Math 332 Winter 2023, Lecture 23: Modules

website: `https://www.cip.ifi.lmu.de/~grinberg/t/23wa`

2. Modules

2.10. 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. Thus, in order to define an R-algebra, it is natural to start by defining a ring and then putting the action on it (and showing that it satisfies the R-module axioms and scale-invariance).

Often, however, it is easier to proceed differently: First, define an R-module, and then define the multiplication and the unity to turn it into an R-algebra. If you do things in this order, you can use the R-module structure as scaffolding for defining the multiplication.

Here is an example of how this can work:

Recall the ring $\mathbb{H}$ of Hamilton quaternions, which were "defined" (in §1.1.2 in Lecture 2) to be "numbers" of the form $a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$ and equipped 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 in the first place?

This is not a vacuous question. For instance, if we replace the rule $k^2 = -1$ by $k^2 = 1$ in the above definition (but still require i^2 and j^2 to be -1), then we get

$$j^2 \underbrace{k^2}_{=1} = j^2 = -1$$

and thus

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

Adding 1 to this equality, we obtain $0 = 2$. Upon multiplication by $1/2$, this becomes $0 = 1$, so that our new ring is actually trivial. Instead of expanding our number system with new numbers, we have inadvertently collapsed it to a trivial ring!

Thus, if we wantonly define rings by inventing new "numbers" and declaring new rules, then we cannot expect our new rings to contain our initial rings (such as $\mathbb{R}$ in the above example) as subrings; instead, they may end up trivial or otherwise smaller than expected. This makes them rather useless.

This is similar to why you cannot divide by 0: If you introduce a new "number" $\infty = \frac{1}{0}$, then $0 \cdot \infty$ equals both 0 and 1, so you get $0 = 1$ and thus your ring is trivial.

Of course, this cautionary tale does not always have to become reality. What it shows is just that we need to be cautious: When creating new "numbers", we must make sure we don't accidentally collapse the old ones.

But how can we make this sure? For instance, how do we know that the ring $\mathbb{H}$ we defined by the above rules actually contains $\mathbb{R}$ as a subring (as opposed, e.g., to being trivial)?

One safe way of defining $\mathbb{H}$ is as follows: We define a quaternion to be a 4-tuple (a, b, c, d) of real numbers (this 4-tuple is supposed to stand for $a + bi + cj + dk$), and we define addition, multiplication and scaling of these 4-tuples explicitly by the formulas

$$(a, b, c, d) + \left(a', b', c', d'\right) = \left(a + a',\ b + b',\ c + c',\ d + d'\right)$$

and

$$\begin{aligned}(a, b, c, d)\left(a', b', c', d'\right) = (\ & aa' - bb' - cc' - dd', \\ & ab' + ba' + cd' - dc', \\ & ac' - bd' + ca' + db', \\ & ad' + bc' - cb' + da'\ )\end{aligned}$$

and

$$r\,(a, b, c, d) = (ra, rb, rc, rd) \qquad \text{for } r \in \mathbb{R}.$$

This is a valid definition, but you have to check that all the ring axioms (and module axioms, and scale-invariance) hold. Associativity of multiplication, in particular, is a lot of work. (The quaternions (a, b, c, d) thus defined can be rewritten as $a + bi + cj + dk$ once we observe that the unity of the ring is $(1, 0, 0, 0)$ and we furthermore set $i := (0, 1, 0, 0)$ and $j := (0, 0, 1, 0)$ and $k := (0, 0, 0, 1)$. Pedants will note that the ring of quaternions thus defined does not literally contain $\mathbb{R}$ as a subring, but merely contains a subring isomorphic to $\mathbb{R}$: namely, the subring $\{(a, 0, 0, 0) \mid a \in \mathbb{R}\}$. But this is not a serious problem, and the same situation occurs when defining the complex numbers.)

This definition of $\mathbb{H}$ does its job well, but as we just said, it is laborious to justify and somewhat inflexible if one is looking to generalize the construction.

A simpler and slicker way to define $\mathbb{H}$ proceeds as follows: First define $\mathbb{H}$ as an $\mathbb{R}$-module (which is easy: it will just be the free $\mathbb{R}$-module $\mathbb{R}^4$), and then build the multiplication on top of it, using the notion of bilinearity. As we saw in §2.9.1 (Lecture 22), the multiplication of an $\mathbb{R}$-algebra is always $\mathbb{R}$-bilinear, and as we saw in §2.7, if we want to define an $\mathbb{R}$-bilinear map $M \times N \to P$ where M and N are two free $\mathbb{R}$-modules, then we can do this simply by specifying its values on all pairs of basis elements (Theorem 2.7.2 in Lecture

22). Thus, instead of defining the product of any two quaternions, we will only need to define the product of two quaternions from a given basis (which we will take to be $(1, i, j, k)$).

Let us do this. We define $\mathbb{H}$ to be the $\mathbb{R}$-module $\mathbb{R}^4$, which is a free $\mathbb{R}$-module of rank 4. Thus, a quaternion is defined to be a 4-tuple (a, b, c, d) of real numbers. The addition and the scaling of quaternions are thus entrywise (since this is how $\mathbb{R}^4$ is defined).

We denote the standard basis (e_1, e_2, e_3, e_4) of $\mathbb{H}$ by $(\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k})$. The four basis vectors $\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ here will eventually be revealed (or renamed) as $1, i, j, k$, but for now I will use boldface letters to avoid suggesting too much.

Now, 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

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

By the universal property of free modules wrt bilinear maps (Theorem 2.7.2 in Lecture 22), there really is a unique such $\mathbb{R}$-bilinear map $\mu : \mathbb{H} \times \mathbb{H} \to \mathbb{H}$; thus, we have defined our μ.

We claim that the $\mathbb{R}$-module $\mathbb{H}$ becomes an $\mathbb{R}$-algebra (and thus a ring) if we endow it with the multiplication μ and the unity $\mathbf{e}$. To prove this, we need to show a few axioms. Some of them (scale-invariance, distributivity and the $0a = a0 = 0$ axiom) follow from the $\mathbb{R}$-bilinearity of μ; others follow from the fact that $\mathbb{H}$ is an $\mathbb{R}$-module. It remains to prove two axioms:

1. The map μ is associative (i.e., we have $\mu(\mu(a,b),c) = \mu(a,\mu(b,c))$ for all $a, b, c \in \mathbb{H}$).
2. The element $\mathbf{e}$ is a neutral element for μ (i.e., we have $\mu(a,\mathbf{e}) = \mu(\mathbf{e},a) = a$ for all $a \in \mathbb{H}$).

Let us prove the first axiom. Here, again, the bilinearity of μ will make our life easier. Indeed, we have the following more general fact:

Lemma 2.10.1. 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. Assume that

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

Then,

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

In other words, this lemma says that in order to prove that an R-bilinear map $f : M \times M \to M$ is associative, it suffices to prove that it is associative on a given family that spans M (i.e., that it holds whenever the three inputs a, b, c are entries of this family).

Proof of Lemma 2.10.1. See Lemma 3.12.1 in the text. In a nutshell: Expand a, b, c as R-linear combinations

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

and plug this into $f(f(a,b),c)$ and $f(a,f(b,c))$. Using the bilinearity of f, you find

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

The right hand sides of these equalities are equal (by assumption); thus, so are the left hand sides.

This proof can be summarized in two words: "by linearity". □

Having proved Lemma 2.10.1, we can now prove the associativity law $\mu(\mu(a,b),c) = \mu(a,\mu(b,c))$ for μ on $\mathbb{H}$: Indeed, by Lemma 2.10.1, it suffices to prove this law in the case when $a, b, c \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$. These are 64 equalities to prove in total (since there are 64 triples $(a,b,c) \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}^3$). This can be reduced down to 27 equalities (by realizing that if any of a, b, c is $\mathbf{e}$, then $\mu(\mu(a,b),c) = \mu(a,\mu(b,c))$ holds for easy reasons), and further down to 9 equalities (using the cyclic symmetry in $\mathbf{i}, \mathbf{j}, \mathbf{k}$ in the definition of μ). One of them is $\mu(\mu(\mathbf{i},\mathbf{k}),\mathbf{k}) = \mu(\mathbf{i},\mu(\mathbf{k},\mathbf{k}))$. By checking this and the 8 other equalities, you can convince yourself that μ is associative, i.e., that associativity holds for $\mathbb{H}$.

What about the axiom saying that $\mathbf{e}$ is a neutral element for μ ? Again, by linearity, it suffices to prove $\mu(\mathbf{e},a) = \mu(a,\mathbf{e}) = a$ only for $a \in \{\mathbf{e}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ (why?). But this is obvious from a look at the definition of μ.

Thus, we have proved that $\mathbb{H}$ is indeed a ring and even an $\mathbb{R}$-algebra. To connect this definition with the informal one we gave in §1.1.2 (Lecture 2), we just have to set

$$i = \mathbf{i}, \qquad j = \mathbf{j}, \qquad k = \mathbf{k}$$

(and observe that the unity of $\mathbb{H}$ is $\mathbf{e}$, so that $1 = \mathbf{e}$), and therefore every quaternion $(a,b,c,d) \in \mathbb{H}$ can be rewritten as

$$(a,b,c,d) = a\underbrace{\mathbf{e}}_{=1} + b\underbrace{\mathbf{i}}_{=i} + c\underbrace{\mathbf{j}}_{=j} + d\underbrace{\mathbf{k}}_{=k} = a + bi + cj + dk.$$

3. Monoid algebras and polynomials

Convention 3.0.1. For this entire chapter, we fix a **commutative** ring R.

In the previous section, we have learned how to define an R-algebra the "quick" way: Define an R-module first, and then define its multiplication as a certain R-bilinear map μ (which you can specify on the basis elements, if the R-module has a basis). Then, associativity and neutrality of the unity can be proved just by verifying them for basis elements.

Let me now apply this strategy to the construction of an important class of algebras: the **monoid algebras**, and in particular the **polynomial rings**.

3.1. Monoid algebras

3.1.1. Definition

Recall the notion of a **monoid**: Roughly speaking, it is a "group without inverses". That is, a **monoid** is a triple $(M, \cdot, 1)$, where M is a set, $\cdot$ is an associative binary operation on M, and 1 is an element of M that is neutral for $\cdot$. We will write mn for $m \cdot n$ when $m, n \in M$, and we will call mn the **product** of m and n. The monoid M is said to be **abelian** if $mn = nm$ for all $m, n \in M$. Given a monoid $(M, \cdot, 1)$, the binary operation $\cdot$ is called the **operation** of M, and the element 1 is called the **neutral element** of M. We say that the monoid M is **written multiplicatively** (or, for short, **multiplicative**) when its operation is denoted by $\cdot$, and we say that it is **written additively** (or, for short, **additive**) when its operation is denoted by $+$. Usually, the neutral element of a multiplicative monoid is denoted by 1, whereas the neutral element of an additive monoid is denoted by 0.

If M is a monoid written multiplicatively, then we can define the **monoid algebra** $R[M]$. Informally, this is the R-algebra obtained by "throwing" the elements of M "into" the ring R. Its elements are "formal R-linear combinations of elements of M", i.e., expressions of the form

$$r_1 m_1 + r_2 m_2 + \cdots + r_k m_k$$

with $k \in \mathbb{N}$ and $m_1, m_2, \ldots, m_k \in M$ and $r_1, r_2, \ldots, r_k \in R$. These expressions are multiplied by distributivity and using the multiplications of R and M: that is,

$$\left(r_1 m_1 + r_2 m_2 + \cdots + r_k m_k\right)\left(s_1 n_1 + s_2 n_2 + \cdots + s_\ell n_\ell\right) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \underbrace{r_i s_j}_{\substack{\text{product} \\ \text{in } R}} \underbrace{m_i n_j}_{\substack{\text{product} \\ \text{in } M}}.$$

In order to make this rigorous, let us recall a few concepts:

If M is any set, then R^M is the R-module

$$\left\{(r_m)_{m\in M} \mid r_m \in R \text{ for each } m \in M\right\}$$

(consisting of all families $(r_m)_{m\in M}$ of elements of R), whereas $R^{(M)}$ is the R-submodule

$$\left\{(r_m)_{m\in M} \in R^M \mid \text{all but finitely many } m \in M \text{ satisfy } r_m = 0\right\}$$

of R^M. If the set M is finite, then $R^{(M)} = R^M$.

The R-module $R^{(M)}$ is free, and the **standard basis** $(e_m)_{m\in M}$ of $R^{(M)}$ is defined as follows: For each $m \in M$, the vector $e_m \in R^{(M)}$ is the family whose m-th entry is 1 and whose all other entries are 0. (If $M = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, then this recovers the classical linear-algebraic standard basis: e.g., if $M = \{1, 2, 3\}$, then $e_1 = (1, 0, 0)$ and $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$.)

The standard basis $(e_m)_{m\in M}$ of $R^{(M)}$ is, of course, a basis of $R^{(M)}$.

We can now give a rigorous definition of the monoid algebra $R[M]$:

Definition 3.1.1. Let M be a monoid, written multiplicatively (so that $\cdot$ denotes its operation, and 1 denotes its neutral element).

The **monoid algebra** of M over R (also known as the **monoid ring** of M over R) is the R-algebra $R[M]$ defined as follows:

As an R-module, it is the free R-module

$$R^{(M)} = \left\{(r_m)_{m\in M} \in R^M \mid \text{all but finitely many } m \in M \text{ satisfy } r_m = 0\right\}.$$

Its multiplication is defined to be the unique R-bilinear map $\mu : R^{(M)} \times R^{(M)} \to R^{(M)}$ that satisfies

$$\mu(e_m, e_n) = e_{mn} \qquad \text{for all } m, n \in M.$$

Here, $(e_m)_{m\in M}$ is the standard basis of $R^{(M)}$ (that is, $e_m \in R^{(M)}$ is the family whose m-th entry is 1 and whose all other entries are 0). The unity of this R-algebra is e_1.

Theorem 3.1.2. This is indeed a well-defined R-algebra.

Proof. By linearity (and associativity of M). See Theorem 4.1.2 in the text for details. □

Recall that any group is a monoid. A group's monoid algebra is called a "group algebra":

Definition 3.1.3. If G is a group, then its monoid algebra $R[G]$ is called a **group algebra** (or **group ring**).

3.1.2. Examples

The above definition of a monoid algebra was rather abstract, so let us give some examples.

Example 3.1.4. Consider the cyclic group C_2 of order 2. We write it multiplicatively as $C_2 = \{1, u\}$ where $u^2 = 1$.

(This group is better known as $\mathbb{Z}/2$, but that would require writing it additively, which we don't want.)

How does the group algebra (= monoid algebra) $\mathbb{Q}[C_2]$ look like?

As a $\mathbb{Q}$-module, it is

$$\begin{aligned}\mathbb{Q}^{(C_2)} &= \mathbb{Q}^{(\{1,u\})} = \mathbb{Q}^{\{1,u\}} \qquad \left(\text{since } R^{(I)} = R^I \text{ when } I \text{ is finite}\right) \\ &= \left\{(r_m)_{m\in\{1,u\}} \mid r_m \in \mathbb{Q} \text{ for all } m \in \{1,u\}\right\}.\end{aligned}$$

A family of the form $(r_m)_{m\in\{1,u\}}$ contains just two entries: r_1 and r_u. By abuse of notation, we can thus identify such a family with the pair (r_1, r_u) (although, formally speaking, there is merely a bijection between the former families and the latter pairs). Thus, we can rewrite our above equality as

$$\mathbb{Q}^{(C_2)} = \{(r_1, r_u) \mid r_1, r_u \in \mathbb{Q}\} = \mathbb{Q}^2.$$

The addition and the action of the group algebra $\mathbb{Q}[C_2]$ are entrywise. What about its multiplication?

Its standard basis is $(e_m)_{m\in\{1,u\}} = (e_1, e_u)$, where the vector e_1 has 1-th entry 1 and u-th entry 0, and where the vector e_u has 1-th entry 0 and u-th entry 1. If we again encode each family $(r_m)_{m\in\{1,u\}}$ as a pair (r_1, r_u), then we can restate this as

$$e_1 = (1,0) \qquad \text{and} \qquad e_u = (0,1).$$

The multiplication of the group algebra $\mathbb{Q}[C_2]$ is given by

$$\mu(e_m, e_n) = e_{mn} \qquad \text{for all } m, n \in C_2.$$

In other words,

$$e_m e_n = e_{mn} \qquad \text{for all } m, n \in C_2.$$

Thus,

$$\begin{aligned} e_1e_1 &= e_{1\cdot 1} = e_1, \qquad & e_1e_u &= e_{1u} = e_u, \\ e_ue_1 &= e_{u\cdot 1} = e_u, \qquad & e_ue_u &= e_{uu} = e_1 \qquad \left(\text{since } uu = u^2 = 1\right). \end{aligned}$$

Since (e_1, e_u) is a basis of $\mathbb{Q}[C_2]$, we can write each element of $\mathbb{Q}[C_2]$ uniquely as $ae_1 + be_u$ for two numbers $a, b \in \mathbb{Q}$. How do we multiply two such elements? Let's see:

$$\begin{aligned}
&(ae_1 + be_u)(ce_1 + de_u) \\
&= ae_1(ce_1 + de_u) + be_u(ce_1 + de_u) \qquad \left(\begin{array}{c}\text{since the multiplication } \mu \\ \text{is } \mathbb{Q}\text{-bilinear}\end{array}\right) \\
&= a\left(c\underbrace{e_1e_1}_{=e_1} + d\underbrace{e_1e_u}_{=e_u}\right) + b\left(c\underbrace{e_ue_1}_{=e_u} + d\underbrace{e_ue_u}_{=e_1}\right) \\
&\qquad\qquad (\text{since the multiplication } \mu \text{ is } \mathbb{Q}\text{-bilinear}) \\
&= a(ce_1 + de_u) + b(ce_u + de_1) \\
&= (ac + bd)e_1 + (ad + bc)e_u. \qquad (1)
\end{aligned}$$

If we again use the notation encode each family $(r_m)_{m\in\{1,u\}}$ as a pair (r_1, r_u), then the element $ae_1 + be_u$ simply becomes (a, b) (since $e_1 = (1, 0)$ and $e_u = (0, 1)$), and thus the multiplication rule (1) rewrites as

$$(a, b)(c, d) = (ac + bd,\ ad + bc). \qquad (2)$$

This is almost the rule for multiplying complex numbers! In fact, the latter rule is

$$(a, b)(c, d) = (ac - bd,\ ad + bc),$$

which differs from (2) only in that it has a minus instead of a plus. Thus, we can think of $\mathbb{Q}[C_2]$ as a "twin brother" of $\mathbb{C}$, except that we are using rational numbers rather than real numbers as our entries (but this is not a conceptual difference; we could play the same game with $\mathbb{R}$ instead of $\mathbb{C}$).

However, it is a much less famous twin, and for a good reason: The ring $\mathbb{C}$ is a field, but the ring $\mathbb{Q}[C_2]$ is not. In fact, (2) yields

$$(1, 1)(1, -1) = (1 \cdot 1 + 1 \cdot (-1),\ 1 \cdot (-1) + 1 \cdot 1) = (0, 0)$$

in $\mathbb{Q}[C_2]$, so that $\mathbb{Q}[C_2]$ is not an integral domain, let alone a field.

Actually, we have

$$\mathbb{Q}[C_2] \cong \mathbb{Q} \times \mathbb{Q} \qquad \text{as rings and even as } \mathbb{Q}\text{-algebras.}$$

Next time, we will see why.