n \in \mathbb{N}$, there exists an $x \in A$ such that $nx = a$. For example, $\mathbb{Q}$ is obviously divisible, but $\mathbb{Z}$ is not, nor is any finite Abelian group. We claim:

A is an injective $\mathbb{Z}$-module if and only if it is a divisible Abelian group.

Indeed, if A is injective, take $y \in A$ and $n \in \mathbb{N}$. Consider the exact sequence

$$0 \to n\mathbb{Z} \to \mathbb{Z}$$

Let $\alpha : n\mathbb{Z} \to A$ be the map $nz \mapsto yz$ Then there exists $\beta : \mathbb{Z} \to A$ extending α. Set $x = \beta(1)$. Then, $y = nx$ so A is divisible. Conversely, if G is divisible, use the criterion for injectivity and reverse the argument just given.

3.10.4. **Lemma.** *If A is an Abelian group, A can be embedded into a divisible Abelian group.*

Proof. Let $\pi : F \twoheadrightarrow A$ be an epimorphism, where F is free (Theorem 3.5.4) so that

$$F = \bigoplus_{i \in I} \mathbb{Z}_i,$$

a direct sum of copies of $\mathbb{Z}$. Embed each $\mathbb{Z}_i$ in a corresponding copy $\mathbb{Q}_i$ of $\mathbb{Q}$, to construct an embedding

$$i : F \hookrightarrow \bigoplus_{i \in I} \mathbb{Q}_i = \bar{F}.$$

Now, $A \cong F/\ker\pi \cong i(F)/i(\ker\pi)$. The right hand side is a submodule of $\bar{F}/i(\ker\pi)$, which is a quotient of a divisible Abelian group hence divisible. We thus obtain the required embedding of A into a divisible group. □

Now we can prove the main result:

Injective embedding theorem. *For any ring R, every (left) R-module M can be embedded into an injective R-module.*

Proof. In the first place, M is an Abelian group so by the lemma we can find a $\mathbb{Z}$-module monomorphism $f : M \hookrightarrow I$ where I is a divisible Abelian group. Consider the map

$$\bar{f} : \mathrm{Hom}_{\mathbb{Z}}(R, M) \to \mathrm{Hom}_{\mathbb{Z}}(R, I), \quad \theta \mapsto \bar{f}(\theta)$$

where $\bar{f}(\theta)(r) = f(\theta(r))$. If for any Abelian group we view $\mathrm{Hom}_{\mathbb{Z}}(R, A)$ as a left R-module by the rule $(r\theta)(s) = \theta(sr)$, $r, s \in R, \theta : R \to A$, then $\bar{f}$ is an R-module homomorphism. Now consider

$$M \to \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_{\mathbb{Z}}(R, M) \xrightarrow{\bar{f}} \mathrm{Hom}_{\mathbb{Z}}(R, I).$$

The first map here is the map $m \mapsto \theta_m$, where $\theta_m(r) = rm$ for $r \in R$. The second map here is the obvious inclusion. The third map $\bar{f}$ is as constructed above. We obtain an R-module monomorphism $M \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, I)$.

It just remains to check that the R-module $\mathrm{Hom}_{\mathbb{Z}}(R, I)$ is injective. We use the criterion for injectivity. So let $f : L \to \mathrm{Hom}_{\mathbb{Z}}(R, I)$ be a homomorphism, where L is a left ideal of R. Then, $\phi : a \mapsto (f(a))(1_R)$ is a $\mathbb{Z}$-module homomorphism $L \to I$. Since I is an injective $\mathbb{Z}$-module, there exists a $\mathbb{Z}$-module homomorphism $\bar{\phi} : R \to I$ extending ϕ. Now define

$$\bar{f} : R \to \mathrm{Hom}_{\mathbb{Z}}(R, I), \quad r \mapsto r\bar{\phi}.$$

Now you check that this is an R-module homomorphism extending f: for $r \in L$ and $s \in R$,

$$\bar{f}(r)(s) = (r\bar{\phi})(s) = (\bar{\phi})(sr) = \phi(sr) = f(sr)(1_R) = f(r)(s)$$

whence $\bar{f}(r) = f(r)$ for all $r \in L$. □

To illustrate the theorem, we can now prove a version of the characterization of projective modules above for injectives (but observe that there is now no case (3)!).

Characterization of injectives. *Let I be an R-module. The following properties are equivalent:*
(1) *I is injective.*
(2) *Every short exact sequence of the form*

$$0 \longrightarrow I \longrightarrow M \longrightarrow Q \longrightarrow 0$$

starting in I is split.

Proof. (1)⇒(2). Suppose I is injective and we are given a short exact sequence starting in I. Then, the identity map $I \to I$ extends to a map $j : M \to I$. This defines a splitting to show that the short exact sequence is split.

(2)⇒(1). Let I be a module with the given property. Apply the injective embedding theorem to find an embedding $i : I \hookrightarrow M$ with M injective. The short exact sequence splits, so $M = I \oplus Q'$ for some R-module Q'. In other words, $I \oplus Q' \cong I \times Q'$ is injective, so I is too, applying Lemma 3.10.3. □

3.11 Adjoint functors

So now we have introduced the tensor functors. The other important functors that arise in studying module categories are the *hom functors*. (Actually, we already made use of hom functors in the proof of the injective embedding theorem in section 3.10 – which should make more sense by the end of the present discussion!)

Let R be a ring and ${}_RM, {}_RN$ be left R-modules. We have already observed that

$$\mathrm{Hom}_R({}_RM, {}_RN)$$

is an Abelian group (R-$\mathfrak{mod}$ is an additive category). If ${}_RM$ has the additional structure of an R, S-bimodule for some new ring S, there ought to be some additional structure on

$$\mathrm{Hom}_R({}_RM_S, {}_RN).$$

Indeed, it is a *left* S-module. The left action of $s \in S$ on $f : M \to N$ is defined so that $sf : M \to N$ is the map with $(sf)(m) = f(ms)$ for all $m \in M$. You check associativity, for instance, with

$$((ss')f)(m) = f(m(ss')) = f((ms)s') = (s'f)(ms) = (s(s'f))(m).$$

Thus, the *right S-module structure on the first argument M leads to left S-module structure on* $\mathrm{Hom}_R(M, N)$. Instead, suppose that ${}_RN$ has the additional structure of an R, T-bimodule for some new ring T. Then this time,

$$\mathrm{Hom}_R({}_RM, {}_RN_T)$$

has additional structure of a *right* T-module. The right action of $t \in T$ on $f : M \to N$ is defined by $(ft)(m) = f(m)t$ for all $m \in M$. Thus, the *right T-module structure on the second argument N leads to right T-module structure on* $\mathrm{Hom}_R(M, N)$. Putting both cases together, if M is an R, S-bimodule and N is an R, T-bimodule, then

$$\mathrm{Hom}_R({}_RM_S, {}_RN_T)$$

is actually an S, T-bimodule. You could also consider this all for *right* R-modules instead. Then, the Abelian group

$$\mathrm{Hom}_R({}_SM_R, {}_TN_R)$$

is actually a T, S-bimodule: the left S-module structure on the first argument M leads to right S-modue structure on the hom space, while the left T-module structure on the second argument N leads to left T-module structure overall.

Example. Let N be an R,T-bimodule. Consider the space

$$\mathrm{Hom}_R({}_RR_R, {}_RN_T)$$

of homomorphisms of left R-modules. We have explained above how to view this as an R,T-bimodule. We claim that in fact,

$$\mathrm{Hom}_R({}_RR_R, {}_RN_T) \cong_R {}_RN_T$$

as an R,T-bimodule. Indeed, the isomorphism is given in the forward direction by the map Ev where $Ev(f) = f(1_R)$ – thus, Ev is "evaluation at 1_R". Let us check that evaluation really is an R,T-bimodule homomorphism:

$$Ev(rft) = (rft)(1_R) = f(1_Rr)t = f(r1_R)t = rf(1_R)t = rEv(f)t$$

for every $f \in \mathrm{Hom}_R(R,N)$, $r \in R$ and $t \in T$, as required. It remains to see that Ev is bijective, which we do by exhibiting an inverse map $\gamma : N \to \mathrm{Hom}_R(R,N)$. Given $n \in N$, define $\gamma(n) : R \to N$ by $\gamma(n)(r) = rn$. Since R is free R-module, $\gamma(n)$ really is an R-module homomorphism from ${}_RR$ to ${}_RN$. Then,

$$Ev(\gamma(n)) = \gamma(n)(1_R) = n$$

and

$$\gamma(Ev(f))(r) = \gamma(f(1_R))(r) = rf(1_R) = f(r1_R) = f(r),$$

showing that γ and Ev are two-sided inverses to each other.

Now we explain how $\mathrm{Hom}_R({}_RM_S, ?)$ is an (additive) functor from the category of left R-modules to the category of left S-modules. We have already explained how given a left R-module ${}_RN$,

$$\mathrm{Hom}_R({}_RM_S, {}_RN)$$

is a left S-module. So it remains to see that if $f :_R N \to_R N'$ is a morphism of left R-modules, we can define a morphism

$$\mathrm{Hom}_R(M,f) : \mathrm{Hom}_R({}_RM_S, {}_RN) \to \mathrm{Hom}_R({}_RM_S, {}_RN')$$

so that $\theta : M \to N$ maps to $f \circ \theta : M \to N'$. Thus, $\mathrm{Hom}_R({}_RM_S, ?)$ is a (covariant) functor from the category of left R-modules to the category of left S-modules.

You should not be surprised now that $\mathrm{Hom}_R(?, {}_RN_S)$ is also a functor from the category of left R-modules, but this time to the category of right S-modules, because each

$$\mathrm{Hom}_R({}_RM, {}_RN_S)$$

is a right S-module. Given a morphism $f :_R M \to_R M'$ of left R-modules, we define

$$\mathrm{Hom}_R(f,N) : \mathrm{Hom}_R({}_RM', {}_RN_S) \to \mathrm{Hom}_R({}_RM, {}_RN_S)$$

to be the right S-module homomorphism sending $\theta : M' \to N$ to the map $\theta \circ f : M \to N$. Thus, $\mathrm{Hom}_R(?,N)$ is a functor – but this time it is a *contravariant functor* because the direction of the morphism $f : M \to M'$ was reversed to give the morphism $\mathrm{Hom}_R(f,N) : \mathrm{Hom}_R(M',N) \to \mathrm{Hom}_R(M,N)$. Putting both constructions together, we obtain a functor

$$\mathrm{Hom}_R(?,?) : (R\text{-}\mathfrak{mod})^{\mathrm{op}} \times R\text{-}\mathfrak{mod} \to \mathfrak{ab}.$$

Thus, the functor $\mathrm{Hom}_R(?,?)$ is contravariant in the first argument and covariant in the second. This is a little more complicated than the tensor functor $?\otimes_R?$ which was covariant in both arguments.

So now we have introduced both the tensor and the hom functors. The connection between the two comes from the following fundamental theorem:

Adjointness of tensor and hom. *Given rings R, S and modules ${}_RM, {}_SP_R, {}_SN$, there is a natural isomorphism*

$$\tau : \mathrm{Hom}_S({}_SP_R \otimes_R {}_RM, {}_SN) \xrightarrow{\sim} \mathrm{Hom}_R({}_RM, \mathrm{Hom}_S({}_SP_R, {}_SN))$$

of Abelian groups.

Proof. Given $f : P \otimes_R M \to N$, define $\tau f : M \to \mathrm{Hom}_S(P, N)$ so that $(\tau f)(m)$ is the function

$$p \mapsto f(p \otimes m)$$

for each $p \in P$.

There are now many things to check! For instance, we need to see that each $(\tau f)(m)$ is a left S-module morphism: $(\tau f)(m)(sp) = f(sp \otimes m) = sf(p \otimes m) = s((\tau f)(m)(p))$. Then we need to see that $\tau f : M \to \mathrm{Hom}_S(P, N)$ is a left R-module homomorphism, so that τf is really an element of $\mathrm{Hom}_R(M, \mathrm{Hom}_S(P, N))$. Well, $(\tau f)(rm)(p) = f(p \otimes rm) = f(pr \otimes m) = (\tau f)(m)(pr) = (r((\tau f)(m)))(p)$ hence $(\tau f)(rm) = r((\tau f)(m))$ as required. Finally we need to see that τ is a bijection. For this, we construct a two-sided inverse

$$\sigma : \mathrm{Hom}_R({}_RM, \mathrm{Hom}_S({}_RP_S, {}_SN)) \to \mathrm{Hom}_S({}_SP_R \otimes_R {}_RM, {}_SM).$$

To define σ, take $f : M \to \mathrm{Hom}_S(P, N)$. Define a map $P \times M \to N$ by $(p, m) \mapsto (f(m))(p)$. This is balanced, so induces by the universal property of tensor product a unique map $\sigma(f) : P \otimes_R M \to N$. This defines σ, and now you check that it is the two-sided inverse to τ. □

The statement of the theorem used the word "natural". I want to explain what this means precisely shortly, but first let us make an abstract categorical definition because I think in this case the abstraction helps to understand the significance of the theorem just proved.

Let $\mathscr{A}, \mathscr{B}$ be two categories and $F : \mathscr{A} \to \mathscr{B}$ and $G : \mathscr{B} \to \mathscr{A}$ be functors. Then, (F, G) is called an *adjoint pair* if for each pair of objects $X \in \mathscr{A}$ and $Y \in \mathscr{B}$, there is a natural bijection

$$\mathrm{Hom}_{\mathscr{B}}(FX, Y) \cong \mathrm{Hom}_{\mathscr{A}}(X, GY).$$

The crucial word here is "natural". To interpret its meaning, we need to view

$$\mathrm{Hom}_{\mathscr{B}}(F?, ?) : \mathscr{A}^{\mathrm{op}} \times \mathscr{B} \to \mathfrak{sets}$$

as a *functor*. Indeed, we understand how to plug an object in $\mathscr{A}$ into the first ? and an object in $\mathscr{B}$ into the second ? to get a set at the end. Moreover, we understand how to plug morphisms in too to get set maps (bearing in mind that $\mathrm{Hom}_{\mathscr{B}}(F?, ?)$ is contravariant in the first argument). Similarly,

$$\mathrm{Hom}_{\mathscr{A}}(?, G?) : \mathscr{A}^{\mathrm{op}} \times \mathscr{B} \to \mathfrak{sets}$$

is a functor. Then, saying that (F, G) is an adjoint pair really means that there is an isomorphism of functors

$$\mathrm{Hom}_{\mathscr{B}}(F?, ?) \cong \mathrm{Hom}_{\mathscr{A}}(?, G?).$$

Again, this means that for each pair of objects $(X, Y) \in \mathscr{A} \times \mathscr{B}$ there exists a bijection (an isomorphism of sets!)

$$\eta_{X,Y} : \mathrm{Hom}_{\mathscr{B}}(FX, Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathscr{A}}(X, GY)$$

such that for arrows $f : X' \to X$ in $\mathscr{A}$ and $g : Y \to Y'$ in $\mathscr{B}$, we have that

$$\mathrm{Hom}_{\mathscr{A}}(f, Gg) \circ \eta_{X,Y} = \eta_{X',Y'} \circ \mathrm{Hom}_{\mathscr{B}}(Ff, g)$$

as maps from $\mathrm{Hom}_{\mathscr{B}}(FX, Y)$ to $\mathrm{Hom}_{\mathscr{A}}(X', GY')$ (draw the diagram!).

We will go back in a moment to the above theorem ("adjointness of tensor and hom"). But let's start with some *easier* examples of adjoint pairs of functors...

Examples of adjoint pairs. (1) Let $i : \mathfrak{ab} \to \mathfrak{groups}$ be the "inclusion" functor. Let $\alpha : \mathfrak{groups} \to \mathfrak{ab}$ be the "Abelianization" functor. Thus, $\alpha(G) = G/G'$, and if $f : G \to H$ is a group homomorphism, $\alpha f : G/G' \to H/H'$ is induced by the map $\pi \circ f : G \to H/H'$ factored through G' (H/H' is Abelian so $\pi \circ f$ maps G' to $\{1\}$). Then, there is a natural isomorphism

$$\mathrm{Hom}_{\mathfrak{groups}}(G, iA) \cong \mathrm{Hom}_{\mathfrak{ab}}(G/G', A)$$

so that (α, i) is an adjoint pair of functors. One calls the left hand functor α the *left* adjoint and the right hand functor i the *right* adjoint in the pair.

(2) Let $F : \mathfrak{sets} \to \mathfrak{groups}$ be the functor sending a set X to the free group on X. Let $G : \mathfrak{groups} \to \mathfrak{sets}$ be the forgetful functor. Then, the universal property of free groups gives us a natural isomorphism

$$\mathrm{Hom}_{\mathfrak{groups}}(FX, H) \cong \mathrm{Hom}_{\mathfrak{sets}}(X, GH)$$

so again (F, G) is an adjoint pair of functors. In fact, most definitions by universal property can be interpreted in terms of an adjoint pair of functors in this way.

(3) Let $F : \mathscr{A} \to \mathscr{B}$ be an equivalence of categories. Recall this means there is another functor $G : \mathscr{B} \to \mathscr{A}$ such that $F \circ G \cong \mathrm{Id}_{\mathscr{B}}$ and $G \circ F \cong \mathrm{Id}_{\mathscr{A}}$. You can check that (F, G) (and similarly (G, F)) is an adjoint pair of functors. This is one reason for being interested in adjoint pairs: if one is trying to construct equivalences of categories, good candidates for the functors should be an adjoint pair to start with.

Idea of proof: The functor G induces a natural transformation between the functors

$$\eta : \mathrm{Hom}_{\mathscr{B}}(F?, ?) \to \mathrm{Hom}_{\mathscr{A}}(?, G?),$$

where for a given objects M, N the map $\eta_{M,N}$ defining the natural transformation η comes from

$$\mathrm{Hom}_{\mathscr{B}}(FM, N) \to \mathrm{Hom}_{\mathscr{A}}(GFM, GN) \cong \mathrm{Hom}_{\mathscr{A}}(M, GN).$$

The first map is defined simply by $\theta \mapsto G\theta$. The second map is defined using the given isomophism $GF \cong \mathrm{Id}$. To see this is an isomorphism, we write down the inverse. This turns out to be

$$\sigma : \mathrm{Hom}_{\mathscr{A}}(?, G?) \to \mathrm{Hom}_{\mathscr{B}}(F?, ?),$$

where for a given objects M, N the map $\sigma_{M,N}$ defining the natural transformation σ comes from

$$\mathrm{Hom}_{\mathscr{A}}(M, GN) \to \mathrm{Hom}_{\mathscr{B}}(FM, FGN) \cong \mathrm{Hom}_{\mathscr{B}}(FM, N).$$

(4) Now consider tensor and hom. The theorem proved above gives that for rings R, S and modules ${}_RM, {}_RP_S, {}_SN$, there is a natural isomorphism

$$\tau : \mathrm{Hom}_S(P \otimes_R M, N) \to \mathrm{Hom}_R(M, \mathrm{Hom}_S(P, N)).$$

This can be reinterpreted as saying that $(P \otimes_R ?, \mathrm{Hom}_S(P, ?))$ *is an adjoint pair of functors.* Informally, "tensor is left adjoint to hom" or "hom is right adjoint to tensor".

Now we have the basic notion of an adjoint pair of functors, I am going just to *state without proof* some of the basic (but remarkably useful) properties which follow by general nonsense just from the additive adjoint pair assumption. I will state these properties just in the case of an adjoint pair of functors between module categories. The correct setting really for these results is that of an additive adjoint pair of functors between arbitrary *Abelian categories.* But this material is taking us far enough as it is...

We need one more definition. Let $F : R\text{-}\mathfrak{mod} \to S\text{-}\mathfrak{mod}$ be a (covariant) functor, for rings R, S. Then, F is called *left exact* if it is additive and moreover exactness of any sequence

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C$$

in R-$\mathfrak{mod}$ implies exactness of the sequence

$$0 \longrightarrow FA \xrightarrow{F\alpha} FB \xrightarrow{F\beta} FC$$

in S-$\mathfrak{mod}$. Similarly, F is called *right exact* if it is additive and exactness of any sequence

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

in R-$\mathfrak{mod}$ implies exactness of the sequence

$$FA \xrightarrow{F\alpha} FB \xrightarrow{F\beta} FC \longrightarrow 0$$

in S-$\mathfrak{mod}$. Finally, F is called *exact* if it is both left and right exact. It is an exercise to check that F being exact as just defined is *equivalent* to saying that F is additive and exactness of any sequence

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

in R-$\mathfrak{mod}$ implies exactness of the sequence

$$0 \longrightarrow FA \xrightarrow{F\alpha} FB \xrightarrow{F\beta} FC \longrightarrow 0$$

in S-$\mathfrak{mod}$.

Properties of adjoint pairs. *Let $F : R\text{-}\mathfrak{mod} \to S\text{-}\mathfrak{mod}$ and $G : S\text{-}\mathfrak{mod} \to R\text{-}\mathfrak{mod}$ be functors such that (F, G) is an adjoint pair. Then:*

(0) *[Maclane[1], p.83] Any other functor G' which is right adjoint to F is isomorphic to G; any other functor F' which is left adjoint to G is isomorphic to F;*

(1) *[Maclane, p.83 and p.193] F and G are automatically additive functors, and the natural bijection*

$$\mathrm{Hom}_S(FM, N) \cong \mathrm{Hom}_R(M, GN)$$

arising from the adjunction (which by definition is only an isomorphism as sets) is necessarily an isomorphism as Abelian groups.

(2) *[Rotman[2], p.39] F is right exact; G is left exact;*

(3) *[Rotman,p.47,p.55] F commutes with arbitrary (not necessarily finite) direct sums; G commutes with arbitrary direct products;*

(4) *[Exercise!] If F is in fact exact (not just right exact), then G sends injective S-modules to injective R-modules; if G is in fact exact (not just left exact), then F sends projective R-modules to projective S-modules.*

Note if F, G are quasi-inverse equivalences, we have observed that both (F, G) and (G, F) are adjoint pairs. So in this special case, these properties imply some things which are not hard to prove directly in this case, but worth knowing: equivalences of categories are exact, they commute with direct sums and products, and they send injectives/projectives to injectives/projectives.

Now let us consider the consequences of these properties in the case which interests us: the adjoint pair $(P\otimes_R ?, \mathrm{Hom}_S(P, ?))$. We get for any rings R, S that:

(1) For any left S-module P, the functor $\mathrm{Hom}_S(P, ?)$ is left exact and commutes with arbitrary direct products (because it is right adjoint to something);

(2) For any right R-module P, the functor $P\otimes_R ?$ is right exact and commutes with arbitrary direct sums (because it is left adjoint to something);

[1]The reference is to Maclane's book "Categories for the working mathematician"

[2]Rotman's book "Introduction to homological algebra"

Remark. Let me note that we have worked here with left modules. We could just as well have worked with right modules, using the adjoint pair $(? \otimes_R {}_R P_S, \mathrm{Hom}_S({}_R P_S, ?))$. Then we would have deduced:

$(1)'$ For any right S-module P, the functor $\mathrm{Hom}_S(P, ?)$ is left exact and commutes with arbitrary products;

$(2)'$ For any left R-module ${}_R P$, the functor $? \otimes_R P$ is right exact and commutes with arbitrary coproducts;

Now it is reasonable to ask the following questions:

(Q1) When is the functor $\mathrm{Hom}_R({}_R P, ?)$ exact (not just left exact)?

(Q2) When is the functor $P_S \otimes_S ?$ exact (not just right exact)?

We can easily answer (Q1):

3.11.1. **Lemma.** *The functor* $\mathrm{Hom}_R({}_R P, ?)$ *is exact if and only if* ${}_R P$ *is a projective left* R*-module.*

Proof. Take a short exact sequence

$$0 \longrightarrow K \longrightarrow M \xrightarrow{f} Q \longrightarrow 0.$$

Since $\mathrm{Hom}_R({}_R P, ?)$ is always left exact, we just need to ask when

$$\mathrm{Hom}_R(P, M) \xrightarrow{\hat{f}} \mathrm{Hom}_R(P, Q)$$

is surjective, where $\hat{f}$ is defined by $(\hat{f})(\theta) = f \circ \theta$. Saying $\hat{f}$ is surjective is equivalent to saying that for every homomorphism $\phi : P \to Q$, there exists a homomorphism $\theta : P \to M$ such that $\phi = f \circ \theta$. In other words, every homomorphism from P to Q *lifts* through f to a homomorphism from P to M. This is exactly the definition of what it means for P to be a projective R-module! □

On the other hand, we do not know an answer to question (Q2) at present. So instead, let us *define* a right S-module P_S to be *flat* if the functor $P_S \otimes_S ?$ is exact (not just right exact). This introduces another very important sort of modules.

Example. $\mathbb{Z}_n$ *is not a flat* $\mathbb{Z}$*-module.* Proof. Take the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{f} \mathbb{Z}$$

where f is the map determined by multiplication by n. Applying the functor $\mathbb{Z}_n \otimes_{\mathbb{Z}} ?$, we get the sequence

$$0 \longrightarrow \mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Z} \xrightarrow{\bar{f}} \mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Z}$$

where the map $\bar{f}$ is multiplication by n. In other words $\bar{f}$ is zero!!! So $\bar{f}$ is no longer injective, and our sequence is no longer exact. So $\mathbb{Z}_n$ is not flat.

We do at least have:

3.11.2. **Lemma.** *If* P_S *is a projective* S*-module, then it is flat.*

Sketch. We need to show that $P_S \otimes_S ?$ sends injective homomorphisms to injective homomorphisms. Since P_S is projective, it is a summand of $\bigoplus_{i \in I} S_S$ for some I. So it suffices (with a little more care than I'm explaining in this sketch) to show that

$$\left(\bigoplus_{i \in I} S_S \right) \otimes_S ?$$

sends injective homomorphisms to injective homomorphisms. In turn, since $\otimes$ commutes with direct sums, this reduces (again you need to think exactly why) to checking just that

$$S_S \otimes_S ?$$

sends injective homomorphisms to injective homomorphisms. But this is clear, since $S_S \otimes_S {}_S M \cong_S M$ for any left S-module ${}_S M$. □

But *the converse to the lemma is false in general.* For example, $\mathbb{Q}$ is a flat $\mathbb{Z}$-module but is not projective.

In a course on homological algebra you will introduce Ext and Tor which in some sense measure how far modules are from being projective or flat...

3.12 Morita equivalence

Throughout the section, let R and S be rings. We want to try to compare R and S by comparing their module categories.

3.12.1. **Lemma.** *Let F and G be functors from R-$\mathfrak{mod}$ to S-$\mathfrak{mod}$ which are right exact and commute with arbitrary direct sums. Suppose that $\eta : F \to G$ is a natural transformation of functors such that $\eta_R : FR \to GR$ is an isomorphism of right S-modules. Then, η is an isomorphism of functors.*

Sketch. Take $M \in R$-$\mathfrak{mod}$. There exists a surjection $S_1 \xrightarrow{f} M$ where S_1 is a direct sum of copies of R. Again, there exists a surjection $S_2 \xrightarrow{g} \ker f$ where S_2 is a direct sum of copies of R. Putting together, we have constructed an exact sequence

$$S_2 \xrightarrow{g} S_1 \xrightarrow{f} M \longrightarrow 0.$$

(This is a *presentation* of the R-module M.) Now apply the right exact functors F and G in turn to obtain two exact sequences

$$FS_2 \longrightarrow FS_1 \longrightarrow FM \longrightarrow 0$$

and

$$GS_2 \longrightarrow GS_1 \longrightarrow GM \longrightarrow 0.$$

Moreover, the natural transformation η gives us vertical maps $\eta_2 : FS_2 \to GS_2$, $\eta_1 : FS_1 \to GS_1$ and $\eta_M : FM \to GM$ so that the resulting diagram commutes. We want to show that η_M is an isomorphism. Using the five lemma, it suffices to show that both η_1 and η_2 are isomorphisms.

Well,

$$FS_2 = F(\bigoplus R) \cong \bigoplus FR$$

since F commutes with arbitrary direct sums. Similarly,

$$GS_2 = G(\bigoplus R) \cong \bigoplus GR.$$

By the assumption η_R is an isomorphism between FR and GR. You get from this using naturality of η that η_2 is an isomorphism between FS_2 and GS_2. Similarly, η_1 is an isomorphism. □

3.12.2. **Lemma.** *A functor $F : R$-$\mathfrak{mod} \to S$-$\mathfrak{mod}$ has a right adjoint if and only if $F \cong {}_S P_R \otimes_R ?$ for some S, R-bimodule P.*

Proof. ($\Leftarrow$). The functor $P \otimes_R ?$ has a right adjoint, namely, $\mathrm{Hom}_S(P, ?)$, by the adjointness of tensor and hom.

($\Rightarrow$). Suppose $F : R\text{-}\mathfrak{mod} \to S\text{-}\mathfrak{mod}$ has a right adjoint $G : S\text{-}\mathfrak{mod} \to R\text{-}\mathfrak{mod}$, i.e. that (F, G) is an adjoint pair. Then, we get from the general properties of adjoint pairs that F is right exact and commutes with arbitrary direct sums.

Let ${}_SP = F({}_RR)$, a left S-module. It is even an S, R-bimodule. Indeed, given $r \in R$, right multiplication by r defines a left S-module homomorphism $\rho_r : {}_RR \to {}_RR$. So applying the functor F, we obtain a left S-module homomorphism

$$F\rho_r : {}_SP \to {}_SP.$$

Now define the right action of $r \in R$ on ${}_SP$ by

$$pr = (F\rho_r)(p)$$

for each $p \in P$. This makes ${}_SP$ into an S, R-bimodule ${}_SP_R$. We know by the general theory of adjoint pairs that the functor ${}_SP_R\otimes?$ is right exact and commutes with arbitrary direct sums.

Now we claim that $F \cong {}_SP_R\otimes_R?$, which will complete the proof. First, we construct a natural transformation between the functors. Given ${}_RM \in R\text{-}\mathfrak{mod}$, we define a map

$$\eta_M \in \mathrm{Hom}_R(M, \mathrm{Hom}_S(P, FM))$$

by the composite

$$M \xrightarrow{\sim} \mathrm{Hom}_R(R, M) \longrightarrow \mathrm{Hom}_S(FR, FM) = \mathrm{Hom}_S(P, FM).$$

Here, the first isomorphism is the canonical one and the second homomorphism is the map induced by the functor F. Now, by adjointness of tensor and hom, there is a canonical isomorphism

$$\mathrm{Hom}_R(M, \mathrm{Hom}_S(P, FM)) \cong \mathrm{Hom}_S(P \otimes_R M, FM)$$

so we can view η_M instead as a map $P \otimes_R M \to FM$. Then, η defines a natural transformation from the functor $P\otimes_R?$ to the functor $F?$.

Now to show that η is an isomorphism of functors it suffices applying Lemma 3.12.1 just to see that $\eta_R : P \otimes_R R \to FR = P$ is an isomorphism. But by definition η_R is just the canonical isorphism between $P \otimes_R R$ and P. □

Now we obtain the main result of the section:

Morita theorem (weak version). *Let R and S be rings. The following properties are equivalent:*

(i) *The categories $R\text{-}\mathfrak{mod}$ and $S\text{-}\mathfrak{mod}$ are equivalent.*

(ii) *The categories $\mathfrak{mod}\text{-}R$ and $\mathfrak{mod}\text{-}S$ are equivalent.*

(iii) *There exists bimodules ${}_SP_R$ and ${}_RQ_S$ with*

$${}_SP_R \otimes_R {}_RQ_S \cong {}_SS_S, \qquad {}_RQ_S \otimes_S {}_SP_R \cong {}_RR_R$$

as bimodules.

Proof. I just prove the equivalence of (i) and (iii), the equivalence of (ii) and (iii) being similar.

(iii)$\Leftarrow$(i). Let $F = {}_SP_R\otimes_R?$ and $G = {}_RQ_S\otimes_S?$. Then,

$$F \circ G = P \otimes_R (Q\otimes_S?) \cong (P \otimes_R Q)\otimes_S? \cong S\otimes_S? \cong \mathrm{id},$$

using associativity of the tensor product functor. Similarly, $G \circ F \cong \mathrm{Id}$. Hence, the categories are equivalent.

(i)$\Rightarrow$(iii). If the categories are equivalent, then there exist functors $F : R\text{-}\mathfrak{mod} \to S\text{-}\mathfrak{mod}$ and $G : S\text{-}\mathfrak{mod} \to R\text{-}\mathfrak{mod}$ with $F\circ G \cong \mathrm{Id}$ and $G\circ F \cong \mathrm{Id}$. We have already observed that equivalences of categories are adjoint pairs (F, G) and (G, F). So applying Lemma 3.12.2 twice, we get bimodules

${}_SP_R$ and ${}_RQ_S$ with $F \cong {}_SP_R \otimes_R ?$ and $G \cong {}_RQ_S \otimes_S ?$. Then, since $FG \cong \mathrm{Id}$, applying to the module ${}_SS_S$ we get that

$$P \otimes_R Q \otimes_S S \cong S$$

hence $P \otimes_R Q \cong S$ and similarly $Q \otimes_S P \cong R$ as bimodules. □

Now call the rings R and S *Morita equivalent* if any of the equivalent properties in the Morita theorem hold.

Example. This is really the crucial example! Let R be any ring and $S = M_n(R)$, the ring of $n \times n$ matrices with entries in R. I claim that R and S are Morita equivalent rings.

Proof. Let ${}_RP_S$ be the R, S-bimodule consisting of all row vectors of the form $(a_1 \dots a_n)$ for $a_i \in R$. It is a left R-module by multiplication, and a right S-module by matrix multiplication. Similarly, let ${}_SQ_R$ be the S, R-bimodule consisting of all column vectors under matrix multiplication.

I claim that

$${}_SQ_R \otimes_R {}_RP_S \cong_S S_S$$

and

$${}_RP_S \otimes_S {}_SQ_R \cong_R R_R$$

as bimodules. Indeed, the isomorphisms in each case are just defined by multiplication. So in the first case, you take the map $Q \times P \to S$ given by sending a pair (c, r) (where c is a column vector and r is a row vector) to their product (which is an $n \times n$ matrix). This map is balanced so extends to a unique S, S-bimodule homomorphism as stated. Similarly in the second case, the map is induced by the universal property of tensor by the map $P \times Q \to R$ given by multiplication, this time the row vector times the column vector which gives a 1×1 matrix! In each case, it remains to check that the maps defined are indeed bijective. I leave this as an exercise.

Hence by (the easy implication in) the Morita theorem, R and S are Morita equivalent rings.

Having defined the relation "Morita equivalence" on the category of rings, it is reasonable to ask what sort of ring theoretic properties are preserved by Morita equivalences. In other words, what are the *Morita invariants* of rings. I give two examples, the first with proof the second without. Hopefully, these are enough to convince you that Morita equivalence – something defined entirely in terms of the modules of a ring not the ring itself – is a useful thing to consider.

First example of a Morita invariant property. *If R and S are Morita equivalent rings, then the lattice of two-sided ideals of R is isomorphic to the lattice of two-sided ideals of S.*

Sketch. By the Morita theorem, there exist bimodules ${}_RP_S$ and ${}_SQ_R$ such that the functors

$$? \otimes_R {}_RP_S \qquad \text{and} \qquad ? \otimes_S {}_SQ_R$$

give the mutually inverse equivalences. Applying $? \otimes_R {}_RP_S$ to a two-sided ideal in ${}_RR_R$ gives an R, S-subbimodule of $R \otimes_R P = P$. Conversely applying $? \otimes_S {}_SQ_R$ to an R, S-subbimodule of P gives a two-sided ideal in R. Since the two functors are inverse to each other, you deduce using naturality that they induce a lattice isomorphism between the lattice of two-sided ideals of R and the lattice of R, S-subbimodules of P.

Now repeat the argument instead with the functors

$${}_RP_S \otimes_S ? \qquad \text{and} \qquad {}_SQ_R \otimes_R ?$$

to deduce this time that there is a lattice isomorphism between the lattice of two-sided ideals of S and the lattice of R, S-subbimodules of P. Combining the conclusion in each of these two paragraphs completes the proof. □

For example if R is a simple ring (only the trivial two-sided ideals) and R is Morita equivalent to S then S is a simple ring too.

Second example of a Morita invariant property. You can also show – with not too much extra work – that if R and S are Morita equivalent rings, then the *centers* of the rings R and S are isomorphic. For instance, since R and $M_n(R)$ are isomorphic, the centers of these rings are isomorphic – in this special case this is easy to prove directly!

(Hint as to the proof: look at the identity functor Id : R-$\mathfrak{mod} \to R$-$\mathfrak{mod}$. There is a natural ring structure on the set of endomorphisms End(Id) of this functor. Now you prove that $Z(R) \cong$ End(Id). This gives a description of the center $Z(R)$ purely in terms of the module category, i.e. something that is preserved under Morita equivalence.)

In particular, two *commutative rings* R, S are Morita equivalent if and only R and S are in fact already isomorphic rings themselves. This shows that for commutative rings, *everything about the ring is encoded in the category of modules over the ring.* Perhaps this convinces you of the value of studying modules as an indirect way to get at the structure of a ring.

By the way it is interesting to point out some things that are *not* preserved by Morita equivalence. For example if you have an equivalence R-$\mathfrak{mod} \to S$-$\mathfrak{mod}$, it need not send *free* R-modules to *free* S-modules... But it does send projectives to projectives which is why projectives are a more natural notion than free objects.