2.4 Algebra of polynomials

([1], p.136-142)

In this section we will give a brief introduction to the algebraic properties of the polynomial algebra $\mathbb{C}[t]$. In particular, we will see that $\mathbb{C}[t]$ admits many similarities to the algebraic properties of the set of integers $\mathbb{Z}$.

Remark 2.4.1. Let us first recall some of the algebraic properties of the set of integers $\mathbb{Z}$.

- **division algorithm**: given two integers $w, z \in \mathbb{Z}$, with $|w| \leq |z|$, there exist $a, r \in \mathbb{Z}$, with $0 \leq r < |w|$ such that
$$z = aw + r.$$
Moreover, the 'long division' process allows us to determine a, r. Here r is the 'remainder'.

- **prime factorisation**: for any $z \in \mathbb{Z}$ we can write
$$z = \pm p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s},$$
where p_i are prime numbers. Moreover, this expression is essentially unique - it is unique up to ordering of the primes appearing.

- **Euclidean algorithm**: given integers $w, z \in \mathbb{Z}$ there exists $a, b \in \mathbb{Z}$ such that
$$aw + bz = \gcd(w, z),$$
where $\gcd(w, z)$ is the 'greatest common divisor' of w and z. In particular, if w, z share no common prime factors then we can write
$$aw + bz = 1.$$
The Euclidean algorithm is a process by which we can determine a, b.

We will now introduce the polynomial algebra in one variable. This is simply the set of all polynomials with complex coefficients and where we make explicit the $\mathbb{C}$-vector space structure and the multiplicative structure that this set naturally exhibits.

Definition 2.4.2. - The $\mathbb{C}$-*algebra of polynomials in one variable*, is the quadruple $(\mathbb{C}[t], \alpha, \sigma, \mu)$[43] where $(\mathbb{C}[t], \alpha, \sigma)$ is the $\mathbb{C}$-vector space of polynomials in t with $\mathbb{C}$-coefficients defined in Example 1.2.6, and
$$\mu : \mathbb{C}[t] \times \mathbb{C}[t] \to \mathbb{C}[t] \; ; \; (f, g) \mapsto \mu(f, g),$$
is the 'multiplication' function.

So, if
$$f = a_0 + a_1 t + ... + a_n t^n, \; g = b_0 + b_1 + ... + b_m t^m \in \mathbb{C}[t],$$
with $m \leq n$ say, then
$$\mu(f, g) = c_0 + c_1 t + ... + c_{m+n} t^{m+n},$$
where
$$c_i = \sum_{\substack{j+k=i,\\ 0 \leq j \leq n,\\ 0 \leq k \leq m}} a_j b_k.$$

[43]This is a particular example of a more general algebraic object called a $\mathbb{C}$-*algebra*: a $\mathbb{C}$-algebra is a set A that is a $\mathbb{C}$-vector space and for which there is a well-defined commutative multiplication map that interacts with addition in a nice way - for example, distributivity, associativity hold. One usually also requires that a $\mathbb{C}$-algebra A has a *multiplicative identity*, namely an element e such that $f \cdot e = e \cdot f = f$, for every $f \in A$. It is common to denote this element by 1.

We write

$$\mu(f,g) = f \cdot g, \quad \text{or simply } fg.$$

μ is nothing more than the function defining the 'usual' multiplication of polynomials with $\mathbb{C}$-coefficients. In particular, for every $f, g \in \mathbb{C}[t]$ we have $fg = gf$.

We will write $\mathbb{C}[t]$ instead of the quadruple above when discussing $\mathbb{C}[t]$ as a $\mathbb{C}$-algebra. Note that the polynomial $1 \in \mathbb{C}[t]$ satisfies the property that $1 \cdot f = f \cdot 1 = f$, for every $f \in \mathbb{C}[t]$.

- A *representation of* $\mathbb{C}[t]$ is a $\mathbb{C}$-linear morphism

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

for some finite dimensional $\mathbb{C}$-vector space V, such that

$$(*) \quad \rho(fg) = \rho(f) \circ \rho(g), \quad \text{and} \quad \rho(1) = \mathrm{id}_V,$$

where we are considering composition of linear endomorphisms of V on the RHS of the first equality.[44]

Remark 2.4.3. Suppose that

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

is a representation of $\mathbb{C}[t]$. Then, for any $f = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n \in \mathbb{C}[t]$, we have

$$\begin{aligned} \rho(f) = \rho(a_0 + a_1 t + ... + a_n t^n) &= a_0\rho(1) + a_1\rho(t) + ... + a_n\rho(t^n), \quad \text{as } \rho \text{ is } \mathbb{C}\text{-linear}, \\ &= a_0 \mathrm{id}_V + a_1\rho(t) + a_2\rho(t)^2 + ... + a_n\rho(t)^n, \quad \text{by } (*), \end{aligned}$$

where we have written $\rho(t)^k = \rho(t) \circ \cdots \circ \rho(t)$, the k-fold composition of $\rho(t)$.

Hence, a represention of $\mathbb{C}[t]$ is the same thing as specifying a $\mathbb{C}$-linear endomorphism $\rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$: the multiplicative property of ρ then implies that $\rho(f)$ only depends on $\rho(t)$, for any $f \in \mathbb{C}[t]$.

Conversely, given a $\mathbb{C}$-linear endomorphism of V, $L \in \mathrm{End}_{\mathbb{C}}(V)$ say, then we can define a representation ρ_L of $\mathbb{C}[t]$ as follows: define

$$\rho_L : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V) \; ; \; a_0 + a_1 t + ... + a_n t^n \mapsto a_0 \mathrm{id}_v + a_1 L + ... + a_n L^n \in \mathrm{End}_{\mathbb{C}}(V),$$

where $L^k = L \circ \cdots \circ L$ and the addition and scalar multiplication on the RHS is occuring in $\mathrm{End}_{\mathbb{C}}(V)$.

We are going to study an endomorphism $L \in \mathrm{End}_{\mathbb{C}}(V)$ by studying the representation ρ_L of $\mathbb{C}[t]$ it defines. If $A \in Mat_n(\mathbb{C})$ then we define ρ_A to be the representation defined by the endomorphism T_A of $\mathbb{C}^n$.

Suppose we are given a representation of $\mathbb{C}[t]$

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

and denote $n = \dim_{\mathbb{C}} V$, $L = \rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$ (so that $\rho = \rho_L$) and suppose that $L \neq \mathrm{id}_V$.[45]

We know that $\mathrm{End}_{\mathbb{C}}(V)$ is n^2-dimensional (since we know that $\mathrm{End}_{\mathbb{C}}(V)$ is isomorphic to $Mat_n(\mathbb{C})$). Therefore, there must exist a nontrivial linear relation

$$\lambda_0 \mathrm{id}_V + \lambda_1 L + \lambda_2 L^2 + ... + \lambda_{n^2} L^{n^2} = 0_{\mathrm{End}_{\mathbb{C}}(V)},$$

[44] This means that ρ is a morphism of (unital) $\mathbb{C}$-algebras.

[45] If $L = \mathrm{id}_V$ then we call the representation ρ_{id_V} the *trivial representation*. In this case, we have that

$$\mathrm{im}\rho = \{c \cdot \mathrm{id}_V \in \mathrm{End}_{\mathbb{C}}(V) \mid c \in \mathbb{C}\} \subset \mathrm{End}_{\mathbb{C}}(V).$$

with $\lambda_i \in \mathbb{C}$, since the set $\{\mathrm{id}_V, L, L^2, \dots, L^{n^2}\}$ contains n^2+1 vectors. Thus, we have

$$\begin{aligned} 0_{\mathrm{End}_{\mathbb{C}}(V)} &= \lambda_0 \mathrm{id}_V + \lambda_1 L + \lambda_2 L^2 + \dots + \lambda_{n^2} L^{n^2} \\ &= \lambda_0 \rho(1) + \lambda_1 \rho(t) + \dots + \lambda_{n^2} \rho(t)^{n^2} \\ &= \rho(\lambda_0 + \lambda_1 t + \dots + \lambda_{n^2} t^{n^2}), \end{aligned}$$

so that the polynomial

$$f = \lambda_0 + \lambda_1 t + \dots + \lambda_{n^2} t^{n^2} \in \ker \rho.$$

In particular, we have that $\ker \rho \neq \{0_{\mathbb{C}[t]}\}$. We will now make a detailed study of the kernel of representations of $\mathbb{C}[t]$.

Keep the same notation as above. We have just seen that $\ker \rho$ is nonzero. Let $m_L \in \ker \rho$ be a nonzero polynomial for which $\rho(m_L) = 0_{\mathrm{End}_{\mathbb{C}}(V)}$ and such that m_L has minimal degree.[46] We must have $\deg m_L \neq 0$, otherwise m_L is a constant polynomial, say $m_L = c \cdot 1$ with $c \in \mathbb{C}$, $c \neq 0$, and $\rho(c \cdot 1) = c\rho(1) = c\mathrm{id}_V \neq 0_{\mathrm{End}_{\mathbb{C}}(V)}$, contradicting that $m_L \in \ker \rho$. Hence, we can assume that $\deg m_L = m > 0$.

Now, let $f \in \ker \rho$ be any other polynomial in the kernel of ρ. Denote $\deg f = p$. Thus, by our choice of m_L (it must have minimal degree) we see that $p \geq m$. Now use the division algorithm for polynomials[47] to find polynomials $g, h \in \mathbb{C}[t]$ such that

$$f = gm_L + h,$$

where $\deg h < m$.

Then, as $f \in \ker \rho$, we must have

$$0_{\mathrm{End}_{\mathbb{C}}(V)} = \rho(f) = \rho(gm_L + h) = \rho(g)\rho(m_L) + \rho(h) = 0_{\mathrm{End}_{\mathbb{C}}(V)} + \rho(h),$$

so that $h \in \ker \rho$. If h were a nonzero polynomial then we have obtained an element in $\ker \rho$ that has strictly smaller degree that m_L, contradicting our choice of m_L. Hence, we must have that $h = 0$ and $f = gm_L$. We say that m_L *divides* f.

We have just shown the following

Proposition 2.4.4. *Suppose that*

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

is a representation of $\mathbb{C}[t]$. *Denote* $L = \rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$ *and suppose that* $m_L \in \ker \rho$ *is nonzero and has minimal degree. Then, for any* $f \in \ker \rho$ *there exists* $g \in \mathbb{C}[t]$ *such that*

$$f = gm_L.$$

Remark 2.4.5. Proposition 2.4.4 is stating the fact that the $\mathbb{C}$-algebra $\mathbb{C}[t]$ is a *principal ideal domain*, namely, every ideal in $\mathbb{C}[t]$ is generated by a single polynomial (ie, 'principal').

Definition 2.4.6. Let $L \in \mathrm{End}_{\mathbb{C}}(V)$ and consider the representation

$$\rho_L : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

defined above. We define the *minimal polynomial of L*, denoted $\mu_L \in \mathbb{C}[t]$, to be the unique nonzero polynomial $\mu_L \in \ker \rho$ that has minimal degree and has leading coefficient 1: this means that

$$\mu_L = a_0 + a_1 t + \dots + a_{m-1} t^{m-1} + t^m.$$

[46] Recall that the degree, $\deg f$, of a polynomial

$$f = a_0 + a_1 t + \dots + a_k t^k \in \mathbb{C}[t],$$

is defined to be $\deg f = k$. We have the property that

$$\deg fg = \deg f + \deg g.$$

[47] If you have not seen this before, don't worry, as I will cover this in class.

This polynomial is well-defined (ie, it's unique) by Proposition 2.4.4: if $m_L \in \ker \rho$ has minimal degree and leading coefficient $a \in \mathbb{C}$, then we have $\mu_L = a^{-1} m_L$. If $f \in \ker \rho$ is any other polynomial of minimal degree and with leading coefficient 1, then we must have $\deg f = \deg \mu_L$ and, by Proposition 2.4.4, we know that there exists $g \in \mathbb{C}[t]$ such that

$$f = g\mu_L.$$

Since $\deg f = \deg(g\mu_L) = \deg g + \deg \mu_L$ we must have that $\deg g = 0$, so that $g = c \cdot 1 \in \mathbb{C}[t]$. As both f and μ_L have leading coefficient 1, the only way this can hold is if $c = 1$, so that $f = \mu_L$.

For $A \in Mat_n(\mathbb{C})$ we write μ_A instead of μ_{T_A} and call it the *minimal polynomial of A*.

Corollary 2.4.7. *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$, μ_L be the minimal polynomial of L. For $f = a_0 + a_1 t + ... + a_k t^k \in \mathbb{C}[t]$ we denote*

$$f(L) = \rho_L(f) = a_0 \mathrm{id}_V + a_1 L + ... + a_k L^k \in \mathrm{End}_{\mathbb{C}}(V).$$

If $f(L) = 0_{\mathrm{End}_{\mathbb{C}}(V)}$ then $f = \mu_L g$, for some $g \in \mathbb{C}[t]$.

Proof: This is simply a restatement of Proposition 2.4.4. □

Example 2.4.8. 1. Consider the endomorphism T_A of $\mathbb{C}^3$ defined by the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & -1 \end{bmatrix}.$$

Then, you can check that the following relation holds

$$-A^3 + 2A^2 - A + 2I_3 = 0_3.$$

Consider the representation ρ_A defined by A. Then, since the above relation holds we must have

$$f = -\lambda^3 + 2\lambda^2 - \lambda + 2 \in \ker \rho_A.$$

You can check that we can decompose f as

$$f = (2 - \lambda)(\lambda - \sqrt{-1})(\lambda + \sqrt{-1}).$$

Hence, we must have that μ_A is one of the following polynomials[48]

$$(\lambda - \sqrt{-1})(\lambda + \sqrt{-1}),\ (2 - \lambda)(\lambda - \sqrt{-1}),\ (2 - \lambda)(\lambda + \sqrt{-1}),\ f.$$

In fact, we have $\mu_A = f$.

You may have noticed that $f = \chi_A(\lambda)$ - this is the *Cayley-Hamilton Theorem* (to be proved later and in homework): if $A \in Mat_n(\mathbb{C})$ then $\chi_A(\lambda) \in \ker \rho_A$, so that $\chi_A(A) = 0$ (using the above notation from Corollary 2.4.7).

- Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

You can check that we have the relation

$$-A^3 + 3A^2 - 3A + I_3 = 0_3,$$

so that

$$f = -\lambda^3 + 3\lambda^2 - 3\lambda + 1 = (1 - \lambda)^3 \in \ker \rho_A.$$

[48] Why can't we have μ_A be one of $(2 - \lambda)$, $(\lambda - \sqrt{-1})$, $(\lambda + \sqrt{-1})$?

Now, we see that we must have μ_A being one of the following polynomials[49]

$$(1-\lambda)^2,\ f.$$

It can be checked that

$$A^2 - 2A + I_3 = 0_3,$$

so that

$$\mu_A = (1-\lambda)^2.$$

You will notice that

$$\chi_A(\lambda) = (1-\lambda)^3.$$

In both of these examples you can see that the roots of the minimal poynomial of A are precisely the eigenvalues of A (possibly with some repeated multiplicity). In fact, this is always true: for a matrix A the roots of μ_A are precisely the eigenvalues of A. This will be proved in the next section.

Recall that a polynomial $f \in \mathbb{C}[t]$ can be written as a product of linear factors

$$f = a(t-c_1)^{n_1}(t-c_2)^{n_2}\cdots(t-c_k)^{n_k},$$

where $a, c_1, \dots, c_k \in \mathbb{C}$, $n_1, \dots, n_k \in \mathbb{N}$.

This is the analogue in $\mathbb{C}[t]$ of the 'prime factorisation' property of $\mathbb{Z}$ mentioned at the beginning of this section: the 'primes' of $\mathbb{C}[t]$ are degree 1 polynomials.

Definition 2.4.9. We say that the (nonzero) polynomials $f_1, \dots, f_p \in \mathbb{C}[t]$ are *relatively prime* if there is no common linear factor for all of the f_i.

Example 2.4.10. The polynomials $f = t^2+1$ and $g = t^2-1$ are relatively prime. Indeed, we have

$$f = t^2+1 = (t-\sqrt{-1})(t+\sqrt{-1}),\ g = (t-1)(t+1),$$

so that there is no common linear factor of either.

However, the polynomials g and $h = t^n - 1$ are <u>not</u> relatively prime as

$$h = t^n - 1 = (t-1)(t-\omega)(t-\omega^2)\cdots(t-\omega^{n-1}),$$

where $\omega = \cos(2\pi/n) + \sqrt{-1}\sin(2\pi/n) \in \mathbb{C}$. Hence, the linear factor $(t-1)$ appears in both g and h.

We now give another basic algebraic property of the $\mathbb{C}$-algebra $\mathbb{C}[t]$ whose proof you would usually see in Math 113. As such, we will not prove this result here although the proof is exactly the same as the corresponding result for $\mathbb{Z}$ (with the appropriate modifications): it involves the $\mathbb{C}[t]$-analogue of the 'Euclidean algorithm' for $\mathbb{Z}$.

Lemma 2.4.11. *Let $f_1, \dots, f_p \in \mathbb{C}[t]$ be a collection of relatively prime polynomials. Then, there exists $g_1, \dots, g_p \in \mathbb{C}[t]$ such that*

$$f_1g_1 + \dots + f_pg_p = 1 \in \mathbb{C}[t].$$

Example 2.4.12.

1. The polynomials $f = t^2+1, g = t^2-1$ are relatively prime and

$$\frac{1}{2}(t^2+1) - \frac{1}{2}(t^2-1) = 1.$$

2. The polynomials $f = t^2+1, g = t^3-1$ are relatively prime and

$$\frac{1}{2}(t-1)(t^3-1) - \frac{1}{2}(t^2-t-1)(t^2+1) = 1.$$

[49] Why can't we have $1-\lambda$?