Math 533 Winter 2021, Lecture 10: Modules

website: `https://www.cip.ifi.lmu.de/~grinberg/t/21w/`

1. Modules ([DF, Chapter 10])

1.1. The universal property of a free module ([DF, §10.3])

As before, we fix a ring R. Recall the last theorem from Lecture 9:

> **Theorem 1.1.1.** Let M be a left R-module. Let $(m_i)_{i\in I}$ be any family of vectors in M. Consider the map
>
> $$f : R^{(I)} \to M,$$
> $$(r_i)_{i\in I} \mapsto \sum_{i\in I} r_i m_i.$$
>
> Then:
>
> **(a)** This map f is always a left R-module morphism.
>
> **(b)** The map f is injective if and only if the family $(m_i)_{i\in I}$ is linearly independent.
>
> **(c)** The map f is surjective if and only if the family $(m_i)_{i\in I}$ spans M.
>
> **(d)** The map f is an isomorphism if and only if the family $(m_i)_{i\in I}$ is a basis of M.

The map f in Theorem 1.1.1 takes a family $(r_i)_{i\in I}$ of scalars, and uses it to build a linear combination of $(m_i)_{i\in I}$.

The next proposition shows that linear maps respect linear combinations (in the sense that if you apply a linear map to a linear combination of some vectors, then you get the same linear combination of their images):

> **Proposition 1.1.2.** Let M and P be two left R-modules. Let $f : M \to P$ be an R-linear map. Let $(m_i)_{i\in I}$ be any family of vectors in M, and let $(r_i)_{i\in I}$ be a family of scalars in R with the property that
>
> $$\text{all but finitely many } i \in I \text{ satisfy } r_i = 0. \qquad (1)$$
>
> Then,
>
> $$f\left(\sum_{i\in I} r_i m_i\right) = \sum_{i\in I} r_i f(m_i).$$

Proof of Proposition 1.1.2. We give a proof by example: We assume that $I = \{1, 2, 3\}$. Thus, the claim we need to prove is saying that

$$f(r_1 m_1 + r_2 m_2 + r_3 m_3) = r_1 f(m_1) + r_2 f(m_2) + r_3 f(m_3).$$

But this is a consequence of the linearity of f (applied several times):

$$\begin{aligned}
& f\left(r_1 m_1 + r_2 m_2 + r_3 m_3\right) \\
& = f\left(r_1 m_1 + r_2 m_2\right) + f\left(r_3 m_3\right) \qquad (\text{since } f \text{ respects addition}) \\
& = f\left(r_1 m_1\right) + f\left(r_2 m_2\right) + f\left(r_3 m_3\right) \qquad (\text{since } f \text{ respects addition}) \\
& = r_1 f\left(m_1\right) + r_2 f\left(m_2\right) + r_3 f\left(m_3\right) \qquad (\text{since } f \text{ respects scaling}).
\end{aligned}$$

The same reasoning applies to an arbitrary finite set I. (To be fully rigorous, this is a proof by induction on $|I|$.)

The case when I is infinite can be reduced to the case when I is finite using the assumption (1). Indeed, because of (1), there is a finite subset J of I such that all $i \in I \setminus J$ satisfy $r_i = 0$. Choosing such a J, we then have

$$\sum_{i\in I} r_i m_i = \sum_{i\in J} r_i m_i \qquad \text{and} \qquad \sum_{i\in I} r_i f\left(m_i\right) = \sum_{i\in J} r_i f\left(m_i\right), \tag{2}$$

since vanishing addends in a sum can be discarded. But since we have already proved Proposition 1.1.2 in the case of a finite set I, we can apply Proposition 1.1.2 to J, and thus obtain $f\left(\sum\limits_{i\in J} r_i m_i\right) = \sum\limits_{i\in J} r_i f\left(m_i\right)$. In view of (2), this rewrites as $f\left(\sum\limits_{i\in I} r_i m_i\right) = \sum\limits_{i\in I} r_i f\left(m_i\right)$, so we are done. □

One useful feature of bases is that they make it easy to define linear maps out of a module: Namely, if M is a module with a basis $\left(m_i\right)_{i\in I}$, and you want to define a linear map f out of M, then it suffices to specify the values $f\left(m_i\right)$ of the map on each vector of the basis. These values can be specified arbitrarily; each possible specification yields a unique linear map f. Here is the theorem that underlies this strategy:

> **Theorem 1.1.3** (Universal property of free modules)**.** Let M be a free left R-module with basis $\left(m_i\right)_{i\in I}$. Let P be a further left R-module (not necessarily free). Let $p_i \in P$ be a vector for each $i \in I$. Then, there exists a **unique** R-linear map $f : M \to P$ such that
>
> $$\text{each } i \in I \text{ satisfies } f\left(m_i\right) = p_i. \tag{3}$$

Proof. **Uniqueness:** If $f : M \to P$ is an R-linear map satisfying (3), then any R-linear combination $\sum\limits_{i\in I} a_i m_i$ of $\left(m_i\right)_{i\in I}$ (with $a_i \in R$ and with all but finitely many $i \in I$ satisfy $a_i = 0$) satisfies

$$\begin{aligned}
f\left(\sum_{i\in I} a_i m_i\right) &= \sum_{i\in I} a_i \underbrace{f\left(m_i\right)}_{\substack{=p_i \\ (\text{by } (3))}} \qquad (\text{by Proposition 1.1.2}) \\
&= \sum_{i\in I} a_i p_i. \qquad (4)
\end{aligned}$$

This equality uniquely determines the value of f on each R-linear combination of $(m_i)_{i\in I}$. But each element of M can be written as an R-linear combination of $(m_i)_{i\in I}$ (since $(m_i)_{i\in I}$ is a basis of M and thus spans M). Thus, the equality (4) uniquely determines the value of f on each element of M. In other words, it uniquely determines f. Hence, the R-linear map f satisfying (3) is unique.

Existence: Consider the map

$$\begin{aligned} g : R^{(I)} &\to M, \\ (r_i)_{i\in I} &\mapsto \sum_{i\in I} r_i m_i. \end{aligned}$$

This is the map we called f in Theorem 1.1.1 (of course, we cannot call it f right now, since we need the letter for something else). Theorem 1.1.1 **(d)** yields that the map g is an isomorphism (since the family $(m_i)_{i\in I}$ is a basis of M). In particular, this means that g is bijective. Hence, any element of M can be written as an R-linear combination $\sum\limits_{i\in I} r_i m_i$ of $(m_i)_{i\in I}$ for a **unique** family $(r_i)_{i\in I} \in R^{(I)}$.

Thanks to this, we can define a map

$$\begin{aligned} f : M &\to P, \\ \sum_{i\in I} r_i m_i &\mapsto \sum_{i\in I} r_i p_i \qquad \left(\text{with } (r_i)_{i\in I} \in R^{(I)}\right). \end{aligned}$$

Now, it is easy to see that this map f is R-linear and satisfies (3). Hence, the R-linear map f satisfying (3) exists.

Having proved both existence and uniqueness, we are now done proving Theorem 1.1.3. □

In the proof of the "Uniqueness" part above, we have not used the assumption that the family $(m_i)_{i\in I}$ is a basis of M; we have only used that it spans M. Thus, the uniqueness of f holds even under this weaker condition. Let us isolate this into a separate theorem:

> **Theorem 1.1.4** (Linear maps are determined on a spanning set)**.** Let M be a left R-module. Let $(m_i)_{i\in I}$ be a family of vectors in M that spans M. Let P be a further left R-module. Let $f, g : M \to P$ be two R-linear maps such that
>
> $$\text{each } i \in I \text{ satisfies } f(m_i) = g(m_i).$$
>
> Then, $f = g$.

This theorem is often used to prove that two linear maps are equal.

1.2. Bilinear maps

When R is a commutative ring, the addition map

$$\text{add} : R \times R \to R, \qquad (a, b) \mapsto a + b$$

is R-linear (where the domain is the direct product of two copies of the left R-module R). In fact, if $(a, b) \in R \times R$ and $(c, d) \in R \times R$ are any two pairs, then

$$\operatorname{add}\left(\underbrace{(a,b)+(c,d)}_{=(a+c,b+d)}\right) = \operatorname{add}((a+c, b+d)) = (a+c)+(b+d) \qquad \text{and}$$

$$\operatorname{add}((a,b)) + \operatorname{add}((c,d)) = (a+b)+(c+d) = (a+c)+(b+d)$$

are clearly the same thing. (This just shows that add respects addition; but the other axioms are just as easy.)

However, the multiplication map

$$\operatorname{mul} : R \times R \to R, \qquad (a,b) \mapsto ab$$

is **not** R-linear. However, it is linear in the first argument if we fix the second. In other words, for any given $b \in R$, the map

$$R \to R, \qquad a \mapsto ab$$

is R-linear. Likewise, the multiplication map $\operatorname{mul} : R \times R \to R$ is linear in the second argument if we fix the first. Such maps have a name:

Definition 1.2.1. Let R be a commutative ring. Let M, N and P be three R-modules. A map $f : M \times N \to P$ is said to be R**-bilinear** (or just **bilinear**) if it satisfies the following two conditions:

- For any $n \in N$, the map

  $$M \to P, \qquad m \mapsto f(m,n)$$

  is R-linear. That is, we have

  $$f(m_1 + m_2, n) = f(m_1, n) + f(m_2, n) \qquad \text{for all } m_1, m_2 \in M \text{ and } n \in N;$$
  $$f(rm, n) = rf(m,n) \qquad \text{for all } r \in R,\ m \in M \text{ and } n \in N;$$
  $$f(0,n) = 0 \qquad \text{for all } n \in N.$$

  This is called **linearity in the first argument**.

- For any $m \in M$, the map

  $$N \to P, \qquad n \mapsto f(m,n)$$

  is R-linear. That is, we have

  $$f(m, n_1 + n_2) = f(m, n_1) + f(m, n_2) \qquad \text{for all } m \in M \text{ and } n_1, n_2 \in N;$$
  $$f(m, rn) = rf(m,n) \qquad \text{for all } r \in R,\ m \in M \text{ and } n \in N;$$
  $$f(m,0) = 0 \qquad \text{for all } m \in M.$$

  This is called **linearity in the second argument**.

Here are some examples of bilinear maps:[1]

- As I said, the multiplication map $R \times R \to R, \ (a,b) \mapsto ab$ is R-bilinear.
- For any $n \in \mathbb{N}$, the **standard scalar product** (also known as the **dot product**)
$$\begin{aligned} R^n \times R^n &\to R, \\ ((a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n)) &\mapsto a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \end{aligned}$$
is R-bilinear.
- Consider the field $\mathbb{C}$ of complex numbers. For any $n \in \mathbb{N}$, the **standard inner product**
$$\begin{aligned} \mathbb{C}^n \times \mathbb{C}^n &\to \mathbb{C}, \\ ((a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n)) &\mapsto a_1 \overline{b_1} + a_2 \overline{b_2} + \cdots + a_n \overline{b_n} \end{aligned}$$
(where $\overline{z}$ denotes the complex conjugate of a $z \in \mathbb{C}$) is $\mathbb{R}$-bilinear but not $\mathbb{C}$-bilinear (since it is antilinear rather than linear in the second argument). However, it becomes $\mathbb{C}$-bilinear if you view it as a map $\mathbb{C}^n \times \overline{\mathbb{C}}^n \to \mathbb{C}$ (with $\overline{\mathbb{C}}$ being the "twisted" $\mathbb{C}$-module $\mathbb{C}$ from Lecture 9).
- The determinant map
$$\begin{aligned} \det : R^2 \times R^2 &\to R, \\ ((a,b), (c,d)) &\mapsto ad - bc \end{aligned}$$
is R-bilinear. (This is called the determinant map because it sends a 2×2-matrix – encoded as pair of pairs – to its determinant.)
- Matrix multiplication is bilinear. That is: For any $m, n, p \in \mathbb{N}$, the map
$$\begin{aligned} R^{m \times n} \times R^{n \times p} &\to R^{m \times p}, \\ (A, B) &\mapsto AB \end{aligned}$$
is R-bilinear.
- The **cross product** map
$$\begin{aligned} R^3 \times R^3 &\to R^3, \\ ((a,b,c), (d,e,f)) &\mapsto (bf - ce, cd - af, ae - bd) \end{aligned}$$
is R-bilinear.

[1] In all these examples, R is assumed to be commutative.

- For any R-module M, the action map

  $$R \times M \to M,$$
  $$(r, m) \mapsto rm$$

  is R-bilinear. In fact, it is linear in its first argument since

  $$(r_1 + r_2)\, m = r_1 m + r_2 m \qquad \text{for all } r_1, r_2 \in R \text{ and } m \in M;$$
  $$(rs)\, m = r\,(sm) \qquad \text{for all } r, s \in R \text{ and } m \in M;$$
  $$0_R \cdot m = 0_M \qquad \text{for all } m \in M;$$

  and it is linear in its second argument since

  $$r\,(m_1 + m_2) = rm_1 + rm_2 \qquad \text{for all } r \in R \text{ and } m_1, m_2 \in M;$$
  $$r\,(sm) = s\,(rm) \qquad \text{for all } r, s \in R \text{ and } m \in M;$$
  $$r \cdot 0_M = 0_M \qquad \text{for all } r \in R.$$

  (Here, the equality $r\,(sm) = s\,(rm)$ follows from $r\,(sm) = \underbrace{(rs)}_{=sr} m = (sr)\, m = s\,(rm)$. Note how we relied on the commutativity of R here!)

We have always been assuming that R is commutative in this section. Noncommutative rings R would be a distraction at this point, but will appear later on.

Theorem 1.1.3 gave us a way to construct linear maps out of a free module by specifying their values on a basis. We can do the same for bilinear maps:

Theorem 1.2.2 (Universal property of free modules wrt bilinear maps). Let R be a commutative ring. Let M be a free R-module with basis $(m_i)_{i \in I}$. Let N be a free R-module with basis $(n_j)_{j \in J}$. Let P be a further R-module (not necessarily free). Let $p_{i,j} \in P$ be a vector for each pair $(i, j) \in I \times J$. Then, there exists a **unique** R-bilinear map $f : M \times N \to P$ such that

$$\text{each } (i, j) \in I \times J \text{ satisfies } f(m_i, n_j) = p_{i,j}.$$

Proof. Similar to the proof of Theorem 1.1.3. □

1.3. Multilinear maps

Linear and bilinear maps are the first two links in a chain of notions. Here is the general case:

> **Definition 1.3.1.** Let R be a commutative ring. Let $M_1, M_2, \ldots, M_n$ be finitely many R-modules. Let P be any R-module. A map $f : M_1 \times M_2 \times \cdots \times M_n \to P$ is said to be R-**multilinear** (or just **multilinear**) if it satisfies the following condition:
>
> - For any $i \in \{1, 2, \ldots, n\}$ and any $m_1, m_2, \ldots, m_{i-1}, m_{i+1}, m_{i+2}, \ldots, m_n$ in the respective modules (meaning that $m_k \in M_k$ for each $k \neq i$), the map
>
> $$\begin{aligned} M_i &\to P, \\ m_i &\mapsto f(m_1, m_2, \ldots, m_n) \end{aligned}$$
>
> is R-linear. That is, if we fix all arguments of f other than the i-th argument, then f is R-linear as a function of the i-th argument. This is called **linearity in the** i**-th argument**.

Thus, "bilinear" is just "multilinear for $n = 2$", whereas "linear" is "multilinear for $n = 1$".

The most famous example of a multilinear map is the determinant function

$$\begin{aligned} \det : \underbrace{R^n \times R^n \times \cdots \times R^n}_{n \text{ times}} &\to R, \\ (v_1, v_2, \ldots, v_n) &\mapsto \det(v_1, v_2, \ldots, v_n), \end{aligned}$$

where $\det(v_1, v_2, \ldots, v_n)$ means the determinant of the $n \times n$-matrix whose columns are $v_1, v_2, \ldots, v_n$.

There is a universal property of free modules with respect to multilinear maps (extending Theorem 1.1.3 and Theorem 1.2.2), which says that a multilinear map from a product of free R-modules can be defined by specifying its values on all combinations of basis elements (i.e., on all n-tuples whose all entries belong to the respective bases). I leave it to you to state and prove it.

1.4. Algebras over commutative rings ([DF, §10.1])

In this section, we fix a **commutative** ring R.

The notion of an R-**algebra** combines the notions of a ring and of an R-module, as well as connecting them by an extra axiom:

> **Definition 1.4.1.** An R-**algebra** is a set A that is endowed with
>
> - two binary operations (i.e., maps from $A \times A$ to A) that are called **addition** and **multiplication** and denoted by $+$ and $\cdot$,
> - a map $\cdot$ from $R \times A$ to A that is called **action** of R on A (and should not be confused with the multiplication map, which is also denoted by $\cdot$), and

- two elements of A that are called **zero** and **unity** and denoted by 0 and 1,

such that the following properties (the "**algebra axioms**") hold:

- The addition, the multiplication, the zero and the unity satisfy all the ring axioms (so that A becomes a ring when equipped with them).
- The addition, the action and the zero satisfy all the module axioms (so that A becomes an R-module when equipped with them).
- **Scale-invariance of multiplication:** We have

  $$r(ab) = (ra)b = a(rb) \qquad \text{for all } r \in R \text{ and } a, b \in A.$$

  Here (and in the following), we omit the $\cdot$ signs for multiplication and action (so "ab" means "$a \cdot b$", and "$r(ab)$" means "$r \cdot (ab)$").

Thus, an R-algebra is an R-module that is also a ring at the same time, with the same addition (i.e., the addition of the R-module must be identical with the addition of the ring) and the same zero, and satisfying the "scale-invariance" axiom. In other words, you get the definition of an R-algebra by throwing the definitions of an R-module and a ring together (without duplicating the addition and the zero) and requiring that the multiplication plays nice with the scaling (in the sense that scaling a product is equivalent to scaling one of its factors). Hence, in order to specify an R-algebra, it is enough to provide a set with both a ring structure and an R-module structure and show that it satisfies the "scale-invariance" axiom.

The "scale-invariance" axiom can also be restated as "the multiplication map $A \times A \to A$ is R-bilinear". More precisely, requiring that the multiplication map $A \times A \to A$ is R-bilinear includes both the scale-invariance axiom and some of the ring and module axioms.

Examples of R-algebras include the following:

- The commutative ring R itself is an R-algebra (with both multiplication and action being the usual multiplication of R).
- The zero ring $\{0\}$ is an R-algebra.
- The matrix ring $R^{n \times n}$ is an R-algebra for any $n \in \mathbb{N}$.
- The ring $\mathbb{C}$ is an $\mathbb{R}$-algebra.
- The ring $\mathbb{R}$ is a $\mathbb{Q}$-algebra.

- More generally: If a commutative ring R is a subring of a **commutative** ring S, then S becomes an R-algebra in a natural way. In fact, we already know from Lecture 8 that S becomes an R-module, and it is easy to see that this R-module can be combined with the ring structure on S to form an R-algebra.

- Even more generally: If R and S are two **commutative** rings, and if $f : R \to S$ is a ring morphism, then S becomes an R-algebra in a natural way. In fact, we already know from Lecture 8 that S becomes an R-module (this is the R-module structure on S induced by f), and it is easy to see that this R-module can be combined with the ring structure on S to form an R-algebra. This R-algebra structure on S is said to be **induced** by the morphism f.

- Yet more generally: If R and S are two commutative rings, and if $f : R \to S$ is a ring morphism, then any S-algebra A becomes an R-algebra in a natural way. In fact, we already know from Lecture 8 that A becomes an R-module (this is the R-module obtained by restricting the S-module A to R), and it is easy to see that this R-module can be combined with the ring structure on A to form an R-algebra. This is called the R-algebra obtained by **restricting** the S-algebra A to R.

  For example, the matrix ring $\mathbb{C}^{2\times 2}$ is a $\mathbb{C}$-algebra, and thus becomes an $\mathbb{R}$-algebra (since the inclusion map $\mathbb{R} \to \mathbb{C}$ is a ring morphism).

- The quaternion ring $\mathbb{H}$ is an $\mathbb{R}$-algebra. But it is not a $\mathbb{C}$-algebra, even though it contains $\mathbb{C}$ as a subring. Indeed, the "scale-invariance" axiom for $\mathbb{H}$ to be a $\mathbb{C}$-algebra would say that

  $$r\left(ab\right) = \left(ra\right)b = a\left(rb\right) \qquad \text{for all } r \in \mathbb{C} \text{ and } a, b \in \mathbb{H};$$

  but this is **not true** in general since $ji \neq ij$.

- The polynomial ring $R\left[x\right]$ (to be defined soon) is an R-algebra.

Particularly common are the $\mathbb{Z}$-algebras: In fact, every ring is a $\mathbb{Z}$-algebra in a natural way:

Proposition 1.4.2. Let R be any ring. Then, R is an abelian group (with respect to addition), so R becomes a $\mathbb{Z}$-module (since we have seen in Lecture 8 that every abelian group naturally becomes a $\mathbb{Z}$-module). This $\mathbb{Z}$-module structure can be combined with the ring structure on R, turning R into a $\mathbb{Z}$-algebra.

Proof. You have to check "scale-invariance". This is easy and LTTR. □

Thus, every ring becomes a $\mathbb{Z}$-algebra (similarly to how any abelian group becomes a $\mathbb{Z}$-module). This allows us to equate rings with $\mathbb{Z}$-algebras. We shall do this whenever convenient.

Every R-algebra A has an underlying ring (i.e., the ring obtained from A by forgetting the action) and an underlying R-module (i.e., the R-module obtained from A by forgetting the multiplication and the unity); we will refer to these simply as the "ring A" and the "R-module A". So, for example, if A and B are two R-algebras, then the "ring morphisms from A to B" will simply mean the ring morphisms from the underlying ring of A to the underlying ring of B. Similarly the "R-module morphisms from A to B" are to be understood.

Definition 1.4.3. Let A and B be two R-algebras. An **R-algebra morphism** (or, short, **algebra morphism**) from A to B means a map $f : A \to B$ that is both a ring morphism and an R-module morphism (i.e., that respects addition, multiplication, zero, unity and scaling).

An R-algebra is said to be **commutative** if its underlying ring is commutative (i.e., if its multiplication is commutative).

Algebras have subalgebras; they are defined exactly as you would expect:

Definition 1.4.4. Let A be an R-algebra. An **R-subalgebra** of A means a subset of A that is simultaneously a subring and an R-submodule of A.

In pedestrian terms, this means that an R-subalgebra of A is a subset of A that is closed under addition, multiplication and scaling and contains the zero and the unity. Such an R-subalgebra of A clearly becomes an R-algebra in its own right (since we can restrict all relevant operations from A to this subalgebra).

The direct product of several R-algebras is defined just as you would expect: addition, multiplication and scaling are all entrywise. Just for the sake of completeness, let me give its precise definition:

Proposition 1.4.5. Let I be any set. Let $(A_i)_{i\in I}$ be any family of R-algebras. Then, their Cartesian product $\prod\limits_{i\in I} A_i$ becomes an R-algebra if we endow it with the entrywise addition (i.e., we set $(m_i)_{i\in I} + (n_i)_{i\in I} = (m_i + n_i)_{i\in I}$ for any two families $(m_i)_{i\in I}, (n_i)_{i\in I} \in \prod\limits_{i\in I} A_i$) and the entrywise multiplication (i.e., we set $(m_i)_{i\in I} \cdot (n_i)_{i\in I} = (m_i \cdot n_i)_{i\in I}$ for any two families $(m_i)_{i\in I}, (n_i)_{i\in I} \in \prod\limits_{i\in I} A_i$) and the entrywise scaling (i.e., we set $r(m_i)_{i\in I} = (rm_i)_{i\in I}$ for any $r \in R$ and any family $(m_i)_{i\in I} \in \prod\limits_{i\in I} A_i$) and with the zero $(0)_{i\in I}$ and the unity $(1)_{i\in I}$. The underlying ring of this R-algebra $\prod\limits_{i\in I} A_i$ is the direct product of the rings A_i, whereas the underlying R-module of this R-algebra $\prod\limits_{i\in I} A_i$ is the direct product of the R-modules A_i.

> **Definition 1.4.6.** This R-algebra is denoted by $\prod\limits_{i\in I} A_i$ and called the **direct product** of the R-algebras A_i.
>
> The usual notations apply to these direct products: For example, if $I = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, then the direct product $\prod\limits_{i\in I} A_i$ is also denoted by $A_1 \times A_2 \times \cdots \times A_n$; we further set $A^I = \prod\limits_{i\in I} A$ and $A^n = A^{\{1,2,\ldots,n\}}$ for each $n \in \mathbb{N}$.

1.5. Defining algebras: the case of $\mathbb{H}$

An R-algebra carries more information than a ring (namely, it has the extra structure of an action). But often it is easier to define the whole R-algebra than just to define the underlying ring, because this extra structure can serve as scaffolding! Let us see an example.

Recall the ring $\mathbb{H}$ of quaternions, which were "defined" to be "numbers" of the form $a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$ and with the multiplication rules

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$

It is clear how to calculate in $\mathbb{H}$ using these rules. But why does this ring $\mathbb{H}$ exist?

Here is a cautionary tale to show why this is a question: Let's replace $k^2 = -1$ by $k^2 = 1$ in our above "definition" of $\mathbb{H}$. Then, $j^2 \underbrace{k^2}_{=1} = j^2 = -1$, so that

$$-1 = j^2k^2 = j \underbrace{jk}_{=i} k = j \underbrace{ik}_{\substack{=-j \\ (\text{since } -ik=j)}} = j(-j) = -\underbrace{j^2}_{=-1} = -(-1) = 1.$$

Adding 1 to this equality, we find $0 = 2$, so that $0 = 1$ (upon division by 2). Therefore, the ring is trivial – i.e., all its elements are 0.

Ouch. We tried to expand our number system by introducing new "numbers" i, j, k, but instead we ended up collapsing it (making all numbers equal to 0).

It should not surprise you that this can happen; after all, the same happens if you introduce the "number" $\infty := \dfrac{1}{0}$ and start doing algebra with it. But why doesn't it happen with the quaternions? Why is $\mathbb{H}$ actually an extension of our number system rather than a collapsed version of it?

The simplest way to answer this question is to throw away the wishy-washy definition of $\mathbb{H}$ we gave above (what does "numbers of the form $a + bi + cj + dk$" really mean?), and redefine $\mathbb{H}$ rigorously.

We want $\mathbb{H}$ to be an $\mathbb{R}$-algebra. First, we introduce its underlying $\mathbb{R}$-module (i.e., $\mathbb{R}$-vector space) structure. This underlying $\mathbb{R}$-module will be a 4-dimensional $\mathbb{R}$-vector space, i.e., a free $\mathbb{R}$-module of rank 4. So let me **define** $\mathbb{H}$ to be $\mathbb{R}^4$ as an $\mathbb{R}$-module. Let me denote its standard basis by $(\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k})$. These four basis vectors $\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ will later become the quaternions $1, i, j, k$, but I'm being cautious for now and avoiding any names that might be too suggestive. The basis vector $\mathbf{e}$ will be the unity of $\mathbb{H}$. Next, we define the multiplication of $\mathbb{H}$ to be the $\mathbb{R}$-bilinear map $\mu : \mathbb{H} \times \mathbb{H} \to \mathbb{H}$ that satisfies[2]

$$\begin{array}{cccc}
\mu(\mathbf{e}, \mathbf{e}) = \mathbf{e}, & \mu(\mathbf{e}, \mathbf{i}) = \mathbf{i}, & \mu(\mathbf{e}, \mathbf{j}) = \mathbf{j}, & \mu(\mathbf{e}, \mathbf{k}) = \mathbf{k}, \\
\mu(\mathbf{i}, \mathbf{e}) = \mathbf{i}, & \mu(\mathbf{i}, \mathbf{i}) = -\mathbf{e}, & \mu(\mathbf{i}, \mathbf{j}) = \mathbf{k}, & \mu(\mathbf{i}, \mathbf{k}) = -\mathbf{j}, \\
\mu(\mathbf{j}, \mathbf{e}) = \mathbf{j}, & \mu(\mathbf{j}, \mathbf{i}) = -\mathbf{k}, & \mu(\mathbf{j}, \mathbf{j}) = -\mathbf{e}, & \mu(\mathbf{j}, \mathbf{k}) = \mathbf{i}, \\
\mu(\mathbf{k}, \mathbf{e}) = \mathbf{k}, & \mu(\mathbf{k}, \mathbf{i}) = \mathbf{j}, & \mu(\mathbf{k}, \mathbf{j}) = -\mathbf{i}, & \mu(\mathbf{k}, \mathbf{k}) = -\mathbf{e}.
\end{array}$$

Theorem 1.2.2 guarantees that there is a unique such bilinear map μ. We set $ab = \mu(a, b)$ for all $a, b \in \mathbb{H}$.

Why is this a ring? All but two of the ring axioms are obvious (they follow either from the bilinearity of μ or from the module axioms for the $\mathbb{R}$-module $\mathbb{H} = \mathbb{R}^4$). The two axioms that are not obvious are the following:

1. Associativity of multiplication.
2. Neutrality of 1 (i.e., the claim that $a \cdot \mathbf{e} = \mathbf{e} \cdot a = a$ for each $a \in \mathbb{H}$).

Fortunately, the bilinearity of μ will make both of these axioms straightforward to check. Indeed, let me explain how to check the associativity of multiplication. In other words, let me prove that the map μ is associative – i.e., that

$$\mu(\mu(a, b), c) = \mu(a, \mu(b, c)) \qquad \text{for all } a, b, c \in \mathbb{H}. \tag{5}$$

The trick to this is that when a map like μ is bilinear, its associativity can be checked on a basis – or, more generally, on a spanning set:

Lemma 1.5.1. Let R be a commutative ring. Let M be an R-module. Let $(m_i)_{i \in I}$ be a family of vectors in M that spans M. Let $f : M \times M \to M$ be an R-bilinear map. If we have

$$f(f(m_i, m_j), m_k) = f(m_i, f(m_j, m_k)) \qquad \text{for all } i, j, k \in I, \tag{6}$$

then we have

$$f(f(a, b), c) = f(a, f(b, c)) \qquad \text{for all } a, b, c \in M. \tag{7}$$

[2] These equations are not chosen at random, of course; they are simply the equations

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j$$

(as well as the equations $1 \cdot 1 = 1$, $1i = i$, $1j = j$, $1k = k$, $i \cdot 1 = i$, $j \cdot 1 = j$ and $k \cdot 1 = k$), with $1, i, j, k$ renamed as $\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}$.

Proof of Lemma 1.5.1. Let $a, b, c \in M$. Since the family $(m_i)_{i \in I}$ spans M, we can write the three vectors a, b, c as

$$a = \sum_{i \in I} a_i m_i, \qquad b = \sum_{j \in I} b_j m_j, \qquad c = \sum_{k \in I} c_k m_k$$

for some coefficients $a_i, b_j, c_k \in R$. Consider these coefficients. Then,[3]

$$\begin{aligned}
f\left(f\left(a,b\right),c\right) &= f\left(f\left(\sum_{i \in I} a_i m_i, \sum_{j \in I} b_j m_j\right), \sum_{k \in I} c_k m_k\right) \\
&= \sum_{k \in I} c_k f\left(f\left(\sum_{i \in I} a_i m_i, \sum_{j \in I} b_j m_j\right), m_k\right) \\
&\qquad \text{(since } f \text{ is linear in its second argument)} \\
&= \sum_{k \in I} c_k f\left(\sum_{i \in I} a_i f\left(m_i, \sum_{j \in I} b_j m_j\right), m_k\right) \\
&\qquad \text{(since } f \text{ is linear in its first argument)} \\
&= \sum_{k \in I} c_k f\left(\sum_{i \in I} a_i \sum_{j \in I} b_j f\left(m_i, m_j\right), m_k\right) \\
&\qquad \text{(since } f \text{ is linear in its second argument)} \\
&= \sum_{k \in I} c_k \sum_{i \in I} a_i \sum_{j \in I} b_j f\left(f\left(m_i, m_j\right), m_k\right) \\
&\qquad \text{(since } f \text{ is linear in its first argument)} \\
&= \sum_{i \in I} \sum_{j \in I} \sum_{k \in I} a_i b_j c_k f\left(f\left(m_i, m_j\right), m_k\right)
\end{aligned}$$

and similarly

$$f\left(a, f\left(b,c\right)\right) = \sum_{i \in I} \sum_{j \in I} \sum_{k \in I} a_i b_j c_k f\left(m_i, f\left(m_j, m_k\right)\right).$$

The right hand sides of these two equalities are equal by our assumption (6). Hence, the left hand sides are equal. In other words, $f\left(f\left(a,b\right),c\right) = f\left(a, f\left(b,c\right)\right)$. This proves Lemma 1.5.1. □

Let us now return to $\mathbb{H}$. We want to prove that

$$\mu\left(\mu\left(a,b\right),c\right) = \mu\left(a, \mu\left(b,c\right)\right) \qquad \text{for all } a, b, c \in \mathbb{H}.$$

By Lemma 1.5.1 (applied to $R = \mathbb{R}$, $M = \mathbb{H}$, $(m_i)_{i \in I} = (\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k})$ and $f = \mu$), it suffices to show that

$$\mu\left(\mu\left(a,b\right),c\right) = \mu\left(a, \mu\left(b,c\right)\right) \qquad \text{for all } a, b, c \in \left\{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\right\}.$$

[3] We will use Proposition 1.1.2 multiple times in this computation.

This is a finite computation: There are only 64 triples (a, b, c) with $a, b, c \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$, and we can check the equality $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ for each of these triples directly by computation (using the definition of μ).

A computer could do this in the blink of an eye, but we can also do this by hand. There are some tricks that reduce our work. The first is to notice that $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ is obvious when one of a, b, c is $\mathbf{e}$ (because $\mu(\mathbf{x}, \mathbf{e}) = \mu(\mathbf{e}, \mathbf{x}) = \mathbf{x}$ for each $\mathbf{x} \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$). Thus, it suffices to prove the equality $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ in the case when $a, b, c \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. This leaves 27 triples (a, b, c) to check.

The next trick is to observe a cyclic symmetry. Indeed, the definition of μ is invariant under cyclic rotation of $\mathbf{i}, \mathbf{j}, \mathbf{k}$, in the sense that if we replace $\mathbf{i}, \mathbf{j}, \mathbf{k}$ by $\mathbf{j}, \mathbf{k}, \mathbf{i}$ (respectively), then the definition remains unchanged (for example, $\mu(\mathbf{j}, \mathbf{i}) = -\mathbf{k}$ becomes $\mu(\mathbf{k}, \mathbf{j}) = -\mathbf{i}$). Thus, when we are proving $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ for all $a, b, c \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, we can WLOG assume that $a = \mathbf{i}$ (since otherwise, we can achieve this by rotating all of a, b, c until a becomes $\mathbf{i}$). This leaves 9 triples (a, b, c) to check.

Let me just check one of them: namely, $(a, b, c) = (\mathbf{i}, \mathbf{k}, \mathbf{k})$. In this case, we have

$$\begin{aligned}\mu(\mu(a, b), c) &= \mu\left(\underbrace{\mu(\mathbf{i}, \mathbf{k})}_{=-\mathbf{j}}, \mathbf{k}\right) = \mu(-\mathbf{j}, \mathbf{k}) = -\underbrace{\mu(\mathbf{j}, \mathbf{k})}_{=\mathbf{i}} \qquad \text{(since } \mu \text{ is bilinear)} \\ &= -\mathbf{i}\end{aligned}$$

and

$$\begin{aligned}\mu(a, \mu(b, c)) &= \mu\left(\mathbf{i}, \underbrace{\mu(\mathbf{k}, \mathbf{k})}_{=-\mathbf{e}}\right) = \mu(\mathbf{i}, -\mathbf{e}) = -\underbrace{\mu(\mathbf{i}, \mathbf{e})}_{=\mathbf{i}} \qquad \text{(since } \mu \text{ is bilinear)} \\ &= -\mathbf{i}.\end{aligned}$$

Thus, $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ is proved for this triple. Similarly, the remaining $9 - 1 = 8$ triples can be checked. Thus, associativity of multiplication is proved for $\mathbb{H}$.

It remains to prove the neutrality of 1. In other words, it remains to prove that $a \cdot \mathbf{e} = \mathbf{e} \cdot a = a$ for each $a \in \mathbb{H}$. Once again, the bilinearity of the multiplication of $\mathbb{H}$ can be used to reduce this to the case when $a \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ (here we need to use Theorem 1.1.4 instead of Lemma 1.5.1); but in this case, the claim follows from our definition of μ. The details are LTTR.