Polynomials

1 More properties of polynomials

Recall that, for \( R \) a commutative ring with unity (as with all rings in this course unless otherwise noted), we define \( R[x] \) to be the set of expressions \( \sum_{i=0}^{n} a_i x^i \), where \( a_i \in R \), with the understanding that two such expressions agree if they differ by terms of the form \( 0x^k \). Alternatively, we could identify a polynomial with an infinite sequence \( a_0, a_1, \ldots \), such that \( a_i \in R \) and only finitely many of the \( a_i \) are non-zero. Addition and multiplication of polynomials are defined as follows:

\[
\sum_i a_i x^i + \sum_i b_i x^i = \sum_i (a_i + b_i)x^i; \\
\left( \sum_i a_i x^i \right) \left( \sum_i b_i x^i \right) = \sum_k \left( \sum_{i+j=k} a_i b_j \right).
\]

Note that, with this definition, \( x^i x^j = x^{i+j} \), and hence \( x^i = \underbrace{x \cdot x \cdots x}_{i \text{ times}} \).

Thus the two meanings of \( x^i \) are consistent. We will use symbols such as \( f, g, p, q \) for polynomials, unlike the more usual notations \( f(x) \), etc. in order to emphasize that polynomials are formal or symbolic objects. We will discuss the various ways in which we can think of polynomials as functions later.

It is routine to check that \( (R[x], +) \) is an abelian group. To check that it is a ring, we must check that multiplication is associative and commutative, that the left distributive law holds (we don’t have to check both laws since multiplication is commutative), and that there is a unity. We will just check associativity. With

\[
f = \sum_i a_i x^i; \quad g = \sum_i b_i x^i; \quad h = \sum_i c_i x^i,
\]

1
a calculation shows that

\[(fg)h = \sum_{\ell} \left( \sum_{i+j+k=\ell} (a_i b_j) c_k \right) x^\ell,\]

and similarly

\[f(gh) = \sum_{\ell} \left( \sum_{i+j+k=\ell} a_i (b_j c_k) \right) x^\ell.\]

Thus \((fg)h = f(gh)\) since multiplication in \(R\) is associative. Note that \(R\) is a subring of \(R[x]\), with

\[r \left( \sum_i a_i x^i \right) = \sum_i ra_i x^i,\]

and in particular \(1 \in R\) is the unity in \(R[x]\).

**Remark 1.1.** The definition of addition and multiplication in \(R[x]\) is essentially forced by requiring associativity, commutativity, and distributivity. For example, we must have \(ax^i + bx^i = (a + b)x^i\). Likewise, we must have \((ax^i)(bx^j) = abx^{i+j}\), provided that we interpret \(x^i\) as \((x)^i\), the product of the ring element \(x\) with itself \(i\) times.

As previously stated, the *degree* of a polynomial \(f = \sum_i a_i x^i\) is the largest integer \(d\) such that \(a_d \neq 0\). The degree of the zero polynomial 0 is undefined. Note that, if \(a \in R, a \neq 0\), then \(\deg a = 0\), and in fact the subring \(R\) of \(R[x]\) is given by

\[R = \{ f \in R[x] : \deg f = 0 \text{ or } f = 0\} .\]

We shall sometimes refer to the elements of \(R\) as *constant polynomials* or *constants*.

If \(f = \sum_{i=0}^d a_i x^i\) is a polynomial of degree \(d\) and \(g = \sum_{i=0}^e b_i x^i\) is a polynomial of degree \(e\) and, say, \(d > e\), then clearly the degree of \(f \cdot g\) is \(d\). But if \(f\) and \(g\) both have the same degree \(d\), the term \((a_d + b_d)x^d\) might be 0, if \(b_d = -a_d\), and hence in this case \(\deg(f + g) < d\), or is undefined if \(g = -f\). Thus, if \(f, g, f + g \neq 0\), then

\[\deg(f + g) \leq \max(\deg f, \deg g).\]

Similarly, if \(f\) and \(g\) are as above, then the highest degree term of \(fg\) is \(a_d b_e x^{d+e}\), unless \(a_d b_e = 0\). Hence, if \(f, g, fg \neq 0\), then

\[\deg(fg) \leq \deg f + \deg g.\]
Example 1.2. In \((\mathbb{Z}/6\mathbb{Z})[x]\),
\[(2x + 1)(3x^2 + 1) = 3x^2 + 2x + 1,
\]since the “leading term” \((2x)(3x^2) = 0\), and hence the product does not have the expected degree \(1 + 2 = 3\). Even worse, \(2(3x^2 + 3) = 0\).

In \((\mathbb{Z}/4\mathbb{Z})[x]\), \((2x + 1)^2 = 4x^2 + 4x + 1 = 1\). Thus, not only does \((2x + 1)^2\) not have the expected degree, but we also see that \(2x + 1\) is a unit, i.e. there are rings \(R\) such that the group of units \((R[x])^*\) is larger than the units \(R^*\) in \(R\).

Polynomials in several variables can be defined similarly. For example, an element of \(R[x_1,x_2]\), i.e. a polynomial in the two variables \(x_1, x_2\), is an expression of the form \(\sum_{i,j \geq 0} a_{ij} x_1^i x_2^j\), where the \(a_{ij} \in R\), and only finitely many are nonzero. By grouping such terms in powers of \(x_2\), we see that \(R[x_1,x_2] \cong R[x_1][x_2]\). In other words, a polynomial in \(x_1\) and \(x_2\) is the same thing as a polynomial in \(x_2\) whose coefficients are polynomials in \(x_1\). Similarly \(R[x_1,x_2] \cong R[x_2][x_1]\) by grouping in powers of \(x_1\). Inductively, we can define polynomials in \(n\) variables via
\[
R[x_1,\ldots,x_n] = R[x_1,\ldots,x_{n-1}][x_n].
\]

2 Polynomials as functions

A polynomial with real coefficients \(f = \sum_i a_i x^i\) defines a function \(f: \mathbb{R} \to \mathbb{R}\) by defining \(f(t) = \sum_i a_i t^i\). (Typically, we speak of \(x\) as the “variable,” not just some formal symbol.) We can do the same thing in a general ring: given \(r \in R\), we define the evaluation \(\text{ev}_r\) of a polynomial \(f = \sum_i a_i x^i\) at \(r\), and write it as \(\text{ev}_r(f)\) or sometimes as \(f(r)\), by the formula
\[
\text{ev}_r(f) = \sum_i a_i r^i \in R.
\]
Informally, \(\text{ev}_r(f)\) is obtained from \(f\) by “plugging in \(r\ for \(x\).” In this way, an element \(f \in R[x]\) also defines a function from \(R\) to \(R\), which we denote by \(E(f)\), via the formula
\[
E(f)(r) = f(r) = \text{ev}_r(f).
\]
For example, if \(a \in R \leq R[x]\) is a constant polynomial, then \(\text{ev}_r(a) = a\) and \(E(f)\) is the constant function from \(R\) to itself whose value is always \(a\). Likewise, \(\text{ev}_r(x) = x\) and \(E(x): R \to R\) is the identity function. To tie this in with ring theory, we have
Proposition 2.1. (i) For all \( r \in R \), the function \( \text{ev}_r : R[x] \to R \) is a homomorphism.

(ii) The function \( E \) is a ring homomorphism from \( R[x] \) to \( R^R \), the ring of all functions from \( R \) to itself (with the operations of pointwise addition and multiplication).

Proof. (i) We must check that, for all \( f, g \in R[x] \),
\[
\text{ev}_r(f + g) = \text{ev}_r(f) + \text{ev}_r(g); \quad \text{ev}_r(fg) = \text{ev}_r(f) \text{ev}_r(g).
\]

With \( f = \sum_i a_i x^i \) and \( g = \sum_i b_i x^i \),
\[
\text{ev}_r(f) + \text{ev}_r(g) = \sum_i a_i r^i + \sum_i b_i r^i = \sum_i (a_i + b_i) r^i = \text{ev}_r(f + g).
\]

Here of course we can add as many terms of the form \( 0x^k \) as are needed to make sure that the sums for \( f \) and \( g \) have the same limits.

For multiplication, with \( f \) and \( g \) as above,
\[
\text{ev}_r(f) \text{ev}_r(g) = \left( \sum_i a_i r^i \right) \left( \sum_i b_i r^i \right) = \sum_{i,j} a_i b_j r^{i+j} = \sum_k \left( \sum_{i+j=k} a_i b_j \right) r^k = \text{ev}_r(fg).
\]

Finally, \( \text{ev}_r(1) = 1 \), so \( \text{ev}_r \) takes the unity in \( R[x] \) to the unity in \( R \).

(ii) We must check that, for all \( f, g \in R[x] \),
\[
E(f + g) = E(f) + E(g); \quad E(fg) = E(f)E(g).
\]

To check for example that the functions \( E(f + g) \) and \( E(f) + E(g) \) are equal, we must check that they have the same value at every \( r \in \), i.e. that
\[
E(f + g)(r) = (E(f) + E(g))(r) = E(f)(r) + E(g)(r)
\]
for every \( r \in R \), where the second equality is just the definition of pointwise addition of functions. By definition, \( E(f + g)(r) = \text{ev}_r(f + g) = \text{ev}_r(f) + \text{ev}_r(g) \), by Part (i), and so
\[
E(f + g)(r) = \text{ev}_r(f) + \text{ev}_r(g) = E(f)(r) + E(g)(r)
\]
as claimed. Finally we must check that \( E(1) = 1 \), where the right hand 1 is the unity in \( R^R \). Here \( E(1)(r) = \text{ev}_r(1) = 1 \) for all \( r \), and hence \( E(1) \) is the constant function \( f : R \to R \) whose value at every \( r \in R \) is 1. This is the unity in \( R^R \). \( \square \)
In more down to earth terms, Part (ii) above just says that every polynomial in \( R[x] \) defines a function from \( R \) to \( R \), and that the operations of polynomial addition and multiplication correspond to pointwise addition and multiplication respectively (and that the constant polynomial 1 corresponds to the constant function 1). One reason (among many) that we want to be somewhat pedantic about this setup is the following observation: For \( R = \mathbb{R} \), the homomorphism \( E \): \( \mathbb{R}[x] \rightarrow \mathbb{R}^{\mathbb{R}} \) is injective: this just says that a polynomial function determines the polynomial itself (i.e. its coefficients) uniquely. We will give an algebraic argument for this fact, in much more generality, soon. Of course, the homomorphism \( E \): \( \mathbb{R}[x] \rightarrow \mathbb{R}^{\mathbb{R}} \) is definitely not surjective, since most functions from \( \mathbb{R} \) to \( \mathbb{R} \) are not polynomials.

There are various generalizations of the homomorphism \( ev_r \):

1. In the case of the polynomial ring \( R[x_1, \ldots, x_n] \) in \( n \) variables, given \( r_1, \ldots, r_n \in R \), we can evaluate \( f \in R[x_1, \ldots, x_n] \) at \((r_1, \ldots, r_n)\). This gives a homomorphism \( ev_{r_1, \ldots, r_n} : R[x_1, \ldots, x_n] \rightarrow R \), as well as a homomorphism \( E : R[x_1, \ldots, x_n] \rightarrow R^{\mathbb{R}^n} \). In other words, a polynomial in \( n \) variables defines a function “of \( n \) variables,” i.e. a function \( R^n \rightarrow R \). Note that \( ev_{r_1, \ldots, r_n} \) can be defined inductively: viewing \( R[x_1, \ldots, x_n] \) as \( R[x_1, \ldots, x_{n-1}][x_n] \) and \( r_n \in R \leq R[x_1, \ldots, x_{n-1}] \), \( ev_{r_n} \) is a homomorphism

\[
ev_{r_n} : R[x_1, \ldots, x_{n-1}][x_n] \rightarrow R[x_1, \ldots, x_{n-1}],
\]

and by repeating this construction successively we get

\[
ev_{r_1, \ldots, r_n} = ev_{r_1} \circ \cdots \circ ev_{r_n} : R[x_1, \ldots, x_n] \rightarrow R.
\]

2. Suppose that \( R \) is a subring of a ring \( S \) and that \( s \in S \). Then we can restrict \( ev_s \) to the subring \( R[x] \) of \( S[x] \) to define a homomorphism \( ev_s : R[x] \rightarrow S \). For example, we might want to evaluate a polynomial with \textit{real} coefficients on a complex number such as \( i \). As we have seen, the image of \( ev_s \) is a subring of \( S \), and is denoted \( R[s] \). By definition, since \( ev_s(a) = a \) for all \( a \in R \) and \( ev_s(x) = s \), the subring \( R[s] \) of \( S \) contains \( R \) and \( s \). In fact,

\[
R[s] = \left\{ \sum_i a_i s^i : a_i \in R \right\}.
\]

Clearly, every subring of \( S \) containing \( R \) and \( s \) contains \( s^i \) for all nonnegative integers \( i \), hence contains \( a_i s^i \) for all \( a_i \in R \) and thus
contains $R[s]$. Thus: $R[s]$ is the smallest subring of $S$ containing $R$ and $s$. For example, the rings $\mathbb{Z}[i], \mathbb{Z}[\sqrt{2}]$ are of this type. Of course, since $i^2 = -1$, given $a_n \in \mathbb{Z}$, we can rewrite $\sum_n a_n i^n$ as a sum only involving actual integers ($n$ even) as well as integers times $i$ ($n$ odd), so every expression of the form $\sum_n a_n i^n$ is actually of the form $a + bi$ where $a, b \in \mathbb{Z}$. A similar remark holds for $\mathbb{Z}[\sqrt{2}]$, using the fact that $(\sqrt{2})^n$ is always of the form $a, b\sqrt{2}$, or $c(\sqrt{2})^2$ for integers $a, b, c$ depending on whether $n$ is congruent to 0, 1, or 2 mod 3.

More generally, given $s_1, \ldots, s_n \in S$, we can define $\text{ev}_{s_1, \ldots, s_n} : R[x_1, \ldots, x_n] \to S$. The image of $\text{ev}_{s_1, \ldots, s_n}$ is a subring of $S$, denoted by $R[s_1, \ldots, s_n]$, and it is the smallest subring of $S$ containing $R$ and $s_1, \ldots, s_n$.

3. Suppose that $\varphi : R \to S$ is a homomorphism. Then we can define a homomorphism from $R[x]$ to $S[x]$, which for simplicity we also denote by $\varphi$, by “applying $\varphi$ to all of the coefficients of $f$.” Explicitly, if $f = \sum_i a_i x^i$, we set $\varphi(f) = \sum_i \varphi(a_i) x^i$. It is easy to check from the definition of polynomial multiplication and the fact that $\varphi$ preserves addition and multiplication that $\varphi : R[x] \to S[x]$ is also a ring homomorphism. We have tacitly used one example of this already: if $R$ is a subring of $S$, then $R[x]$ is a subring of $S[x]$. For another important example, let $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ be the projection of an integer to its congruence class mod $n$. Then we get a homomorphism $\pi : \mathbb{Z}[x] \to (\mathbb{Z}/n\mathbb{Z})[x]$, which consists in reducing the coefficients of an integer polynomial mod $n$.

4. We can also amalgamate the examples above: given a $\varphi : R \to S$ and an element $s \in S$, we can define

$$\text{ev}_{\varphi, s} = \text{ev}_s \circ \varphi.$$ 

In other words, given the polynomial $f \in R[x]$, first apply the homomorphism $\varphi$ to the coefficients of $f$ to view it as a polynomial in $S[x]$, then evaluate it at $s$.

One general theme of this course is as follows: let $F$ be a field (typically $F$ is $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{F}_p$) and let $f \in F[x]$. Then we want to find a root or zero of $f$ (sometimes we say we want to “solve the equation $f = 0$). This means we want to find an element $r \in F$ such that $\text{ev}_r(f) = f(r) = 0$. By experience,
such as with the polynomial $x^2 + 1 \in \mathbb{R}[x]$ or $x^2 - 2 \in \mathbb{Q}[x]$, sometimes we cannot find such an $r$ within $F$. In this case, we look for a larger field $E$, i.e. a field containing $F$ as a subfield, and an element $s \in E$ such that $ev_s(f) = 0$. 

7