Polynomials and Rings

We often consider polynomials in matrices or linear maps. We write

$$\mathbb{F}[x] = \left\{ \sum_{k=0}^{n} a_k x^k \,|\, a_i \in \mathbb{F},\, n \in \mathbb{N} \right\}$$

as the set of polynomials over $\mathbb{F}$.

Definition 1 (Field) *A field $(F, +, \times)$ satisfies*
- *$(F, +)$ is an abelian group.*
- *$(F^*, \times)$ is an abelian group.*
- *$a(b+c) = ab + ac$ and $(a+b)c = ac + bc$.*

Definition 2 (Ring) *A ring $(R, +, \times)$ satisfies*
- *$(R, +)$ is an Abelian group.*
- *$\times$ is associative.*
- *$a(b+c) = ab + ac$ and $(a+b)c = ac + bc$.*

Any field is a commutative ring. A particular class of rings of interest is the integral domains, which are the nonzero commutative rings for which $ab = 0$ implies that $a = 0$ or $b = 0$. We generally consider mainly commutative rings that have an identity element (and therefore only lack multiplicative inverses).

We say that a field is algebraically closed if every non-constant $p \in \mathbb{F}[x]$ has a root in $\mathbb{F}$. We can always embed a field $\mathbb{F}$ into a minimal algebraically closed field $\overline{\mathbb{F}}$, for example we can embed $\mathbb{R}$ in $\mathbb{C}$.

Definition 3 *A map $\phi : R \to S$ between two rings is a ring homomorphism if for all $r, r' \in R$*

$$\phi(r + r') = \phi(r) + \phi(r')$$
$$\text{and } \phi(rr') = \phi(r)\phi(r').$$

If this is bijective we call it a ring isomorphism.

Definition 4 *A non-empty subset I of a ring R is an ideal if for all $s, t \in I$ and $r \in R$ we have $s - t \in I$ and $sr, rs \in I$.*

We write the above fact as $I \trianglelefteq R$, and as consequence $R/I = \{r + I \,|\, r \in R\}$ is a commutative ring, which just follows by showing that multiplication and addition are well defined and satisfy the ring properties. From following basic pattern matching over from group theory we then get the following isomorphism theorem:

Theorem 1 (Isomorphism theorem) *For any ring homomorphism $\phi : R \to S$ we have*

$$R/\ker\phi \cong \operatorname{im}\phi$$
$$\operatorname{im}\phi \leq S$$
$$\ker\phi \trianglelefteq R$$

via $r + \ker\phi \mapsto \phi(r)$.

We write $\langle a \rangle$ as the principal ideal in R containing a. That is, it is the set of all ar or ra for $r \in \mathbb{R}$.

Polynomials

Any field $\mathbb{F}$ induces the commutative ring $\mathbb{F}[x]$, which is the set of polynomials with coefficients in $\mathbb{F}$. Potentially surprisingly, it turns out that we can deal with most elements of $\mathbb{F}[x]$ in very similar ways to $\mathbb{Z}$. Writing $f(x) \,|\, g(x)$ where $\frac{g(x)}{f(x)} \in \mathbb{F}[x]$, we get the division algorithm:

Theorem 2 *For any $f, g \in \mathbb{F}[x]$ with $g(x) \neq 0$, there exist $q, r \in \mathbb{F}$ with $\deg r < \deg g$ such that*

$$f(x) = q(x)g(x) + r(x)$$

We can prove this by induction on the degree of $f(x)$ until $\deg f < \deg g$. From this statement we then quickly get Bezout's lemma.

For an arbitrary matrix A, we write $m_A(x)$ as the least degree polynomial with $m_A(A) = 0$. This polynomial must exist, as we must eventually find that a linearly dependent A^n is introduced to the sequence I, A, A^2, ..., meaning by definition we can have $m_A(A) = 0$. Indeed this also shows that the smallest degree occurs for the first A^n linearly dependent on $\{I, \cdots, A^{n-1}\}$. Note that for any polynomial f, $f(P^{-1}AP) = P^{-1}f(A)P$, so $m_{P^{-1}AP}(A) = 0$. Further as each minimum polynomial is monic (has the same highest coefficient), thus the minimum polynomial of A is unique so $m_{P^{-1}AP} = m_A$.

Theorem 3 *Where for A, a polynomial f is annihilating if $f(A) = 0$. If f is annihilating, then $m_A(x) \,|\, f(x)$.*

$f(A) = q(A)m_A(A) + r(A)$ with $\deg r < \deg m_A$, so $r(A) = 0$, leaving the only possibility that $r(x) = 0$. An alternative way of looking at this is to say that with fixed A, the map $f(x) \mapsto f(A)$ is a ring homomorphism with kernel the set of annihilating polynomials – this set is an ideal, and $\mathbb{F}[x]$ is a principal ideal domain, so thus each annihilating polynomial is generated by a single one, m_A.

For v an eigenvector of A, note that we get $f(A)v = \sum_{k=0}^{n} a_k \lambda^k v = f(\lambda)v$. Thus every root of $\chi_A(x)$ (each eigenvalue) must be a root of $m_A(x)$. Further, for λ a root of $m_A(x)$, $m_A(x) = (x - \lambda)g(x)$, and $g(A) \neq 0$ so there is v such that $0 = m_A(A)v = (A - \lambda I)g(A)v$ but $g(A)v \neq 0$, so λ is an eigenvalue and thus a root of $\chi_A(x)$.

Theorem 4 (Cayley-Hamilton) *If $T : V \to V$ is a linear transformation and V is a finite dimensional vector space, then $\chi_T(T) = 0$. In particular, $m_T(x) \,|\, \chi_T(x)$.*

As proof just take the upper triangular matrix A representing T with each eigenvalue occurring as many times in the diagonal as they occur in χ_T (if $\mathbb{F}$ is not closed just use $\overline{\mathbb{F}}$, and we will still have conjugacy), and use $\prod_{i=1}^{n}(A - \lambda_i I) = 0$ for such a matrix.

Quotient Spaces

Definition 5 (Coset) *A coset of a subspace U of a vector space V is a set $v + U = \{v + u \,|\, u \in U\}$ for some $v \in V$.*

Definition 6 (Quotient space) *The quotient space V/U is the set of cosets of U in V.*

We carry over from group theory the key fact that $v_1 + U = v_2 + U$ iff $v_1 - v_2 \in U$. We should also get relatively immediately that V/U is a vector space with operations defined as expected. Further, we can also get that if $\{u_1, \cdots, u_k\}$ is a basis for U, extending this to a basis $\{u_1, \cdots, u_k, v_{k+1}, \cdots, v_n\}$, $\{v_{k+1}, \cdots, v_n\}/U$ is a basis for V/U. We can then also see conversely that if B_1 is a basis for U and B_2/U a basis for V/U, then $B_1 \cup B_2$ is a basis for V.

Theorem 5 (Isomorphism theorem for vector spaces) *Let $T : V \to W$ be linear, then $V/\ker T \cong \operatorname{im} T$ via $v + \ker T \mapsto \operatorname{im} T$. This also serves as a proof for rank-nullity where V and $\operatorname{im} T$ are finite.*

Theorem 6 *A linear map $T : V \to V$ induces a linear map $\overline{T} : V/U \to V/U$ by $v + U \mapsto Tv + U$ iff T is U-invariant.*

This is clear from the definition, and we get from this that where we have a transformation U-invariant for some U, extending a basis for U to a basis for V gives a matrix of the form

$$\begin{bmatrix} A & * \\ \mathbf{0} & B \end{bmatrix}$$

where A is a representation of $T|_U$ and B is a representation of $\overline{T}$.

Theorem 7 (Triangular form) *Let $T : V \to V$. Suppose $\chi_T(x)$ is the product of linear (possibly repeated) factors. Then there is a basis for which T is upper triangular.*

Note that this holds immediately in any algebraically closed field, and so we can use this for any linear transformation T by passing from $\mathbb{F}$ to its algebraic closure $\overline{\mathbb{F}}$.

To prove this, note that T is E_λ invariant for any eigenvalue λ. From here we just show that by taking each $(x - \lambda)$ out of χ_T one by one we can form a matrix where each eigenvalue occurs in equal number to its geometric multiplicity.

Decomposition

In order to find more informative ways to conjugate matrices, we consider how, where $T : V \to V$ for V the direct sum of T-invariant subspaces $W_i = \langle \mathcal{B}_i \rangle$, we can get a matrix for T:

$${}_{\mathcal{B}}T_{\mathcal{B}} = \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_r \end{bmatrix}$$

where $A_i = {}_{\mathcal{B}_i}\left[T|_{W_i}\right]_{\mathcal{B}_i}$.

Lemma 8 *Assume $f(x) = a(x)b(x)$ with $\gcd(a, b) = 1$ and $f(T) = 0$. Then*

$$V = \ker(a(T)) \oplus \ker(b(T))$$

is a T-invariant decomposition. Furthermore, if $f = m_T$ and a and b are monic, then a and b are the minimal polynomials of T restricted to $\ker(a(T))$ and $\ker(b(T))$ respectively.

We have $p, q \in \mathbb{F}[x]$ such that $ap + bq = 1$, so $a(T)p(T) + b(T)q(T)$ is the identity. Thus we immediately decompose V into $\operatorname{im}(a(T)p(T))$ and $\operatorname{im}(b(T)q(T))$, and see that $\operatorname{im}(a(T)p(T)) \subseteq \ker b(T)$ and $\operatorname{im}(b(T)q(T)) \subseteq \ker a(T)$, and additionally that if $a(T)v = b(T)v = 0$ then $v = 0$ so it is a direct sum (giving that the images are actually equal to the kernels). It is T-invariant as $T(a(T)p(T)v) = a(T)p(T)(Tv)$ and $T(b(T)q(T)v) = b(T)q(T)(Tv)$. The last claim follows immediately with similar approaches as to the others.

Theorem 9 (Primary Decomposition) *Let m_T be the minimal polynomial and write it in the form*

$$m_T(x) = \prod_{i=1}^{r} f_i^{q_i}(x)$$

where f_i are distinct monic irreducible polynomials. With $W_i = \ker(f_i^{q_i}(T))$ Then

$$V = \bigoplus_{i=1}^{r} W_i$$
$$T[W_i] \subseteq W_i$$
$$m_{T|_{W_i}} = f_i^{q_i}.$$

These all follow immediately from the above lemma. Consequently we have that every T has a primary decomposition into r matrices, and where $\mathbb{F}$ is closed r is the number of eigenvalues. We then desire to move forward by finding out the specific form of the matrices ${}_{\mathcal{B}_i}[T|_{W_i}]_{\mathcal{B}_i}$. In fact, we get from later results that for closed $\mathbb{F}$, these are all of the form $J_i(\lambda_j) = \lambda_j I_i + J_i$.

Theorem 10 (Nilpotency) *If T is nilpotent, so there is some $n > 0$ such that $T^n = 0$, then $m_T(x) = x^m$ for some $m \leq n$, and there is a basis $\mathcal{B}$ such that*

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{bmatrix} 0 & * & & 0 \\ & \ddots & \ddots & \\ & & \ddots & * \\ 0 & & & 0 \end{bmatrix} \qquad \text{with each } * \in \{0, 1\}$$

The proof of this theorem follows from observing the strictly increasing (for $\{0, \ldots, m\}$) sequence $\ker(T^k)$. Take $\mathcal{B}_i$ such that $\mathcal{B}_i/\ker(T^{i-1})$ is a basis for $\ker(T^i)/\ker(T^{i-1})$, and collectively for $i \in 1, \ldots, m$ this forms a basis for V. Additionally, note that $T(\mathcal{B}_{i+1})/\ker(T^{i-1})$ is linearly independent in $\ker(T^i)/\ker(T^{i-1})$, so $|\mathcal{B}_{i+1}| \leq |\mathcal{B}_i|$.

To form a basis we begin with an arbitrary $\mathcal{B}_m$, then noting the above independence observation take $T(\mathcal{B}_m)$ and extend it to form $\mathcal{B}_{m-1} = \mathcal{E}_{m-1} \cup T(\mathcal{B}_m)$. We then note that we can reorder the basis as

$$\begin{aligned} \mathcal{B} &= \bigcup_{i=1}^{m} \mathcal{B}_i \\ &= \bigcup_{i=1}^{m} \bigcup_{j=i}^{m} T^{j-i}(\mathcal{E}_j) \\ &= \bigcup_{i=1}^{m} \left(\bigcup_{v \in \mathcal{E}_i} \{T^k v \,|\, k \in \{0, \ldots, i-1\}\} \right) \end{aligned}$$

From here we can see both that each $\{T^k v \,|\, k \in \{0, \cdots, i-1\}\}$ is T-invariant, and that the matrix representing T restricted to each set is a Jordan block J_i of size $i \times i$ (1s above the diagonal).

The most immediate consequence of this theorem is that if a transformation has $m_T(x) = (x - \lambda)^m$, $T - \lambda I$ is nilpotent, so there is a basis such that it is a block diagonal matrix with blocks $J_i(\lambda) = \lambda I + J_i$. Given the primary decomposition theorem, we thus get that (in a complete field) every matrix is conjugate to a block diagonal matrix where each block is itself a block diagonal of Jordan blocks.

As $\mathcal{B}_1 = \bigcup_{i=0}^{m-1} T^i(\mathcal{E}_{i+1})$ is a basis for $\ker T$ and $|\mathcal{B}_1| = \sum_{i=1}^{m} |\mathcal{E}_i|$, which is the number of Jordan blocks, thus the number of Jordan blocks corresponding to any eigenvalue is the dimension of its eigenspace. Additionally, we get obvious results such as that the sum of Jordan block dimensions for an eigenvalue is its geometric multiplicity, and the largest dimension of a Jordan block is its multiplicity in the minimal polynomial.

Dual Spaces

Definition 7 *For a vector space V over $\mathbb{F}$, $V' = \operatorname{Hom}(V, \mathbb{F})$ (the space of linear maps from V to $\mathbb{F}$) is its dual, and its elements are called linear functionals.*

Theorem 11 *With $\{e_1, \ldots, e_n\}$ a basis for V, a finite-dimensional vector space, define the dual of e_i' of e_i by*

$$e_i' = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Then $\{e_1', \cdots, e_n'\}$ is a basis for V'. In particular, $e_i \mapsto e_i'$ induces an isomorphism between V and V'.

To prove this just construct the isomorphism $\varphi(\sum a_i e_i)(\sum b_j e_j) = \sum a_i b_i$.

Theorem 12 *Let V be a finite dimensional vector space. Then, $V \to V''$ defined by $v \mapsto (f \mapsto f(v))$ is a natural (independent of basis) linear isomorphism.*

This is a definition chase in almost all regards.

Definition 8 (Annihilators) *For $U \subseteq V$ a subspace of V, the annihilator of U is*

$$U^0 = \{f \in V' \,|\, f[U] = \{0\}\}$$

We can very quickly get that U^0 is a subspace of V', as for $f, g \in U^0$, $u \in U$, $\lambda \in \mathbb{F}$, $(f + \lambda g)(u) = 0$, and $0 \in U^0$.

Theorem 13 *For V finite dimensional, $U \subseteq V$ a subspace,*

$$\dim U^0 = \dim V - \dim U$$

This follows by extending a basis for U to a basis for V, then taking the dual basis. If $\langle e_1, \ldots, e_m \rangle = U$, $\langle e_1, \ldots, e_n \rangle = V$, then we have $e_j' \in U^0$ for all $j \in \{m+1, \ldots, n\}$. Conversely if $f \in U^0$, $f = \sum_{i=1}^{n} a_i e_i'$, and $a_i = 0$ necessarily for $i \in \{1, \ldots, m\}$ so we get $U^0 = \langle e_{m+1}', \ldots, e_n' \rangle$.

Inner Products

We have from prelims the definition of bilinear forms: an operator $F : V \times V \to \mathbb{F}$ linear in both arguments. Previously we had the notions of symmetric and positive definite bilinear forms, as well as how they relate to orthogonality and the spectral theorem. Introduced here we firstly have the notion of a *non-degenerate* bilinear form, which is one where $F(v, w) = 0$ for all $v \in V$ implies that $w = 0$.

Also seen briefly in prelims linear algebra is the notion of sesquilinear forms, for vector spaces V over $\mathbb{C}$ where we have additive linearity, but $F(\overline{\lambda}v, w) = \lambda F(v, w) = F(v, \lambda w)$. We say that a sesquilinear form is conjugate symmetric if for all $v, w \in V$, $F(v, w) = \overline{F(w, v)}$. We say that a conjugate symmetric form F is positive definite if $F(v, v) > 0$ (note $F(v, v) = \overline{F(v, v)}$ so $F(v, v) \in \mathbb{R}$).

To recall, a set is orthonormal if $\langle w_i, w_j \rangle = 0$ for each $i \neq j$ (mutual orthogonality), and $\langle w_i, w_i \rangle = 1$ for each i. This straightforwardly gives that $\{w_1, \ldots, w_n\}$ is linearly independent if orthonormal.

For any basis $\{v_1, \ldots, v_n\}$ in an inner product space, via the induction from $w_1 = v_1$ to w_n determined by

$$w_k = v_k - \sum_{i=1}^{k-1} \frac{\langle w_i, v_k \rangle}{\langle w_i, w_i \rangle} w_i$$

we get an orthogonal basis, transformed to an orthonormal basis via normalisation. Thus every finite dimensional inner product space has an orthonormal basis.

Theorem 14 *The map $v \mapsto \langle v, \cdot \rangle$ is a natural injective map from V to V', linear in $\mathbb{R}$ and isomorphic for V finite dimensional.*

$\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$ for all $u, v, w \in V$, $\lambda \in \mathbb{R}$, so we get linearity in $\mathbb{R}$ (and conjugate linearity in $\mathbb{C}$). If $\langle u, \cdot \rangle = \langle v, \cdot \rangle$, then for all $w \in V$, $\langle u - v, w \rangle = 0$. As any inner product is non-degenerate, thus $u = v$, so we have injectivity and thus as $\dim V = \dim V'$ we get that the map is an isomorphism.

We define the orthogonal complement of a subspace $U \subseteq V$ of an inner product space as

$$U^\perp = \{v \in V \,|\, \langle U, v \rangle = \{0\}\}.$$

This is a subspace, for which $U \cap U^\perp = \{0\}$, $U \oplus U^\perp = V$ (for finite dimensional V), $(U + W)^\perp = U^\perp \cap W^\perp$, $(U \cap W)^\perp \supseteq U^\perp + W^\perp$, and $U \subseteq (U^\perp)^\perp$ (the last two both holding with equality for finite dimensional V).

The first should be clear from definitions. The second follows from writing any $v \in V$ as $v - \sum \langle v, u_i \rangle u_i + \sum \langle v, u_i \rangle u_i$ for $\{u_1, \ldots, u_k\}$ a basis of U. The third follows from $\langle U + W, v \rangle = \{0\}$ iff $\langle U, v \rangle = \{0\}$ and $\langle W, v \rangle = \{0\}$. The fourth follows from $\langle U, v_1 \rangle = \{0\}$, $\langle W, v_2 \rangle = \{0\}$ implying that $\langle U \cap W, v_1 + v_2 \rangle = \langle U \cap W, v_1 \rangle + \langle U \cap W, v_2 \rangle = \{0\}$, with the reverse implication by taking an orthonormal basis of U, extending it to an orthonormal basis of $U + W$, and then for any v orthogonal to $U \cap W$ we can write $v = v - \sum \langle v, u_i \rangle u_i + \sum \langle v, u_i \rangle u_i$, and $\langle u_i, w_j \rangle = 0$ so v is the sum of an element orthogonal to U, and an element orthogonal to W. The final statement follows as if $v \in U$, $u \in U^\perp$, $\langle v, u \rangle = 0$, so $v \in (U^\perp)^\perp$. For finite dimensional V take an orthonormal basis of U extended to an orthonormal basis of V, and clearly the basis of $(U^\perp)^\perp$ is the basis of U.

Note also that the $v \mapsto \langle v, \cdot \rangle$ isomorphism maps $U^\perp$ to U^0 isomorphically for finite vector spaces, as $\dim U^\perp = \dim U^0$.

Definition 9 *Given a linear map $T : V \to V$, $T^* : V \to V$ is its adjoint if for all $v, w \in V$,*

$$\langle v, Tw \rangle = \langle T^* v, w \rangle$$

Immediately we get that if an adjoint exists it is unique, and further for finite dimensional V we have a linear adjoint by taking $w \mapsto \langle v, Tw \rangle$, noting that there is u such that $\langle v, T\cdot \rangle = \langle u, \cdot \rangle$ (by $v \mapsto \langle v, \cdot \rangle$ being an isomorphism for finite dimensional V), and thus $T^* v = u$ is the adjoint. Its linearity pops out from the adjoint equation it satisfies.

In fact, we can get the matrix of the adjoint as the conjugate transpose of the matrix for T with respect to the same basis. This is clear just from $\langle e_i, Te_j \rangle = \langle T^* e_i, e_j \rangle = \overline{\langle e_j, T^* e_i \rangle}$, and noting that these are the entries of the matrices of T and T^* with respect to that basis.

We say that T is self-adjoint where $T = T^*$. This essentially just translates to T having a real symmetric matrix with respect to any basis, so any eigenvalue it has will be real (proof follows from algebra). Further, if $U \subseteq V$ is T-invariant, then so is $U^\perp$, as for $w \in U^\perp$, $\langle u, Tw \rangle = \langle Tu, w \rangle = 0$ by $Tu \in U$.

Theorem 15 *If $T : V \to V$ is self-adjoint and V finite dimensional, then there is an orthonormal basis of eigenvectors for T.*

For any eigenvalue we have two T-invariant subspaces ($\langle v \rangle$ and $\langle v \rangle^\perp$ for the eigenvector v), orthogonal to each other. By induction we then build an orthogonal basis of eigenvectors, which is then straightforwardly transformed to an orthonormal basis.

Definition 10 *If for a finite dimensional inner product space V, a linear transformation $T : V \to V$ has $T^* = T^{-1}$, then if the vector space is real then T is orthogonal, and if the vector space is complex then T is unitary.*

In either case, where we have $T^* = T^{-1}$, we immediately get $\langle v, w \rangle = \langle v, T^{-1}Tw \rangle = \langle Tv, Tw \rangle$, and thus $||v|| = ||Tv||$. In fact we also get $\langle v, w \rangle = \langle Tv, Tw \rangle$ implies that $T^*Tv = v$, so $T^* = T^{-1}$ and the statements are equivalent. Further, we find that length preservation determines the inner product from manipulation of $||v + w||$ and $||v + iw||$ to determine $\langle v, w \rangle$, so in fact all three statements are equivalent.

Theorem 16 *If $T : V \to V$ is unitary and V is a finite dimensional complex vector space, then there exists an orthonormal basis of eigenvectors.*

By algebraic closure of $\mathbb{C}$ we can get an eigenvector v for which $\langle v \rangle$ is T-invariant, so $\langle v \rangle^\perp$ is also T-invariant (see by using m_T to write T^{-1} as a polynomial in T). We can induct using this to prove the theorem.

Theorem 17 *If $T : V \to V$ is orthogonal and V is a finite dimensional real vector space, then there exists an orthonormal basis such that the matrix of T is block diagonal formed of I, $-I$, and 2-dimensional rotation matrices.*

Write $S = T + T^{-1} = T + T^*$, so S is self-adjoint, meaning there is an orthonormal basis of eigenvectors for S. We can thus decompose V into the direct sum of $V_i = E_{\lambda_i}$ with λ_i the ith eigenvalue of S. S commutes with T so each of these is also T-invariant. We can get the eigenvalues of $T|_{V_i}$ as the roots of the polynomial $x^2 - \lambda_i x + 1$. If $\lambda_i \neq \pm 2$ then by orthogonality the eigenvalues are complex, so $\{v, Tv\}$ are LI, and $\langle v, Tv \rangle$ is T-invariant so we can use a 2D rotation matrix for this subspace.