Math 533 Winter 2021, Lecture 11: Polynomial rings

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

1. Monoid algebra and polynomials ([DF, Chapter 9])

Let R be a commutative ring. (This will be a standing assumption throughout this chapter.)

In Lecture 10, we have learned how to define an R-algebra the "slick" way: Define an R-module first, and then define an R-bilinear map on it, which will serve as the multiplication of the algebra. Then show that the multiplication is associative (this is best done "by linearity", i.e., using the last lemma from Lecture 10) and has a unity (this can again be simplified using linearity).

I illustrated this method on the example of the ring of quaternions (an $\mathbb{R}$-algebra).

Now let me apply it to define a more important class of algebras: the monoid algebras, and, as a particular case, the polynomial rings.

1.1. Monoid algebras

Recall the notion of a **monoid**: in a nutshell, 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 $m \cdot n$ whenever $m, n \in M$. We will write M for the monoid $(M, \cdot, 1)$ if $\cdot$ and 1 are clear from the context. The monoid M is said to be **abelian** if $mn = nm$ for all $m, n \in M$. (This generalizes the notion of an abelian group.)

Here is the **idea** behind the notion of a monoid algebra**:** The monoid algebra $R[M]$ is the R-algebra obtained by "adjoining" the monoid M to the ring R, which means "inserting" the elements of M "into" R. That is, the algebra $R[M]$ consists of "formal products" rm with $r \in R$ and $m \in M$, as well as their formal sums. These products are multiplied using the multiplications of R and M:

$$(r_1 m_1) \cdot (r_2 m_2) = (r_1 r_2) \cdot (m_1 m_2) .$$

Let us formalize this:[1]

Definition 1.1.1. 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

[1] We recall that R is a commutative ring.

the R-algebra $R[M]$ defined as follows: As an R-module, it is the free R-module $R^{(M)}$. Its multiplication is defined to be the unique R-bilinear map $\mu : R^{(M)} \times R^{(M)} \to R^{(M)}$ that satisfies

$$\mu(e_m, e_n) = e_{mn} \qquad \text{for all } m, n \in M. \tag{1}$$

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 $R[M]$ is e_1.

Theorem 1.1.2. This is indeed a well-defined R-algebra.

Proof. All we need to show is that μ is associative, and that e_1 is a unity. I will only show the first statement, and leave the second to you.

We need to show that $\mu(\mu(a,b),c) = \mu(a,\mu(b,c))$ for all $a, b, c \in R[M]$. According to the last lemma from Lecture 10, it suffices to prove that

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

Let us do this: If $m, n, p \in M$, then

$$\mu\left(\underbrace{\mu(e_m, e_n)}_{=e_{mn}}, e_p\right) = \mu(e_{mn}, e_p) = e_{(mn)p} = e_{mnp}$$

and similarly $\mu(e_m, \mu(e_n, e_p)) = e_{mnp}$, so we indeed have $\mu(\mu(e_m, e_n), e_p) = \mu(e_m, \mu(e_n, e_p))$ as desired. This completes the proof that μ is associative. Thus, Theorem 1.1.2 is proven. □

Since the bilinear map μ in Definition 1.1.1 is used as the multiplication of $R[M]$, we can rewrite the equality (1) as follows:

$$e_m \cdot e_n = e_{mn} \qquad \text{for all } m, n \in M. \tag{2}$$

When a monoid M is a group, its monoid algebra $R[M]$ is called a **group algebra** (or **group ring**).

Let me show a few examples.

Example 1.1.3. Consider the order-2 cyclic group $C_2 = \{1, u\}$ with $u^2 = 1$. This group is better known as $\mathbb{Z}/2$, and its operation is commonly written as addition, not as multiplication; but we want to write it multiplicatively here, in order to match the way M is written in Definition 1.1.1.

How does the group algebra $\mathbb{Q}[C_2]$ look like? As a $\mathbb{Q}$-module (i.e., $\mathbb{Q}$-vector space), it has a basis $(e_m)_{m \in C_2} = (e_1, e_u)$. Thus, any element of $\mathbb{Q}[C_2]$ can be written as $a \underbrace{e_1}_{=1} + b e_u = a + b e_u$ for some unique $a, b \in \mathbb{Q}$.

The multiplication on $\mathbb{Q}[C_2]$ is $\mathbb{Q}$-bilinear and given on the basis by

$$e_1 e_1 = e_{1\cdot 1} = e_1, \qquad e_1 e_u = e_{1\cdot u} = e_u,$$
$$e_u e_1 = e_{u\cdot 1} = e_u, \qquad e_u e_u = e_{u\cdot u} = e_{u^2} = e_1.$$

Let us use this to compute some products in $\mathbb{Q}[C_2]$:

$$\begin{aligned}
(3+2e_u)(1+2e_u) &= 3\cdot 1 + 3\cdot 2e_u + 2e_u\cdot 1 + 2e_u\cdot 2e_u \\
&= 3 + 6e_u + 2e_u + 4\underbrace{e_u e_u}_{=e_1=1} \\
&= 3 + 6e_u + 2e_u + 4 = 7 + 8e_u; \\
(1+e_u)^2 &= 1 + 2e_u + \underbrace{e_u^2}_{=e_u e_u = e_1 = 1} = 1 + 2e_u + 1 = 2 + 2e_u; \\
(1-e_u)(1+e_u) &= 1 - \underbrace{e_u^2}_{=e_u e_u = e_1 = 1} \\
&\qquad \Big(\text{since } (1-x)(1+x) = 1 - x^2 \text{ for any } x \text{ in any ring}\Big) \\
&= 1 - 1 = 0.
\end{aligned}$$

The last of these computations shows that $\mathbb{Q}[C_2]$ is not an integral domain. In general, for any $a, b, c, d \in \mathbb{Q}$, we have

$$\begin{aligned}
(a+be_u)(c+de_u) &= ac + ade_u + b\underbrace{e_u c}_{\substack{=ce_u \\ \text{(since the multiplication of } \mathbb{Q}[C_2] \\ \text{is } \mathbb{Q}\text{-bilinear)}}} + b\underbrace{e_u d}_{\substack{=de_u \\ \text{(since the multiplication of } \mathbb{Q}[C_2] \\ \text{is } \mathbb{Q}\text{-bilinear)}}} e_u \\
&= ac + ade_u + bce_u + bd\underbrace{e_u e_u}_{=e_1=1} = ac + ade_u + bce_u + bd \\
&= (ac+bd) + (ad+bc)e_u.
\end{aligned} \tag{3}$$

How does $\mathbb{Q}[C_2]$ "look like"? Meaning, what known $\mathbb{Q}$-algebras is $\mathbb{Q}[C_2]$ isomorphic to (if any)?

I **claim** that

$$\mathbb{Q}[C_2] \cong \mathbb{Q}^2 = \mathbb{Q}\times\mathbb{Q} \qquad (\text{as } \mathbb{Q}\text{-algebras}). \tag{4}$$

[*Proof of (4):* First, we observe that $\mathbb{Q}[C_2]$ is commutative (this is easy to check), and that the element $z := \dfrac{1+e_u}{2}$ of $\mathbb{Q}[C_2]$ is idempotent (since an easy computation shows $z^2 = z$). Hence, homework set #1 exercise 3 **(d)** shows that the map

$$\begin{aligned}
f : (z\mathbb{Q}[C_2]) \times ((1-z)\mathbb{Q}[C_2]) &\to \mathbb{Q}[C_2], \\
(a,b) &\mapsto a+b
\end{aligned}$$

is a ring isomorphism; thus, this map f is invertible. This map f is furthermore $\mathbb{Q}$-linear and thus is a $\mathbb{Q}$-algebra morphism. Hence, f is a $\mathbb{Q}$-algebra isomorphism (since it is invertible). Now, what are $z\mathbb{Q}[C_2]$ and $(1-z)\mathbb{Q}[C_2]$? A general element of $\mathbb{Q}[C_2]$ has the form $a+be_u$ for some $a,b\in\mathbb{Q}$. Thus, a general element of $z\mathbb{Q}[C_2]$ has the form $z(a+be_u)$ for some $a,b\in\mathbb{Q}$. Since

$$\begin{aligned} z(a+be_u) &= \frac{1+e_u}{2}(a+be_u) = \frac{1}{2}\underbrace{(1+e_u)(a+be_u)}_{\substack{=a+e_ua+be_u+e_ube_u \\ =a+ae_u+be_u+be_ue_u}} \\ &= \frac{1}{2}\left(a+ae_u+be_u+b\underbrace{e_ue_u}_{=e_1=1}\right) = \frac{1}{2}\left((a+b)+(a+b)e_u\right) \\ &= \underbrace{(a+b)}_{\in\mathbb{Q}}z, \end{aligned}$$

we see that any such element is a **scalar** multiple of z (that is, an element of the form λz for some $\lambda\in\mathbb{Q}$, not just a multiple of z in the ring $\mathbb{Q}[C_2]$). In other words, any such element belongs to the $\mathbb{Q}$-submodule (= $\mathbb{Q}$-vector subspace)

$$\mathbb{Q}z := \{\lambda z \mid \lambda\in\mathbb{Q}\} \qquad \text{of } \mathbb{Q}[C_2].$$

Thus, $z\mathbb{Q}[C_2]\subseteq\mathbb{Q}z$. Since we also have $\mathbb{Q}z\subseteq z\mathbb{Q}[C_2]$ (since $\mathbb{Q}\subseteq\mathbb{Q}[C_2]$), this entails $z\mathbb{Q}[C_2]=\mathbb{Q}z$. Hence, in particular, $\mathbb{Q}z$ is a $\mathbb{Q}$-algebra with unity z. However, the map

$$\mathbb{Q}\to\mathbb{Q}z,\ \lambda\mapsto\lambda z$$

is a $\mathbb{Q}$-algebra morphism (indeed, it is clearly $\mathbb{Q}$-linear; it respects multiplication since $(\lambda z)(\mu z)=\lambda\mu\underbrace{z^2}_{=z}=\lambda\mu z$ for any $\lambda,\mu\in\mathbb{Q}$; its respects the unity since $1z=z$ is the unity of $\mathbb{Q}z$), and thus is a $\mathbb{Q}$-algebra isomorphism (since it is easily seen to be bijective). Thus, $\mathbb{Q}z\cong\mathbb{Q}$ as $\mathbb{Q}$-algebras. Combining this with $z\mathbb{Q}[C_2]=\mathbb{Q}z$, we obtain $z\mathbb{Q}[C_2]=\mathbb{Q}z\cong\mathbb{Q}$ as $\mathbb{Q}$-algebras. Similarly, we can prove that $(1-z)\mathbb{Q}[C_2]\cong\mathbb{Q}$ (indeed, a simple computation shows that $1-z=\dfrac{1-e_u}{2}$, and thus we can mostly repeat our above argument with $1-z$ instead of z, with the main difference being that some plus signs become minus signs).

So the isomorphism f results in

$$\mathbb{Q}[C_2]\cong\underbrace{(z\mathbb{Q}[C_2])}_{\cong\mathbb{Q}}\times\underbrace{((1-z)\mathbb{Q}[C_2])}_{\cong\mathbb{Q}}\cong\mathbb{Q}\times\mathbb{Q}=\mathbb{Q}^2.$$

This proves (4).]

Retracing our proof of (4), we actually get an explicit $\mathbb{Q}$-algebra isomorphism

$$\begin{aligned}
&\mathbb{Q}^2 \to \mathbb{Q}[C_2], \\
&(\lambda, \mu) \mapsto f(\lambda z, \mu(1-z)) = \lambda z + \mu(1-z) = \lambda \cdot \frac{1+e_u}{2} + \mu \cdot \frac{1-e_u}{2} \\
&\qquad = \frac{\lambda+\mu}{2} + \frac{\lambda-\mu}{2} e_u.
\end{aligned}$$

Example 1.1.4. (a) We can easily repeat Example 1.1.3 using the field $\mathbb{R}$ (or $\mathbb{C}$) instead of $\mathbb{Q}$. Everything works just as it did for $\mathbb{Q}$. For example, we get an $\mathbb{R}$-algebra isomorphism $\mathbb{R}^2 \to \mathbb{R}[C_2]$.

(b) Now, let us try to repeat Example 1.1.3 using the ring $\mathbb{Z}$ instead of $\mathbb{Q}$. The multiplication rule (3) still holds (but now for $a, b, c, d \in \mathbb{Z}$). What about the isomorphism (4) ? The idempotent z no longer exists (since we had to divide by 2 to construct it, but we cannot divide by 2 in $\mathbb{Z}$), so our proof of (4) does not work. And indeed, (4) does not hold for $\mathbb{Z}$. The $\mathbb{Z}$-algebra

$$\mathbb{Z}[C_2] = \{a + be_u \mid a, b \in \mathbb{Z}\}$$

is **not** isomorphic to any direct product of nontrivial $\mathbb{Z}$-algebras. This can be proved by showing that $\mathbb{Z}[C_2]$ has no idempotents other than 0 and 1. (In fact, if $a + be_u \in \mathbb{Z}[C_2]$ is an idempotent, then $(a+be_u)^2 = a + be_u$. But (3) yields $(a+be_u)^2 = \left(a^2+b^2\right) + 2abe_u$, so this idempotency results in $\left(a^2+b^2\right) + 2abe_u = a + be_u$, and thus $a^2 + b^2 = a$ and $2ab = b$ (since $e_1 = 1$ and e_u are $\mathbb{Z}$-linearly independent). But the only integer solutions (a, b) of this system of two equations are $(0, 0)$ and $(1, 0)$ (check this!); thus, the only idempotents of $\mathbb{Z}[C_2]$ are $0 + 0e_u = 0$ and $1 + 0e_u = 1$.)

Example 1.1.5. Now, let us take the order-3 cyclic group $C_3 = \{1, u, v\}$ with $u^3 = 1$ and $v = u^2$. (Again, this group is better known as $\mathbb{Z}/3$, but we write it multiplicatively.) Then, $\mathbb{Q}[C_3]$ has an idempotent $z := \dfrac{1+e_u+e_v}{3}$; this leads to a $\mathbb{Q}$-algebra isomorphism

$$\mathbb{Q}[C_3] \cong \mathbb{Q} \times S,$$

where the $\mathbb{Q}$ factor is

$$z\mathbb{Q}[C_3] = \mathbb{Q}z = \{a + ae_u + ae_v \mid a \in \mathbb{Q}\}$$

and where the S factor is

$$(1-z)\mathbb{Q}[C_3] = \{a + be_u + ce_v \mid a + b + c = 0\}.$$

The $\mathbb{Q}$ factor is 1-dimensional (as a $\mathbb{Q}$-vector space), while the S factor is 2-dimensional. Can S be decomposed further? How does S "look like"? We will later see.

Example 1.1.6. Here is a **non-example**: The ring of quaternions $\mathbb{H}$ is **not** a monoid algebra. It is pretty close, in that it has a basis $(1, i, j, k)$ (over $\mathbb{R}$) with the property that the product of any two basis elements is either a basis element again (for example, $ij = k$) or the negative of a basis element (for example, $ji = -k$). However, for it to be a monoid algebra, it would need a basis such that the product of any two basis elements is always a basis element (never the negative of a basis element).[2] Such a basis does not exist for $\mathbb{H}$.

If we remove all the minus signs in the definition of $\mathbb{H}$ (that is, we replace the multiplication rules by $i^2 = j^2 = k^2 = 1$ and $ij = ji = k$ and $jk = kj = i$ and $ki = ik = j$), then we actually do obtain a monoid algebra (namely, the group algebra of the Klein four-group).

We can find another group algebra closely related to $\mathbb{H}$. Indeed, we define the **quaternion group** Q_8 to be the subgroup $\{1, i, j, k, -1, -i, -j, -k\}$ of the group of units of $\mathbb{H}$. Then, consider the group algebra $\mathbb{H}' := \mathbb{R}[Q_8]$ of this group Q_8. This group algebra $\mathbb{H}'$ is 8-dimensional as an $\mathbb{R}$-vector space, whereas $\mathbb{H}$ is 4-dimensional; thus, $\mathbb{H}'$ is not quite $\mathbb{H}$ (but rather close). The main difference between $\mathbb{H}$ and $\mathbb{H}'$ is that the elements e_1 and e_{-1} of $\mathbb{H}'$ are two different basis elements (thus linearly independent), whereas the elements 1 and -1 of $\mathbb{H}$ are negatives of each other. Even though $\mathbb{H}'$ is not commutative, we can define a principal ideal $(e_1 + e_{-1})\,\mathbb{H}'$ of $\mathbb{H}'$ (since the element $e_1 + e_{-1}$ commutes with every element of $\mathbb{H}'$), and then it is not hard to show that the quotient ring $\mathbb{H}'/(e_1 + e_{-1})\,\mathbb{H}'$ is isomorphic to $\mathbb{H}$. Thus, while $\mathbb{H}$ itself is not a group ring, we can obtain $\mathbb{H}$ from the group ring $\mathbb{H}' = \mathbb{R}[Q_8]$ by "setting e_{-1} equal to the negative of e_1" (that is, quotienting out the principal ideal generated by $e_1 + e_{-1}$).

Convention 1.1.7. Let R be a commutative ring. Let M be a monoid. The elements e_m of the standard basis $(e_m)_{m \in M}$ of $R[M]$ will often be just denoted by m (by abuse of notation). Thus, for example, the element $a + be_u + ce_v$ of $\mathbb{Q}[C_3]$ (from Example 1.1.5) will be written as $a + bu + cv$. With this notation, an element of $R[M]$ is (at least notationally) really just a sum of products of elements of R with elements of M.

Do **not** use this convention when it causes a danger of confusion! In particular, do not use it when some elements of M include minus (or plus) signs, such as the elements $-1, -i, -j, -k$ in Example 1.1.6. (Indeed, in Example 1.1.6, it is crucial that e_1 and e_{-1} are two different basis elements of $\mathbb{H}'$, not negatives of each other. Denoting them by 1 and -1 would obscure this and risk confusing the nonzero element $e_1 + e_{-1}$ for the zero sum $1 + (-1) = 0$.)

The following properties of monoid algebras are easy:

[2] In general, you can describe a monoid algebra as an algebra that has a basis that contains the unity (i.e., the unity of the algebra belongs to the basis) and is closed under multiplication (i.e., the product of any two basis elements is again a basis element).

Proposition 1.1.8. Let M be an **abelian** monoid. Then, the monoid ring $R[M]$ is commutative.

Proof. We must prove that $ab = ba$ for all $a, b \in R[M]$. This is a typical linearity argument (just as the proof of the last lemma of Lecture 10): Since $(e_m)_{m \in M}$ is a basis of the R-module $R[M]$, we can write a and b as R-linear combinations of this family $(e_m)_{m \in M}$. That is, there exist scalars $a_m \in R$ and $b_m \in R$ for all $m \in M$ such that

$$a = \sum_{m \in M} a_m e_m \qquad \text{and} \qquad b = \sum_{m \in M} b_m e_m$$

(and such that $a_m = 0$ for all but finitely many $m \in M$, and likewise for the b_m). Multiplying these two equalities, we find

$$\begin{aligned} ab &= \left(\sum_{m \in M} a_m e_m \right) \left(\sum_{m \in M} b_m e_m \right) = \left(\sum_{m \in M} a_m e_m \right) \left(\sum_{n \in M} b_n e_n \right) \\ &\qquad \text{(here, we renamed } m \text{ as } n \text{ in the second sum)} \\ &= \sum_{m \in M} \sum_{n \in M} a_m b_n \underbrace{e_m e_n}_{\substack{= e_{mn} \\ \text{(by (2))}}} \\ &\qquad \text{(since the multiplication of the } R\text{-algebra } R[M] \text{ is } R\text{-bilinear)} \\ &= \sum_{m \in M} \sum_{n \in M} a_m b_n \underbrace{e_{mn}}_{\substack{= e_{nm} \\ \text{(since } M \text{ is abelian,} \\ \text{so that } mn = nm\text{)}}} = \sum_{m \in M} \sum_{n \in M} a_m b_n e_{nm} \end{aligned}$$

and (if we multiply them in the opposite order)

$$\begin{aligned} ba &= \left(\sum_{m \in M} b_m e_m \right) \left(\sum_{m \in M} a_m e_m \right) = \left(\sum_{n \in M} b_n e_n \right) \left(\sum_{m \in M} a_m e_m \right) \\ &\qquad \text{(here, we renamed } m \text{ as } n \text{ in the first sum)} \\ &= \underbrace{\sum_{n \in M} \sum_{m \in M}}_{= \sum_{m \in M} \sum_{n \in M}} \underbrace{b_n a_m}_{\substack{= a_m b_n \\ \text{(since } R \text{ is commutative)}}} \underbrace{e_n e_m}_{\substack{= e_{nm} \\ \text{(by (2))}}} \\ &\qquad \text{(since the multiplication of the } R\text{-algebra } R[M] \text{ is } R\text{-bilinear)} \\ &= \sum_{m \in M} \sum_{n \in M} a_m b_n e_{nm}. \end{aligned}$$

The right hand sides of these two equalities are equal; thus, so are the left hand sides. In other words, $ab = ba$. This completes the proof of Proposition 1.1.8. $\square$

Proposition 1.1.9. Let M be a monoid. The map

$$R \to R[M], \\ r \mapsto r \cdot e_1$$

is an injective R-algebra morphism.

Proof. First of all, this map is clearly injective, because the family $(e_m)_{m \in M}$ is a basis of $R[M]$ and thus is R-linearly independent (so $r \cdot e_1 \neq s \cdot e_1$ for any two distinct $r, s \in R$). It remains to prove that this map is an R-algebra morphism. But this is a particular case of the following general fact: If A is an R-algebra, then the map

$$R \to A, \\ r \mapsto r \cdot 1_A$$

is an R-algebra morphism. This fact is easy to show (for example, the map respects multiplication, since any $r, s \in R$ satisfy $(r \cdot 1_A) \cdot (s \cdot 1_A) = rs \cdot 1_A \cdot 1_A = rs \cdot 1_A$), and we can apply it to $A = R[M]$ (recalling that $1_{R[M]} = e_1$) to obtain precisely the claim we are trying to prove. □

Convention 1.1.10. If M is a monoid, then we identify each $r \in R$ with $r \cdot e_1 \in R[M]$. This identification is harmless[3], and turns R into an R-subalgebra of $R[M]$.

An element of $R[M]$ will be called **constant** if it lies in this subalgebra (i.e., if it is of the form $r \cdot e_1$ for some $r \in R$). Thus, we have identified each constant element of $R[M]$ with the corresponding element of R.

A **warning** might be in order: In Example 1.1.3, we have seen that $\mathbb{Q}[C_2] \cong \mathbb{Q} \times \mathbb{Q}$ as $\mathbb{Q}$-algebras. Now, in Convention 1.1.10, we have identified $\mathbb{Q}$ with a $\mathbb{Q}$-subalgebra of $\mathbb{Q}[C_2]$. But this subalgebra is not one of the two $\mathbb{Q}$ factors in $\mathbb{Q}[C_2] \cong \mathbb{Q} \times \mathbb{Q}$. Indeed, as a $\mathbb{Q}$-subalgebra, it contains the unity of $\mathbb{Q}[C_2]$, but none of the two $\mathbb{Q}$ factors does.

Proposition 1.1.11. Let M be a monoid. The map

$$M \to R[M], \\ m \mapsto e_m$$

is a monoid morphism from M to $(R[M], \cdot, 1)$.

[3] Indeed, Proposition 1.1.9 shows that the map $R \to R[M]$ sending each $r \in R$ to $r \cdot e_1$ is an injective R-algebra morphism. Thus, this map keeps distinct elements of R distinct in $R[M]$ (since it is injective), and respects addition and multiplication (since it is an R-algebra morphism).

Proof. This map respects multiplication (because of (2)) and sends the neutral element of M to the unity of $R[M]$ (since e_1 is the unity of $R[M]$). Thus, it is a monoid morphism. □

Note that if we use Convention 1.1.7, then the "$m \mapsto e_m$" in Proposition 1.1.11 can be rewritten as "$m \mapsto m$", so the map from Proposition 1.1.11 looks like an inclusion map. This is merely an artefact of our notation. In truth, the element m of the monoid M is not literally the same as the corresponding basis element e_m of the monoid algebra $R[M]$; we have just agreed to call both of them m for brevity. But Proposition 1.1.11 shows that using the same letter for these two elements is a mostly harmless abuse of notation. The only possible problem it can cause is when the map in Proposition 1.1.11 fails to be injective, so we might accidentally equate two distinct elements m, n of M whose corresponding basis elements e_m and e_n are equal. Fortunately, this can only happen if the ring R is trivial (indeed, for any nontrivial ring R, the basis elements e_m for $m \in M$ are distinct), and this is not a very interesting case. (This is also an issue that rarely comes up in practice. The purpose of Convention 1.1.7 is to simplify computations in $R[M]$, not to "pull" them back into M.)

1.2. Polynomial rings

Now, we can effortlessly define polynomial rings. Recall that R is a commutative ring. Recall also that $\mathbb{N} = \{0, 1, 2, \ldots\}$ (so $0 \in \mathbb{N}$).

Definition 1.2.1. Let C be the free monoid with a single generator x. This is the monoid whose elements are countably many distinct symbols named

$$x^0, \quad x^1, \quad x^2, \quad x^3, \quad \ldots;$$

its monoid operation is defined by

$$x^i \cdot x^j = x^{i+j} \qquad \text{for all } i, j \in \mathbb{N}.$$

We write this monoid multiplicatively, but of course it is just the well-known monoid $(\mathbb{N}, +, 0)$ in new clothes (we have renamed each $i \in \mathbb{N}$ as x^i; we have renamed addition as multiplication). Its neutral element is x^0. We set $x = x^1$ (so that x^i really is the i-th power of x).

The elements of C are called **monomials** (in the variable x). The specific element x is called the **indeterminate**.

Now, the **univariate polynomial ring** $R[x]$ over R is defined to be the monoid algebra $R[C]$. Following Convention 1.1.7, we simply write m for e_m when $m \in C$ (that is, we write x^i for the basis element e_{x^i}); thus, $R[x]$ is a free R-module with basis

$$\left(x^0, x^1, x^2, x^3, \ldots\right) = \left(1, x, x^2, x^3, \ldots\right).$$

This means that any $p \in R[x]$ can be written as a finite R-linear combination of powers of x. That is:

$$p = a_0x^0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for some $n \in \mathbb{N}$ and some $a_0, a_1, \ldots, a_n \in R$. This representation is unique up to trailing zeroes (meaning that we can always add $0x^{n+1}$ addends – e.g., rewriting $4x^0 + 3x^1$ as $4x^0 + 3x^1 + 0x^2$ –, but other than that it is unique).

Elements of $R[x]$ are called **polynomials** in x over R.

Thus, up to notation, the univariate polynomial ring $R[x]$ is just the monoid ring $R[\mathbb{N}]$ of the abelian monoid $\mathbb{N} = (\mathbb{N}, +, 0)$. Hence, this ring $R[x]$ is commutative (by Proposition 1.1.8, since the monoid $\mathbb{N}$ is abelian).

Example 1.2.2. (a) Here is an example of a polynomial:

$$1 + 3x^2 + 6x^3 = 1e_{x^0} + 3e_{x^2} + 6e_{x^3} \in R[x].$$

(b) A non-example: The infinite sum $1 + x + x^2 + x^3 + \cdots$ is **not** in $R[x]$. Indeed, polynomials are linear combinations of powers of x, and linear combinations are finite (by definition); even if you write them as infinite sums, they are de-facto finite because all but finitely many addends are 0. Infinite sums $1 + x + x^2 + x^3 + \cdots$ are thus not polynomials; they are known as **formal power series**. There is a way to define an R-algebra of formal power series, too, but we won't do so now.

So we have defined **univariate** polynomial rings (i.e., polynomial rings in a single variable). Likewise, we can define **multivariate** polynomial rings – i.e., polynomial rings in several variables. For simplicity, let me restrict myself to finitely many variables:

Definition 1.2.3. Let $n \in \mathbb{N}$. Let $C^{(n)}$ be the free abelian monoid with n generators $x_1, x_2, \ldots, x_n$. This is the monoid whose elements are the distinct symbols

$$x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n} \qquad \text{with } (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n;$$

its monoid operation is defined by

$$\left(x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}\right) \cdot \left(x_1^{b_1}x_2^{b_2}\cdots x_n^{b_n}\right) = x_1^{a_1+b_1}x_2^{a_2+b_2}\cdots x_n^{a_n+b_n}$$
$$\text{for all } (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n \text{ and } (b_1, b_2, \ldots, b_n) \in \mathbb{N}^n.$$

We write it multiplicatively, but of course this is just the monoid $\mathbb{N}^n = (\mathbb{N}^n, +, 0)$ in disguise (where the addition on $\mathbb{N}^n$ that we are calling "+" here is entrywise, and 0 means the n-tuple $(0, 0, \ldots, 0)$), with each element $(a_1, a_2, \ldots, a_n)$ renamed as $x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$ and with addition renamed

as multiplication. The elements of $C^{(n)}$ are called **monomials**. For each $i \in \{1, 2, \ldots, n\}$, we define a monomial x_i by

$$x_i = x_1^0 x_2^0 \cdots x_{i-1}^0 x_i^1 x_{i+1}^0 x_{i+2}^0 \cdots x_n^0.$$

These specific elements $x_1, x_2, \ldots, x_n$ are called the **indeterminates**. It is easy to see that any monomial $x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \in C^{(n)}$ is indeed the product of the powers $x_1^{a_1}, x_2^{a_2}, \ldots, x_n^{a_n}$, just as the notation suggests.

Now, the R-algebra $R[x_1, x_2, \ldots, x_n]$ is defined to be the monoid algebra $R\left[C^{(n)}\right]$. It is commonly called the **polynomial ring in n variables $x_1, x_2, \ldots, x_n$ over R**. Following Convention 1.1.7, we simply write m for e_m whenever $m \in C^{(n)}$; thus, $R[x_1, x_2, \ldots, x_n]$ is a free R-module with basis

$$\left(x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}\right)_{(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n}.$$

This means that any $p \in R[x_1, x_2, \ldots, x_n]$ can be uniquely written as an R-linear combination

$$p = \sum_{(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n} p_{a_1, a_2, \ldots, a_n} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

with $p_{a_1, a_2, \ldots, a_n} \in R$ (where all but finitely many of these coefficients $p_{a_1, a_2, \ldots, a_n}$ are 0).

Elements of $R[x_1, x_2, \ldots, x_n]$ are called **polynomials** in $x_1, x_2, \ldots, x_n$.

Thus, up to notation, the multivariate polynomial ring $R[x_1, x_2, \ldots, x_n]$ is just the monoid algebra $R[\mathbb{N}^n]$ of the abelian monoid $\mathbb{N}^n = (\mathbb{N}^n, +, 0)$.

The multivariate polynomial ring $R[x_1, x_2, \ldots, x_n]$ is commutative (by Proposition 1.1.8, since the monoid $\mathbb{N}^n$ is abelian).

The univariate polynomial ring $R[x]$ can be viewed as a particular case of the multivariate polynomial ring $R[x_1, x_2, \ldots, x_n]$ (obtained by taking $n = 1$ and renaming x_1 as x).

Polynomials as formal linear combinations are already useful and nice. But they become a much stronger tool once you learn how to evaluate them, i.e., substitute things into them. Unlike a function, a univariate polynomial over R does not have a fixed domain; you can substitute an element of R into it, but also a square matrix over R or even another polynomial, and more generally, any element of an R-algebra:

Definition 1.2.4. Let $p \in R[x]$ be a univariate polynomial. Let A be any R-algebra. Let $a \in A$.

We define the element $p(a) \in A$ as follows: Write p as

$$p = \sum_{i \in \mathbb{N}} p_i x^i$$

with $p_i \in R$ (where $p_i = 0$ for all but finitely many $i \in \mathbb{N}$), and set

$$p(a) = \sum_{i \in \mathbb{N}} p_i a^i.$$

This element $p(a)$ is called the **evaluation** of p at a; we say that it is obtained by **substituting** a for x in p.

Sometimes I will denote it by $p[a]$ instead of $p(a)$ (for reasons explained below).

Note that the $p_i \in R$ in Definition 1.2.4 are unique, since $\left(x^0, x^1, x^2, \ldots\right)$ is a basis of the R-module $R[x]$.

As I said, A can be any R-algebra in Definition 1.2.4: for example, R itself, or a matrix ring $R^{n \times n}$, or the polynomial ring $R[x]$. In particular, we can substitute x for x in p, obtaining $p(x) = p$.

Warning: The notation $p(a)$ in Definition 1.2.4 has potential for confusion: Is $p(p+1)$ the evaluation of p at $p+1$ or the product of p with $p+1$? This is why I prefer the notation $p[a]$ instead of $p(a)$. I also recommend using $\cdot$ for products whenever such confusion could arise (thus, write $p \cdot (p+1)$ if you mean the product of p with $p+1$). When reading algebra literature, be aware that you will sometimes have to look at the context and make sanity checks.

Example 1.2.5. Let $R = \mathbb{Z}/2$, and let p be the polynomial $x^2 + x = x \cdot \left(x + \overline{1}\right) \in R[x]$. Let us evaluate p at elements of R:

$$p\left(\overline{0}\right) = \overline{0}^2 + \overline{0} = \overline{0};$$
$$p\left(\overline{1}\right) = \overline{1}^2 + \overline{1} = \overline{1} + \overline{1} = \overline{2} = \overline{0}.$$

Thus, the polynomial p gives $\overline{0}$ when evaluated at any element of $\mathbb{Z}/2$, even though p is not zero as a polynomial. If you want a nonzero evaluation of p, one thing you can do is to evaluate it on a square matrix:

$$p\begin{pmatrix} \overline{0} & \overline{1} \\ \overline{1} & \overline{0} \end{pmatrix} = \begin{pmatrix} \overline{0} & \overline{1} \\ \overline{1} & \overline{0} \end{pmatrix}^2 + \begin{pmatrix} \overline{0} & \overline{1} \\ \overline{1} & \overline{0} \end{pmatrix} = \begin{pmatrix} \overline{1} & \overline{1} \\ \overline{1} & \overline{1} \end{pmatrix} \neq 0_{2 \times 2}.$$

(Or you can evaluate it at x, getting $p(x) = p \neq 0$.)

Theorem 1.2.6. Let A be an R-algebra. Let $a \in A$. Then, the map

$$\begin{aligned} R[x] &\to A, \\ p &\mapsto p[a] \end{aligned}$$

is an R-algebra morphism. In particular, for any two polynomials $p, q \in R[x]$,

we have

$$(pq)[a] = p[a] \cdot q[a];$$
$$(p+q)[a] = p[a] + q[a].$$

The proof of this theorem will be easiest to do after showing the following simple lemma (compare with the last lemma of Lecture 10):

Lemma 1.2.7. Let R be a commutative ring. Let A and B be two R-algebras. Let $f : A \to B$ be an R-linear map. Let $(m_i)_{i \in I}$ be a family of vectors in A that spans A. If we have

$$f(m_i m_j) = f(m_i) f(m_j) \qquad \text{for all } i, j \in I, \tag{5}$$

then we have

$$f(ab) = f(a) f(b) \qquad \text{for all } a, b \in A. \tag{6}$$

Proof of Lemma 1.2.7. Let $a, b \in A$. Since the family $(m_i)_{i \in I}$ spans A, we can write the two vectors a and b as

$$a = \sum_{i \in I} a_i m_i \qquad \text{and} \qquad b = \sum_{j \in I} b_j m_j \tag{7}$$

for some coefficients a_i and b_j in R. Consider these coefficients. Hence,

$$ab = \left(\sum_{i \in I} a_i m_i \right) \left(\sum_{j \in I} b_j m_j \right) = \sum_{i \in I} \sum_{j \in I} a_i b_j m_i m_j$$

(since the multiplication of A is R-bilinear) and thus

$$f(ab) = f\left(\sum_{i \in I} \sum_{j \in I} a_i b_j m_i m_j \right) = \sum_{i \in I} \sum_{j \in I} a_i b_j \underbrace{f(m_i m_j)}_{\substack{= f(m_i) f(m_j) \\ \text{(by (5))}}} \qquad \text{(since } f \text{ is } R\text{-linear)}$$

$$= \sum_{i \in I} \sum_{j \in I} a_i b_j f(m_i) f(m_j) = \left(\sum_{i \in I} a_i f(m_i) \right) \left(\sum_{j \in I} b_j f(m_j) \right).$$

Comparing this with

$$f(a) f(b) = f\left(\sum_{i \in I} a_i m_i \right) f\left(\sum_{j \in I} b_j m_j \right) \qquad \text{(by (7))}$$
$$= \left(\sum_{i \in I} a_i f(m_i) \right) \left(\sum_{j \in I} b_j f(m_j) \right) \qquad \text{(since } f \text{ is } R\text{-linear)},$$

we obtain $f(ab) = f(a) f(b)$. But this is precisely what we wanted to prove. □

Proof of Theorem 1.2.6. Let f be the map

$$\begin{aligned} R[x] &\to A, \\ p &\mapsto p[a]. \end{aligned}$$

We must show that f is an R-algebra morphism. It is easy to see that f is R-linear. (For example, in order to show that it respects addition, you need to check that $(p+q)[a] = p[a] + q[a]$ for any $p, q \in R[x]$. But this is done exactly as you would think: Write p and q as $p = \sum_{i \in \mathbb{N}} p_i x^i$ and $q = \sum_{i \in \mathbb{N}} q_i x^i$, and conclude that

$$p + q = \sum_{i \in \mathbb{N}} p_i x^i + \sum_{i \in \mathbb{N}} q_i x^i = \sum_{i \in \mathbb{N}} \left(p_i x^i + q_i x^i \right) = \sum_{i \in \mathbb{N}} (p_i + q_i) x^i,$$

so that

$$\begin{aligned} (p+q)[a] &= \sum_{i \in \mathbb{N}} (p_i + q_i) a^i \qquad \text{(by the definition of } (p+q)[a]\text{)} \\ &= \sum_{i \in \mathbb{N}} p_i a^i + \sum_{i \in \mathbb{N}} q_i a^i; \end{aligned}$$

but it is just as easy to see that $p[a] + q[a]$ gives the same result.)

It is furthermore clear that the map f respects the unity; indeed, $f(1) = 1[a] = 1$ (since substituting a for x in the polynomial $1 = 1x^0 + 0x^1 + 0x^2 + \cdots$ results in $1a^0 + 0a^1 + 0a^2 + \cdots = 1$). All that now remains is to show that f respects multiplication. In other words, it remains to show that $f(pq) = f(p) f(q)$ for all $p, q \in R[x]$. Lemma 1.2.7 gives us a shortcut to proving this: Since the family $\left(x^i\right)_{i \in \mathbb{N}}$ is a basis of the R-module $R[x]$ (and thus spans this R-module), and since we already know that f is R-linear, it suffices to show that

$$f\left(x^i x^j\right) = f\left(x^i\right) f\left(x^j\right) \qquad \text{for all } i, j \in \mathbb{N} \tag{8}$$

(because if we can show this, then Lemma 1.2.7 will yield that $f(pq) = f(p) f(q)$ for all $p, q \in R[x]$).

So let us prove (8). Fix $i, j \in \mathbb{N}$. Then, $x^i[a] = a^i$ (because substituting a for x in the polynomial $x^i = 0x^0 + 0x^1 + \cdots + 0x^{i-1} + 1x^i + 0x^{i+1} + 0x^{i+2} + \cdots$ results in $0a^0 + 0a^1 + \cdots + 0a^{i-1} + 1a^i + 0a^{i+1} + 0a^{i+2} + \cdots = a^i$) and similarly $x^j[a] = a^j$ and $x^{i+j}[a] = a^{i+j}$. But $x^i x^j = x^{i+j}$, so that

$$\begin{aligned} f\left(x^i x^j\right) = f\left(x^{i+j}\right) &= x^{i+j}[a] \qquad \text{(by the definition of } f\text{)} \\ &= a^{i+j} = \underbrace{a^i}_{\substack{=x^i[a] \\ =f\left(x^i\right)}} \underbrace{a^j}_{\substack{=x^j[a] \\ =f\left(x^j\right)}} = f\left(x^i\right) f\left(x^j\right). \end{aligned}$$

(by the definition of f) (by the definition of f)

This proves (8), and thus concludes the proof of Theorem 1.2.6. □

Likewise, we can substitute multiple elements into a multivariate polynomial, as long as these elements commute:

Definition 1.2.8. Let $n \in \mathbb{N}$. Let $p \in R[x_1, x_2, \ldots, x_n]$ be a multivariate polynomial. Let A be any R-algebra. Let $a_1, a_2, \ldots, a_n \in A$ be n elements of A that mutually commute (i.e., that satisfy $a_i a_j = a_j a_i$ for each $i, j \in \{1, 2, \ldots, n\}$).

We define the element $p(a_1, a_2, \ldots, a_n) \in A$ as follows: Write the polynomial p as

$$p = \sum_{(i_1, i_2, \ldots, i_n) \in \mathbb{N}^n} p_{i_1, i_2, \ldots, i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$$

with $p_{i_1, i_2, \ldots, i_n} \in R$ (where $p_{i_1, i_2, \ldots, i_n} = 0$ for all but finitely many $(i_1, i_2, \ldots, i_n) \in \mathbb{N}^n$), and set

$$p(a_1, a_2, \ldots, a_n) = \sum_{(i_1, i_2, \ldots, i_n) \in \mathbb{N}^n} p_{i_1, i_2, \ldots, i_n} a_1^{i_1} a_2^{i_2} \cdots a_n^{i_n}.$$

This element $p(a_1, a_2, \ldots, a_n)$ is called the **evaluation** of p at $a_1, a_2, \ldots, a_n$; we say that it is obtained by **substituting** $a_1, a_2, \ldots, a_n$ for $x_1, x_2, \ldots, x_n$ in p.

The analogue to Theorem 1.2.6 now is the following:

Theorem 1.2.9. Let $n \in \mathbb{N}$. Let A be an R-algebra. Let $a_1, a_2, \ldots, a_n \in A$ be n elements of A that mutually commute. Then, the map

$$\begin{aligned} R[x_1, x_2, \ldots, x_n] &\to A, \\ p &\mapsto p(a_1, a_2, \ldots, a_n) \end{aligned}$$

is an R-algebra morphism.

Proof. Similar to the proof of Theorem 1.2.6, but a bit more notationally sophisticated due to the presence of multiple variables. □

Finally, a few more pieces of notation. We recall the notion of a constant element of a monoid ring (Convention 1.1.10). Since a polynomial ring is a monoid ring, we can apply it to polynomial rings, and obtain the following:

Convention 1.2.10. Let $n \in \mathbb{N}$. Then, we identify each $r \in R$ with $r \cdot 1 \in R[x_1, x_2, \ldots, x_n]$ (where 1 means the monomial $x_1^0 x_2^0 \cdots x_n^0$, which is the unity of $R[x_1, x_2, \ldots, x_n]$). This identification is harmless, and turns R into an R-subalgebra of $R[x_1, x_2, \ldots, x_n]$.

A polynomial $p \in R[x_1, x_2, \ldots, x_n]$ is said to be **constant** if it lies in this subalgebra (i.e., if it satisfies $p = r \cdot 1$ for some $r \in R$).

Example 1.2.11. The polynomial $3 = 3x^0 \in R[x]$ is constant, but the polynomial $3x = 3x^1$ is not.

Definition 1.2.12. Let $p \in R[x_1, x_2, \ldots, x_n]$ be a polynomial. Let $\mathfrak{m} = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$ be a monomial. Then, the **coefficient** of $\mathfrak{m}$ in p is the element $[\mathfrak{m}]\, p$ of R defined as follows: If we write p as

$$p = \sum_{(i_1, i_2, \ldots, i_n) \in \mathbb{N}^n} p_{i_1, i_2, \ldots, i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$$

with $p_{i_1, i_2, \ldots, i_n} \in R$, then

$$[\mathfrak{m}]\, p = p_{a_1, a_2, \ldots, a_n}.$$

Example 1.2.13. (a) For univariate polynomials, we have

$$\left[x^3\right]\left((1+x)^5\right) = 10 \qquad \text{and} \qquad \left[x^7\right]\left((1+x)^5\right) = 0$$

(since $(1+x)^5 = 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$).

(b) For multivariate polynomials, we have

$$\left[x_1^2 x_2^3\right]\left((x_1+x_2)^5\right) = 10 \qquad \text{and} \qquad [x_1]\left((x_1+x_2)^5\right) = 0$$

(since $(x_1+x_2)^5 = x_1^5 + 5x_1^4 x_2 + 10x_1^3 x_2^2 + 10x_1^2 x_2^3 + 5x_1 x_2^4 + x_2^5$).

Often we will want to use symbols other than $x_1, x_2, \ldots, x_n$ for indeterminates. For example, we can rename the indeterminates x_1 and x_2 of the polynomial ring $R[x_1, x_2]$ as x and y, so that the equations in Example 1.2.13 **(b)** become

$$\left[x^2 y^3\right]\left((x+y)^5\right) = 10 \qquad \text{and} \qquad [x]\left((x+y)^5\right) = 0.$$

When we do this, of course, we need to also rename the ring $R[x_1, x_2]$ as $R[x, y]$. More generally, we can have polynomial rings in any set of indeterminates; these are written by putting the names of these indeterminates into the square brackets. For example, $R[a, b, x, y]$ means a polynomial ring in four indeterminates named a, b, x, y.

It is actually helpful to think of polynomial rings with differently named indeterminates as distinct – e.g., the rings $R[x]$ and $R[y]$ are distinct (but isomorphic). This allows us to view them both as subrings of $R[x, y]$ without actually equating x with y.

1.3. Univariate polynomials

1.3.1. Degrees and coefficients

Let us now take a closer look at univariate polynomials (which are the best-behaved among the polynomials).

Definition 1.3.1. Let $p \in R[x]$ be a univariate polynomial.

(a) If $p \neq 0$, then the **degree** of p is defined to be the largest $i \in \mathbb{N}$ such that $\left[x^i\right] p \neq 0$. The degree of the zero polynomial $0 \in R[x]$ is defined to be $-\infty$ (a symbol subject to the rules $-\infty < m$ and $-\infty + m = -\infty$ for any $m \in \mathbb{Z}$).

The degree of p is denoted by $\deg p$.

(b) If $p \neq 0$, then the **leading coefficient** of p is defined to be the coefficient $\left[x^{\deg p}\right] p \in R$.

(c) The polynomial p is said to be **monic** (or, as some say, **normalized**) if its leading coefficient is 1.

For example, the polynomial

$$5x^3 + 2x + 1 \in \mathbb{Q}[x]$$

has degree 3 and leading coefficient 5 and is not monic (since $5 \neq 1$). The polynomial

$$\overline{5}x^3 + \overline{2}x + \overline{1} \in (\mathbb{Z}/n)[x]$$

has

- degree 3 if $n > 5$;
- degree 1 if $n = 5$ (because if $n = 5$, then the $\overline{5}x^3 = \overline{0}x^3$ term disappears);
- degree 3 if $n = 2, 3, 4$; and
- degree $-\infty$ if $n = 1$.

(Degrees are somewhat unstable for trivial rings.)

The polynomial $(1 + x)^3 = 1 + 3x + 3x^2 + x^3$ is monic (i.e., has leading coefficient 1) and has degree 3.