Introduction to Module Theory

Piyush Patil

February 3, 2017

This chapter will cover the basics of a mathematical object known as modules, which are essentially generalizations of the notion of vector spaces. That is, modules are to vector spaces what rings are to fields.

1 Basic Definitions

First recall that a *group* is one of the most basic algebraic structures - it's a set equipped with a binary operator that satisfies the four *group axioms*: closure, associativity, identity, and invertibility. It's the barest algebraic structure, with the fewest axioms, under which the resulting theory is still general and non-pathological enough to merit attention, and many seemingly unrelated structures throughout mathematics are related to groups in some way.
Recall that a *ring* is simple a group with sufficient additional structure to accomodate a second binary operator and still retain all of the nice properties we want the structure to have. Specifically, a ring is a commutative group equipped with a second associative binary operator (which isn't necessarily commutative) that distributes over the first (so it's "compatible" in a sense with it). Rings with a multiplicative identity are called *unital*, and rings in which the second binary operator is commutative are referred to as commutative. Note that the only unital ring where the additive and multiplicative identites coincide is the degenerative ring with one element (namely, the identity). As is convention, we denote the additive identity of a ring R as 0_R, or simply 0 when context is obvious, and the multiplicative identity, should it exist, with 1_R.
Whereas rings are abelian groups under the first operation, and only require that the second operation be compatible with the first, recall that a *field* is a ring with more strict requirements - it must be an commutative group not only over the first binary operation, but over the second as well (after throwing out the additive identity, that is, since we don't want to divide by zero). In other words, both binary operations must satisfy the group axioms, with the caveat that the second need not do so for the additive identity.

Let's jump into the definition of a module.

(Definition) Module: *Let R be a ring. We define a* ***left R-module****, also known as a* ***left module*** *over R, as an abelian (under a binary operator +) group M with an action of R on M, denoted rm for $r \in R$ and $m \in M$, which satisfies*

1. *Distributivity:* $\forall r, s \in R, m, n \in M : (r+s)m = rm + sm \quad and \quad r(m+n) = rm + rn$
2. *Associativity:* $\forall r, s \in R, m \in M : (rs)m = r(sm)$
3. *Identity holds: If R has a 1, then* $\forall m \in M : 1m = m$

• *Motivation*: Modules are meant to generalize the notion of vector spaces over fields, using elements of the underlying ring as their "scalars". Recall that group actions on a set can be viewed as homomorphisms from the group to the permutation group of the set, so that a group action is really just a way of looking at the elements of the group as functions (bijections in fact) on the set, which "act" on the elements of the set. Intuitively, the elements of the group act on the set by moving its members around in different ways, and are constrained to obey the above principles, such as associativity and identity preservation. With vector spaces over a field, the underlying field of scalars acts on vectors through scalar multiplication, in a way that's constrained to obey certain principles, such as associativity and distributivity, which ensure some algebraic structure on the set of actions. A module weakens some of the assumptions the vector space definition makes, by requiring the underlying scalars to only be a ring, not necessarily a field, and generalizes scalar multiplication to any group action.

Right modules are defined the same way, but where ring elements appear on the right in the group action instead of the left. We often refer to the elements of R as *scalars*. For convenience we refer to modules over unital rings as unital. Moreover, when the ring is commutative, we can use the same group action for both right and left modules, in which case we simply drop the distinction and call it an R-module. All the modules we'll study here are assumed to be unital (non-unital modules are somewhat pathological). As one would expect, when R happens to also be a field, the definition of a module coincides with the definition of a vector space. We define *submodules* to be subsets that are also modules. When considering rings as modules over themselves (where the group action is the ordinary multiplication operator), the submodules are precisely the

ring's ideals.
Let's now take a look at some specific modules, which are important enough to merit special attention. The first module is constructed with the ring of integers and any abelian group G. Denoting G's binary operation with $+$, define the group action

$$\forall n \in \mathbb{Z}, g \in G : ng = \begin{cases} g + \cdots + g \ (n \text{ times}) & \text{if } n > 0 \\ 0 \ (G\text{'s identity}) & \text{if } n = 0 \\ (-g) + \cdots + (-g) \ (-n \text{ times}) & \text{if } n < 0 \end{cases}$$

where $-g$ is the inverse of g. This is a direct generalization of integer multiplication that allows integers to act on abstract group elements in the same way. This makes G into a $\mathbb{Z}$-module, and it follows from the definition of a module that this is the only action of the integers on an abelian group for which the resulting module is unital. Because of this, and since every abelian group can be made into a $\mathbb{Z}$-module and every $\mathbb{Z}$-module is obviously an abelian group underneath, it follows that the set of $\mathbb{Z}$-modules and the set of abelian groups are the "same" - every abelian group admits a $\mathbb{Z}$-module structure. Moreover, $\mathbb{Z}$-submodules are the same as subgroups of abelian groups. Note that in contrast to vector spaces, because modules are over rings, a weaker requirement, it's possible to have non-zero elements which, under the group action, produce zero. In $\mathbb{Z}$-modules, this can easily be seen by considering any element of the group multiplied by its order (if it's finite). It also follows that if the group itself has order N, then $Ng = 0$ for every g in the group. This means that G is also a $\mathbb{Z}/N\mathbb{Z}$-module (since taking the product of integers greater than N simply cycles through the first N integers again).

Another important module to consider are those involving polynomials. If F denotes field, then $F[x]$ denotes the set of polynomials with coefficients in F (in the indeterminate x). Let V be a vector space over F, and $T : V \to V$ be a linear transformation. Clearly, V is an F-module. Let's see how we can use T to make V into an $F[x]$-module. For $p(x) = a_n x^n + \cdots a_1 x + a_0 \in F[x]$ and $v \in V$, define

$$p(x)v = a_n T^n(v) + \cdots + a_1 T(v) + a_0 v \text{ where } T^k = T \circ \cdots \circ T \ (k \text{ times})$$

where the product of any a_k and $T^k(v)$ is given by the action of F on V. What we've done here is define x to act on V as the linear transformation T (so that $xv := T(v)$), and then extended this action to polynomials in x in a natural way, replacing multiplication with composition. Because F is a subring of $F[x]$ (if we take elements of F to be constant polynomials), and because this action is consistent with the action of F on V, the definition of the $F[x]$-module is a natural extension of the definition of the F-module. The most intuitive definition for this action sets T to be the identity transformation, so the action of a polynomial on v is defined simply to be the evaluation of the polynomial at v without any powers. We allow the transformation used to vary beyond the identity to generalize this notion, so that in general there are many different $F[x]$-module structures for the same vector space. Thus, module over a polynomial ring is completely determined by a vector space equipped with a linear transformation. The $F[x]$-submodules must be vector spaces contained in V, and T must be closed on the subspace for it to be a valid module. This means that the $F[x]$-submodules of V are precisely the T-stable subspaces of V (recall that subspace W is T-stable if $T(W) \subseteq W$). Because $F[x]$ is a principal ideal domain, and since, as we'll see in the next sections, the structure of a module is dependent on the ideal structure of the underlying ring, $F[x]$-modules are forced to have a relatively uncomplicated structure that reveals a lot about the linear transformation T (such as its rational canonical form and Jordan canonical form).

We conclude the section with an analogous result to the subgroup criterion from group theory (which states that subgroups always contain $x - y$ for every x, y in the subgroup), followed by the definition of an another algebraic object - algebras.

(Theorem) Submodule criterion: *Let R be a ring and M be an R-module. A subset $N \subseteq M$ is a submodule of M if and only if it's non-empty and $\forall r \in R, x, y \in N : x + ry \in N$.*

• *Proof*: If N is a submodule, then because the empty set doesn't admit a group structure, N must be non-empty (in particular it must contain 0). Since N is closed under addition and the action of R, we must have $x + yr \in N$. Conversely, if N is non-empty and $x + yr \in N$ always holds, then clearly N is a subgroup of M by the subgroup criterion (just let $r = -1$). Letting $x = 0$, it follows that N is closed under the action of R. The distributivity and associativity axioms of the action follow from M being a module containing N, so it follows that N is a submodule. □

We now introduce the idea of an *algebra*, which are both rings and vector spaces (recall that the *center* of an algebraic structure is the set of elements which commute with every other element).

(Definition) Algebra over a ring: *Let R be a commutative unital ring. An R-**algebra** is a unital ring A together with a ring homomorphism $f : R \to A$ such that $f(1_R) = 1_A$ and $f(R)$ is contained in the center of A.*

• *Motivation*: R-algebras are known in other contexts as *associative algebras*, because their definition is equivalent to an abelian group A over a commutative ring R satisfying

$$\forall r \in R, x, y \in A : r \cdot (xy) = (r \cdot x)y = x(r \cdot y)$$

Thus, A has the structure of both a ring and an R-module, such that the action of R on A is both associative and commutative. Another way to define associative algebras is to start with an R-module A and equip it with a binary operation on A which is associative. Normally R-modules don't have any binary operators defined on them (eg vector spaces have scalar multiplication defined on them, but not necessarily a binary operation on the vectors). Notice that even when we start with a unital ring A, we can make it an R-module using the homomorphism f if we define the ring action

$$\forall r \in R, x \in A : r \cdot x = f(r)x = xf(r)$$

(since $f(r)$ commutes).

We wrap up the section by extending the definition of homomorphisms to algebras in a natural way.

(Definition) Algebra homomorphism: *Let R be a unital ring, and A and B be two R-algebras. An **R-algebra homomorphism** is a ring homomorphism $\phi : A \to B$ such that*

$$\phi(1_A) = 1_B \text{ and } \forall r \in R, a \in A : \phi(r \cdot a) = r \cdot \phi(a)$$

2 Quotient Modules and Module Homomorphisms

Having defined the basic theory of modules, we'll now begin extending many of the concepts we introduced for groups and rings to modules, starting with quotients and homomorphisms. A homomorphism between modules is defined exactly as one would expect - a mapping which respects the structures of the modules.

(Definition) Module homomorphism: *Let R be a unital ring and M, N be R-modules. An **R-module homomorphism** is a mapping $\phi : M \to N$ such that*

$$\forall x, y \in M : \phi(x + y) = \phi(x) + \phi(y) \text{ and } \forall r \in R, x \in M : \phi(rx) = r\phi(x)$$

An R-module homomorphism is an *isomorphism* of R-modules if it's bijective, in which case we say M and N are isomorphic, denoted $M \cong N$. We also define the *kernel* of a homomorphism to be the set of elements in the domain which map to the identity. Finally, we denote the set of all R-module homomorphisms from M to N with $\text{Hom}_R(M, N)$. Clearly, every R-module homomorphism is also a group homomorphism, though the converse clearly doesn't hold. It's also not difficult to show that the kernel and images of an R-module homomorphism are submodules. Moreover, $\text{Hom}_R(M, N)$ is closed under addition, under which it's an abelian group, and is also closed over the ring action

$$\forall \phi \in \text{Hom}_R(M, N), r \in R, m \in M : (r\phi)(m) = r\phi(m)$$

under which it's an R-module if R is commutative (on the right side we use the action of R on N). The composition of R-module homomorphisms is also an R-module homomorphism. All these properties can be easily proven and are analogs of properties of group homomorphisms. Finally, $\text{Hom}_R(M, N)$ is a unital ring under addition and function composition, and is in fact an R-algebra whenever R is commutative.

(Definition) Endomorphism ring: *The ring $Hom_R(M, M)$ over unital ring R and R-module M, is called the **endomorphism ring of** M, and is denoted $End_R(M)$. Elements of $End_R(M)$ are called **endomorphisms**.*

In other words, the endomorphism ring of M is the ring of all homomorphisms from M onto itself, under the operations of pointwise addition and function composition (hence, endomorphisms are just homomorphisms from a module to itself). The same concept actually applies when M is an abelian group, and not necessarily an R-module. As expected, the multiplicative identity is the identity function. Endomorphism rings, though having an ostensibly contrived definition, will come to be quite important in developing a ring theoretic analog of Cayley's theorem from group theory, which states that every group G can be embedded in (ie is isomorphic to) some subgroup of Sym_G (the symmetric group over G, which is the group of bijections over G under function composition). Endomorphism rings will be critical in showing that, similarly, every ring naturally embeds into the endomorphism ring of its underlying group structure (ie every ring is isomorphic to the endomorphism ring of some abelian group).
Recall that we can make $\text{End}_R(M)$ an R-module by defining the action $\forall r \in R, \phi \in \text{End}_R(M) : (r \cdot \phi)(m) = r \cdot \phi(m)$ for $m \in M$. When R is commutative, we can make $\text{End}_R(M)$ an R-algebra by defining an analogous map $f : R \to \text{End}_R(M)$ by $\forall r \in R : f(r) = rI$ where I is the identity mapping (since R is unital, $f(R)$ lies in the center of M, and the identity commutes). To clarify, rI is defined as the endormorphism given by

$$\forall m \in M : (rI)(m) = rI(m) = rm$$

so this is a very simple action of R onto M's endormorphism ring.

Let's now turn our attention to quotient modules and their construction. First, let's refresh the meaning of a quotient group. The intuition for quotient groups is that we want to take a group G and define a structure-preserving equivalence relation on it, so that similar elements of G collapse to the same point under the equivalence relation. The motivating example is considering the group of integers under addition - to form the quotient group of integers modulo n, the equivalence relation in question is formed by considering elements with the same residue modulo n to be similar, causing each of the sets

$$\{0, 0+n, 0+2n, \cdots\}, \{1, 1+n, 1+2n, \cdots\}, \cdots, \{n-1, n-1+n, \cdots\}$$

to collapse to the points $\overline{0}, \overline{1}, \cdots, \overline{n-1}$ respectively (though we consider these elements in the integers modulo n to be ordinary integers, technically speaking they're equivalence classes and hence sets). When we quotient a group, the equivalence class containing the identity, say N, becomes the identity in the quotient group, and moreover is a normal subgroup (ie it commutes with elements of G, so the left and right cosets coincide) with the rest of the elements of the quotient groups being the left cosets of N. Thus, what it means to quotient G by a subgroup N is to form the left cosets $gN = \{gn, n \in N\}$ for $g \in G$ and treat each coset as its own unique element in a new, smaller group under the operation

$$(gN)(hN) = (gh)N \text{ for } g, h \in G$$

with $1N = N$ forming the identity. What we've done here is collapse all the elements of N into a single point, and then similarly collapse other sets of elements into points in such a way as to preserve the structure of G - we've essentially "divided out" N by considering all its elements to be the same. This is reinforced by the first group isomorphism theorem, which states that if G/N is a quotient group, then N is the kernel of some homomorphism ϕ over G, and moreover G/N is isomorphic to $\phi(G)$. The structure of the quotient group is the structure of $\phi(G)$ - this is presumably because ϕ maps elements of G which are in the same coset to the same element, formalizing the collapsing of elements we alluded to.
To extend all that intuition to some notion of quotient modules, let's first prove that every submodule N of an R-module (where R is a unital ring) is "normal" in the sense that the quotient module M/N is well defined, and the projection $\phi : M \to M/N$ is an R-module homomorphism with kernel N. More concretely, if N is a submodule of M, then M/N is an (abelian) quotient group, which we can make into an R-module, allowing us to call M/N a quotient module. The way we make it an R-module is with the action given by

$$\forall r \in R, m \in M, m + N \in M/N : r \cdot (x + N) = (r \cdot x) + N$$

That is, the action of an element of $r \in R$ on a coset of N in M given by m, which of course is an element of M/N, is defined to be the coset given by the action of r on m. This operation is well defined since if, for $x, y \in M$, $x + N = y + N$, then $r(x + N) = r(y + N)$ since the assumption implies $x - y \in N$ which implies $r \cdot (x - y) = rx - ry \in N$ and hence $rx + N = ry + N$. The axioms of a module are easily seen to be verified.
We conclude the section by stating that the group isomorphism theorems hold for quotient modules. Before doing so, we define the sum of modules in the same way we define a sum of groups: for submodules A, B of an R-module, define $A + B = \{a + b, a \in A, b \in B\}$.

(Theorem) Module isomorphism theorems: *Let R be a unital ring. Then the following hold.*

1. **First Isomorphism Theorem**: *Let M, N be R-modules and $\phi : M \to N$ be an R-module homomorphism. Then the kernel of ϕ is a submodule of M, and $M/ker(\phi) \cong \phi(M)$.*
2. **Second Isomorphism Theorem**: *Let A, B be submodules of some R-module. Then $(A + B)/B \cong A/(A \cap B)$.*
3. **Third Isomorphism Theorem**: *Quotient modules cancel, ie for submodules A, B of R-module M with $A \subseteq B$, $(M/A)/(B/A) \cong M/B$.*
4. **Fourth Isomorphism Theorem**: *Let N be a submodule of the R-module M. Then every submodule of M that contains N bijectively corresponds to a submodule of M/N. This correspondence is given by associating a submodule A of M (with $N \subseteq A$) with the submodule A/N of M/N.*

The first isomorphism theorem, discussed above in the intuition on quotient groups, formalizes much of the intuition behind partitioning a group into substructures which collapse sets of elements to points which preserve the same structure as the original group. The second isomorphism theorem covers the situation wherein we might want to try quotienting out a normal subgroup of a group by another subgroup of the group. Symbolically, if we have a group G, subgroup H, and normal subgroup N, we want to try to quotient N out of H. Of course, this isn't always possible, since N might not be a subgroup of H, or even be contained in H. To get around this, there are two natural "next best" approaches we might take - instead of quotienting N out of H, we could (1) quotient N out of the largest subgroup of G that contains both N and H (which turns out to be $N + H$), or (2) quotient H by the part of N that does lie in H. The second isomorphism theorem states that both approaches leave us in the same place. The third isomorphism theorem isn't particularly deep or insightful, but does cement our intuition of quotienting as an algebraic generalization of the division operation on integers. The fourth isomorphism

theorem, also known as the *lattice theorem* and also as the *correspondence theorem*, seems like an intuitive result one would expect from the first isomorphism theorem - if M/N preserves the structure of M, then we'd expect submodules A of M containing N to "collapse" themselves upon quotienting out N, leaving behind A/N. The lattice part of the name comes from the fact that the bijection detailed in the theorem is also a lattice isomorphism, between the lattice of submodules of M containing N and the lattice of submodules of M/N.

3 Generation of Modules, Direct Sums, and Free Modules

In this section we extend the notion of generators and sums from group and ring theory to module theory. As usual, R is a unital ring, and we use the term module to refer to left modules. First, let's define the concept of the sum of submodules.

(Definition) Sum of submodules: *Let M be an R-module, and $N_1, \cdots, N_n$ be submodules. We define their **sum** to be the set of all finite sums from elements of the N_i, ie*

$$N_1 + \cdots + N_n = \{a_1 + \cdots + a_n \text{ where } \forall i : a_i \in N_i\}$$

Let's also define generators. As in group theory, we'll use the concept of a generating set to refer to a set of elements whose span (in the linear algebra sense of the word) is the module. In the below definition, we first define a natural conception of the "action" of R on a subset of an R-module, which is essentially just the set of R-linear combination of elements in the subset. The subset is a generating set of the containing R-module precisely when its set of R-linear combinations, ie its span, is the whole module.

(Definition) Generating set: *Let M be an R-module, and A be a subset of M. Define*

$$RA = \{r_1 a_1 + \cdots + r_n a_n \text{ where } m \in \mathbb{Z}^+ \text{ and } \forall i : r_i \in R, a_i \in A\}$$

*We call RA the **submodule of** M **generated by** A. Moreover, if N is a submodule of M and $N = RA$ for some $A \subseteq M$, we say A is the **generating set** of N.*

To finish up, we say a submodule is *finitely generated* when its generating set is non-empty and finite; we also say a submodule is *cyclic* if there exists a generating set with a single element. It's easy to see that RA is the smallest submodule of M containing A. Moreover, we can link the above two concepts (sum and generators) - $N_1 + \cdots + N_n$ is precisely the submodule generated by the set $N_1 \cup \cdots \cup N_n$, and is the smallest submodule containing every N_i. Moreover, if N_i is generated by A_i, then the sum is generated by the union of the A_i's. As we'll see in the next chapters, starting with a subset of an R-module and extending it to the submodule of all R-linear combinations of the subset will be our primary way of producing submodules from a module.
The sum of submodules produces another submodule in the same module, so the containing module is "closed" under the sum operation. The *direct product*, on the other hand, is a way to form a new module from old ones, in such a way as to borrow from the structure of each. It simply takes imbues the Cartesian product with the necessary module structure.

(Definition) Direct product: *Let $M_1, \cdots, M_n$ be R-modules. We define the **direct product** of the modules to be the set of all n-tuples with a component from each module, ie*

$$M_1 \times \cdots \times M_n = \{(m_1, \cdots, m_n) \text{ where } \forall i : m_i \in M_i\}$$

where addition and the action of R are defined component-wise in terms of the operations on each of the modules.

Of course, the direct product of modules is itself a module, and is often called the *direct sum*, denoted $M_1 \oplus \cdots \oplus M_n$. We can even extend the definition of a direct product to infinitely many modules. This next result determines when a module is isomorphic to the direct product of some of its submodules. Intuitively, the theorem states that this is the case when the submodules are "disjoint", in the sense that none of them are redundant and hence contain no overlapping elements other than the additive identity.

(Theorem) Isomorphic to direct product of submodules: *Let $N_1, \cdots, N_n$ be submodules of the R-module M. The following are equivalent.*

1. *The map $\phi : N_1 \times \cdots \times N_n \to N_1 + \cdots + N_n$ given by*

$$\phi(a_1, \cdots, a_n) = a_1 + \cdots + a_n$$

is an R-module isomorphism.

2. *For every* $j \in \{1, \cdots, n\}$,

$$N_j \cap \sum_{i \neq j} N_i = \{0\}$$

3. *Every element in the sum can be written uniquely as the sum of elements in* N_i.

• *Proof*: Let's first prove that the first statement implies the second. Suppose ϕ, as defined in the first statement, is an isomorphism. If there were some $a_j \in N_j$ contained in the sum of the other N_i's, then we could write

$$a_j = \sum_{i \neq j} a_i$$

for $a_i \in N_i \neq N_j$. But then $\phi(a_1, \cdots, -a_j, \cdots, a_n) = a_1 + \cdots - \sum_{i \neq j} a_i + \cdots + a_n = 0$, meaning ϕ maps some non-identity element to the identity. Isomorphisms are bijections and hence only map the identity to the identity, a contradiction. Thus, no such a_j exists.
Next, let's prove that the second statement implies the third. If every N_j is disjoint from the sum of the other N_i's, then for any two sums of elements from the submodules

$$a_1 + \cdots + a_n = b_1 + \cdots + b_n \text{ for } a_i \in N_i, b_i \in N_i$$

We could write, for each j,

$$a_j - b_j = \sum_{i \neq j} (b_i - a_i) \in \sum_{i \neq j} N_i$$

Since $a_j - b_j \in N_j$, while the right side is in $\sum_{i \neq j} N_i$, with the intersection of the two containing only 0, it follows that

$$\forall j : a_j - b_j = 0$$

showing that the sums are the same, and hence any such sum for an element in $N_1 + \cdots + N_n$ is unique.
Finally, let's prove that the third statement implies the first. Obviously, ϕ is already surjective, so let's prove that it's injective. But this is precisely what the second statement says - that every image of ϕ corresponds to a unique pre-image. Hence, ϕ is bijective. ϕ is certainly an R-module homomorphism since the action of R distributes over addition, which is abelian. Hence, ϕ is an isomorphism. □

The concept of generators is analogous to the linear algebra concept of the span of a vector space. There's also an analogous definition for the basis of a vector space, which is just a generating set which uniquely generates the module, so that every element of the module can be written as an R-linear combination of terms from the generating set. We'll even extend the definiton of the rank of a matrix.

(Definition) Free: *An* R*-module* M *is* ***free*** *on* $A \subseteq M$ *if*

$$\forall x \in M : x = \sum_{i=1}^{m} r_i a_i \text{ for } m \in \mathbb{Z}^+, r_i \in R, a_i \in A$$

and the r_i, a_i *are unique; that is, this is the only possible expression of* x *as an* R*-linear combination of terms of* A*. Then we say that* A *is a* ***basis*** *or* ***set of free generators*** *of* M*.*
A free module is simple a finitely generated module with a unique generating set, which we call its basis. If R is commutative, the cardinality of A is called the *rank* of M, and is a precursor to the concept of dimension. The extension of the definition of a basis of a vector space to modules is simple - just a generating set which uniquely generates the module. Such a generating set is said to be *linearly independent*, since it can be shown that if there are multiple ways to express elements of M as R-linear combinations of A, then A isn't linearly independent, in the linear algebra sense of the word (ie there are elements whose weighted sum is zero). Notice that while every element in the direct sum of submodules can be expressed as a unique sum of elements from the submodules, the uniqueness of a basis is stronger in that both the elements from the basis and the weights from R are unique. There's only one way, period, to form the linear combination.
We've shown that certain modules have bases, in which case we call them free. A more universal converse is true - every set can be extended to a free module with the set as its basis, and moreover any mapping from the set to some module can be extended to a homomorphism from the corresponding free module to the module.

(Theorem) Universal Property of Free Modules: *For any set* A*, there's a free* R*-module, denoted* $F(A)$*, whose basis is* A*. Moreover,* $F(A)$ *satisfies the following* ***universal property*** *- if* M *is any* R*-module, and* $\phi : A \to M$ *is any mapping, there's a unique* R*-module homomorphism* $\Phi : F(A) \to M$ *such that*

$$\forall a \in A : \Phi(a) = \phi(a)$$

• *Proof*: We can explicitly construct $F(A)$ as follows. Let $F(A)$ be the set of all functions $f : A \to R$ such that $f(a) = 0$ for all but finitely many $a \in A$. We can give this set a module structure by defining pointwise addition and ring multiplication:

$$\forall f, g \in F(A), r \in R : (f + g)(a) = f(a) + g(a), (rf)(a) = rf(a)$$

We can identify A as a subset of $F(A)$ by associating each $a \in A$ with $f_a \in F(A)$, where

$$f_a(x) = \begin{cases} 1, & \text{if } x = a \\ 0, & \text{otherwise} \end{cases}$$

It follows that A is a basis for $F(A)$, if we identify each $f \in F(A)$ with the sum $r_1a_1 + \cdots + r_na_n$, where $f(r_i) = a_i$ and f is zero everywhere else - then

$$f = r_1f_{a_1} + \cdots + r_nf_{a_n}$$

We can prove the universal property by extending ϕ into Φ as follows: define

$$\Phi\left(\sum_{i=1}^{n} r_ia_i\right) = \sum_{i=1}^{n} r_i\phi(a_i)$$

Clearly, the restriction of Φ's domain to A gives us back ϕ, and on the rest of the domain it's easy to prove that Φ is an R-module homomorphism. Φ is the unique extension of ϕ, since the values of a homomorphism on the basis completely determine the values of the homomorphism everywhere. □

It follows as a corollary that free modules with the same basis are isomorphic, and that this isomorphic is the extension of the identity map on the basis. This is what we'd expect, since it should be the case that bases "uniquely" determine their "spanned" modules. This also means that any free module with basis A is isomorphic to $F(A)$, constructed in the above proof, so we can use $F(A)$ as the canonical free module with basis A. In particular, $\mathbb{Z}$-modules with basis A are isomorphic to $\mathbb{Z} + \cdots + \mathbb{Z}$ (n times) .

4 Tensor Product of Modules

The direct product of two modules is more of a set theoretic product than a satisfying notion of a true product between modules. In this section, we introduce the tensor product of two modules, which is a general construction which will give us a module in which we can take products of elements between the two modules. Before introducing the somewhat complicated construction of the tensor product, let's introduce its motivating example - the special case of extending scalars, or of changing base.

Let's consider how we might approach the question of extending a ring R to a larger ring in a way that preserves R-module structure. Let's let R be a subring of the ring S, and further require that R and S share the same multiplicative identity, to ensure that S is a well-defined unital R-module. Then any S-module M is naturally an R-module as well, under the same action. In particular, the following relation follows from the module axioms:

$$\forall s \in S, m \in M, r \in R : (sr)m = s(rm)$$

Let's generalize this to any two arbitrary rings R and S, losing the containment operation above. Then we can make any S-module M an R-module with the use of any homomorphism $\phi : R \to S$ by defining the action of R on M in terms of the action of S:

$$\forall r \in R, m \in M : r \cdot m = f(r) \cdot m$$

In the language of linear algebra, whereas M was previously algebraic structure with the elements of S as scalars, we now view it as an algebraic structure with the elements of R as scalars, with the corresponding scalar multiplication operation defined in terms of elements of S. All we've done is take M's existing S-module structure, and collapse the scalars from S which are the same under ϕ^{-1} (ie they're mapped to the same element of R) to a single scalar. For this reason, we say that the R-module obtained by defining the above action of R on M was obtained by *restriction of scalars* from S to R. Moreover, we view S as an *extension* of R.

This leads us to the question of whether we can, rather than restricting scalars, start with a ring and extend its scalars to another ring, to obtain a module. Specifically, given a ring R that's a subring of ring S, can we start with an R-module M and imbue it with an S-module structure by extending the action of R to the action of S somehow. In general, this is impossible, as we can see by considering the trivial case of R, which is an R-module but not always an S-module (for example, consider $R = \mathbb{Z}$ and $S = \mathbb{Q}$). However, the next best thing we can try here comes from noting that though we can't give M an S-module structure, it might be contained in some S-module. Put another way, we want to find an injection,

more commonly known in this case as an *embedding*, of the R-module M into an S-module, of which M is an R-submodule. More generally, we're interested in the question of what R-module homomorphisms between M and any S-module exist. Unfortunately, it is sometimes the case that there are no non-zero R-module homomorphisms (we use "non-zero" here to refer to a mapping which doen't send everything to zero), and therefore no embeddings of M in any S-module.

Even if we can't, in general, find any S-module to embed M in, we can try to find the best possible target S-module in which we try to embed M. The construction of such an S-module will, as we'll see, determine all possible R-module homomorphisms between M and an S-module, and lead to the construction of the tensor product. Let's begin with the construction of this S-module, by starting with M and manipulating it until all the module axioms for an S-module are satisfied. We begin with M as an abelian group, which must be equipped with an action of S, ie a structure-preserving map from $S \times M$ to S. A good starting point might be to simply take $S \times M$ and make it the basis of a module, ie consider to free $\mathbb{Z}$-module on $S \times M$. This is essentially an abelian group (under addition) consisting of finite and commutative sums of pairs of the form (s, m) for $s \in S$ and $m \in M$. This group's structure is quite bare, in the sense that there are no relations between its elements - every element is completely distinct. In order for it to satisfy the axioms of an S-module it needs to satisfy the relations $(sr)m = s(rm)$ and $(rm) = f(r)m$ that we found in the introduction. This means we need to have

1. $\forall s_1, s_2 \in S, m \in M : (s_1 + s_2, m) = (s_1, m) + (s_2, m)$
2. $\forall s \in S, m_1, m_2 \in M : (s, m_1 + m_2) = (s, m_1) + (s, m_2)$
3. $\forall s \in S, r \in R, m \in M : (sr, m) = (s, rm)$.

We can obtain this structure by imposing equivalence relation that forces terms of the above forms to be equal; this is the same as quotienting our free $\mathbb{Z}$-module over $S \times M$ by the subgroup generated by

$$(s_1 + s_2, m) - (s_1, m) - (s_2, m), (s, m_1 + m_2) - (s, m_1) - (s, m_2), \text{ and } (sr, m) - (s, rm)$$

This will give us the scalar extension that we seek, and defines the tensor product.

To summarize, we define the *tensor product* $S \otimes_R M$ as the quotient of $F_{\mathbb{Z}}(S \times M)$ modulo H, where $H = \langle (s_1 + s_2, m) - (s_1, m) - (s_2, m), (s, m_1 + m_2) - (s, m_1) - (s, m_2), (sr, m) - (s, rm) \text{ for } s, s_1, s_2 \in S, m, m_1, m_2 \in M, r \in R \rangle$. If we let $s \otimes m \in S \otimes M$, for $s \in S, m \in M$, denote the coset containing (s, m), then we get the following relations immediately.

$$\forall s_1, s_2 \in S, m \in M : (s_1 + s_2) \otimes m = s_1 \otimes m + s_2 \otimes m$$

$$\forall s \in S, m_1, m_2 \in M : s \otimes (m_1 + m_2) = s \otimes m_1 + s \otimes m_2$$

$$\forall s \in S, r \in R, m \in M : (sr) \otimes m = s \otimes (rm)$$

We refer to elements in the tensor product as *tensors*, and elements of the form $s \otimes m$ as *simple tensors*. Any tensor can always be written (though not uniquely) as a finite sum of simple tensors. Notice that $F_{\mathbb{Z}}(S \times M)/H$ is an abelian quotient group, and we have yet to imbue it with a module structure. To do so, we'll define the following action of S on $S \otimes_R M$:

$$\forall s, s_i \in S : s \cdot \sum_i s_i \otimes m_i = \sum_i (ss_i) \otimes m_i$$

This is well defined because every element of $S \otimes_R M$ can be written as a finite sum of simple tensors.
What we've done here is taken the R-module M and embedded it in $S \times M$, with the intention of viewing the element (s, m) as the "action" of s on m. To do this, we first make $S \times M$ an abelian group by lifting it to the $\mathbb{Z}$-module for which it's a basis, followed by imposing the appropriate equivalence relations to make the S-module axioms hold, and then defining an action of S on M in accordance with our motivation. Intuitively, our goal was always to embed M into an S-module - we show that this embedding has been preserved by providing a proper inclusion map. Take the natural mapping $\iota : M \to S \otimes_R M$ given by $\forall m \in M : \iota(m) = 1 \otimes m$. In other words ι maps an element m in the original R-module to the element $(1, m)$ in $S \times M$ and then quotients out H. Then ι is an R-module homomorphism. Because in general M can't always be embedded in any S-module, in general ι is not injective. This is the best we could hope for, since we set out to find the best possible target in which to attempt to embed M. The next theorem formalizes the usage of the term "best possible" by showing that any other R-module homomorphism from M to an S-module factors through ι.

(Theorem) Universal property for tensor products: *Let R be a subring of S, and M be an R-module. Define $\iota : M \to S \otimes_R M$ by $\forall m \in M : \iota(m) = 1 \otimes m$. Then if $\phi : M \to N$ is any other R-module homomorphism to S-module N, there's a unique S-module homomorphism $\Phi : S \otimes_R M \to N$ such that*

$$\phi = \Phi \circ \iota$$

Conversely, if $\Phi : S \otimes_R M \to N$ is any S-module homomorphism then $\Phi \circ \iota : M \to N$ is an R-module homomorphism.

• *Proof*: By the universal property of free modules, we can find a $\mathbb{Z}$-module homomorphism from $F_{\mathbb{Z}}(S \times M)$ to N that maps generators $(s, m) \in S \times M$ to $s\phi(m)$. As above, let

$$H = \big\langle (s_1+s_2,m)-(s_1,m)-(s_2,m), (s,m_1+m_2)-(s,m_1)-(s,m_2), (sr,m)-(s,rm) \text{ for } s,s_1,s_2 \in S, m,m_1,m_2 \in M, r \in R \big\rangle$$

TODO

• *Intuition*: Intuitively, the universal property states that ι is the closest we can ever get to an inclusion map between M and an S-module, and hence that $S \otimes_R M$ is the closest we can get to an S-module with M embedded in it. This is because if we took any other S-module, then every homomorphism (and hence every candidate for an inclusion mapping) between M and that S-module would pass through $S \otimes_R M$ in the middle. When it comes to embedding M in an S-module, passing through $S \otimes_R M$ is unavoidable.

• *Corollary*: A corollary of this theorem also gives us the best possible embedding of M in an S-module, in the sense that $M/\ker(\iota)$ is the largest quotient module of M that can be embedded in some S-module. This also means that M itself can be embedded in an S-module if and only if ι is injective.

So far we've discussed how to build an intuitive notion of extending the scalars of one ring to a superring, and finding a module in which to embed a module over the ring, or at least the closest thing to it we can find. This involved defining $S \otimes_R M$ in the special case where R is a subring of S, in which case we used elements $(s, m) \in S \times M$ to act like the product of s and m. Notice that this really only required M to be a left R-module, and for S to be a right R-module (equivalent to being a superring). We'll now take a look at the general tensor product construction, motivated by this special case, in which we'll construct $M \otimes_R N$ for any right R-module M and left R-module N.

Let's take a look at how we can take a general right module M and left module N and define an abelian group $M \otimes_R N$, afer which we can investigate how to imbue the group with a module structure. Recall that the motivation for a tensor product is to take two modules and produce a third in which some notion of multiplying elements of M and N is defined. Towards this goal it makes sense to start with $M \times N$, where we'll consider pairs (m, n); if we can give the set the right structure, we'll be able to get (m, n) to behave like the multiplication of m and n. As before, since $\mathbb{Z}$-modules are equivalent to abelian groups, we start by considering the free $\mathbb{Z}$-module generated by $M \times N$: $F_{\mathbb{Z}}(M \times N)$. We now have an abelian group, but with minimal structure; in particular, there are no relations between the elements. Before we give the group a module structure, let's preemptively ensure that the R-module axioms are satisfied. Again, we can do this by quotienting out the subgroup

$$H := \big\langle (m_1 + m_2, n) - (m_1, n) - (m_2, n), (m, n_1 + n_2) - (m, n_1) - (m, n_2), (mr, n) - (m, rn) \big\rangle$$

for $m_1, m_2, m \in M, n_1, n_2, n \in N, r \in R$. We denote the resulting abelian quotient group $M \otimes_R N$.

We now have an abelian group suitable for the defining of a module structure. Generalizing, we call $M \otimes_R N$ the *tensor product of M and N over R*, and we'll denote elements of $M \otimes_R N$ (ie cosets of $F_{\mathbb{Z}}(M \times N)$) with $m \otimes n$, referring to the coset containing (m, n). We call elements of $M \otimes_R N$ *tensors*, and elements of the form $m \otimes n$ *simple tensors*. Notice that it's possible to have

$$m_1 \otimes n_1 = m_2 \otimes n_2, (m_1, n_1) \neq (m_2, n_2)$$

since all that's required is that (m_1, n_1) and (m_2, n_2) are in the same coset of $F_{\mathbb{Z}}(M \times N)$. Thus, though any element of the tensor product is expressible as a sum of simple tensors, this sum is not unique. Moreover, caution must be taken when defining maps over the tensor product, since maps are only well defined if they map (m_1, n_1) and (m_2, n_2) to the same element if the corresponding simple tensors are equal.

Notice that in the construction, we mapped $M \times N$ to the free $\mathbb{Z}$-module on $M \times N$ and passed to the quotient group; this defined a map

$$\iota : M \times N \to M \otimes_R N \text{ where } \forall (m, n) \in M \times N : \iota(m, n) = m \otimes n$$

In general, ι isn't a group homomorphism, but it is additive, in that $\iota(mr, n) = (mr) \otimes n = m \otimes (rn) = \iota(m, rn)$. This map is important enough that we give it a name, since we'll often use it when talking about tensor products.

(Definition) Balanced with respect to a module: *Let M be a right R-module, N be a left R-module, and L be an abelian group. A map $\phi : M \times N \to L$ is R-**balanced** if*

$$\phi(m_1 + m_2, n) = \phi(m_1, n) + \phi(m_2, n)$$

$$\phi(m, n_1 + n_2) = \phi(m, n_1) + \phi(m, n_2)$$

$$\phi(mr, n) = \phi(m, rn)$$

for $m_1, m_2, m \in M, n_1, n_2, n \in N, r \in R$.

By definition, ι as defined above is R-balanced. Using this terminology, we can extend the previous univeral property's statement to an even stronger statement.

(Theorem) Universal property of tensor products: *Let M be a right R-module and N be a left N-module. Define $\iota : M \times N \to M \otimes_R N$ by $\iota(m, n) = m \otimes n$. Then*

1. *For any group homomorphism Φ from $M \times N$ to an abelian group L, the map $\phi = \Phi \circ \iota$ is an R-balanced map from $M \times N$ to L.*
2. *Conversely, for any R-balanced map ϕ from $M \times N$ to L, there's a unique group homomorphism Φ from $M \otimes_R$ to L, and moreover ϕ factors through ι, ie $\phi = \Phi \circ \iota$.*

• *Proof*: TODO • *Intuition*: This theorem essentially uses the previously stated universal property of tensor products, which showed that $S \otimes_R M$ was the closest we could get to embedding an R (subring of S) module in an S-module, in the sense that R-module homomorphisms over M to an S-module factor through the natural mapping from M to $S \otimes_R M$. It builds on this result by showing that there's R-balanced maps $\phi : M \times N \to L$, although not group homomorphisms themselves, do bijectively correspond with group homomorphisms $\Phi : M \otimes_R N \to L$.

This theorem is useful in defining group homomorphisms on $M \otimes_R N$, since we need only verify that the map is R-balanced rather than going through the tedious check of whether the map is well defined. Since simple tensors $m \otimes n$ generate $M \otimes_R N$, and correspond to (m, n) in $M \times N$, it follows as a corollary to the universal property of tensor products that if there's an R-balanced map $\iota' : M \times N \to G$ for abelian group G such that the image ι' generates G, then $M \otimes_R N$ is isomorphic to G. Intuitively, this is because the generators of the tensor product map to the generators of G, in a way that preserves the "tensor product structure" (ie ι' need not be a homomorphism, but merely must be R-balanced).

So far, all we've done is generalize the process of extending scalars to turning any right module and left module into a joint abelian group, and proven some results about the structure of this abelian group. Let's now return to the full construction of the tensor product, and investigate how we can imbue $M \otimes_R N$ with a module structure. Recall that in extending scalars, all that was required to give $S \otimes_R M$ an S-module structure was a left S-module structure on S