

PMath 245 - Advanced Linear Algebra 2

Benjamin Smith*

This is based on a course by Che Tat Ng (Fall semester 2004). All errors are the responsibility of the author.

Contents

Contents	1
1 The determinant function	2
1.1 Properties of determinants	3
1.2 Effect of row operations on evaluation of determinant	3
1.3 Cramer's rule	5
2 Eigen values & eigen vectors	6
2.1 Diagonalization of a matrix (or an operator)	7
Algorithm for diagonalization	7
2.2 Geometric and algebraic multiplicity	8
2.3 Primary decomposition theorem	9
2.4 Annihilating polynomials, Minimal polynomials and partial diagonalization	10
2.5 Cayley-Hamilton theorem	10
2.6 Canonical decomposition	11
2.7 Cyclic subspaces	11
2.8 Another application of PDT	11
3 Inner product spaces (revisited)	13
3.1 The question of orthogonal diagonalization	13
3.2 The adjoint of an operator	13
3.3 Spectral theorem	14
3.4 Application of spectral theorem; Quadratic forms	15
3.5 Projections	16
3.6 Singular value decomposition	17
4 Canonical decompositions of linear operators	18
4.1 Jordan and rational normal forms	19
5 Review Lectures	21

*Please send any other errors to bh2smith@gmail.com

Chapter 1

The determinant function

Let $A = (a_{ij}) \in M_n(\mathbb{F})$ (be an $n \times n$ matrix with entries from a field $\mathbb{F}$). If A is 1×1 then $\det A = a_{11}$, the entry of A . For $n \geq 2$ for each (i, j) write $\tilde{A}_{ij}$ for the matrix obtained from A by deleting row i and column j . Now, for $A \in M_n(\mathbb{F})$ then $\tilde{A}_{ij} \in M_{n-1}(\mathbb{F})$ and we inductively define

$$\det A = \sum_{j=1}^n (-1)^{j+1} a_{1j} \det \tilde{A}_{1j}. \quad (1.1)$$

This formula is known as the *cofactor expansion* of the determinant of A along row 1.

Example 1.0.1. 1.

$$\begin{aligned} \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= (-1)^{1+1} a_{11} \det(a_{22}) + (-1)^{2+1} a_{12} \det(a_{21}) \\ &= a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

2.

$$\begin{aligned} \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} &= (-1)^{1+1} a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} + (-1)^{2+1} a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + (-1)^{3+1} a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \end{aligned}$$

Note that this 3×3 determinant is often used to compute cross products of vectors $a, b \in \mathbb{F}^3$ as follows;

$$\det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = \mathbf{i}(a_2b_3 - a_3b_2) - \mathbf{j}(a_1b_3 - a_3b_1) + \mathbf{k}(a_1b_2 - a_2b_1)$$

3. If A is 4×4 we proceed with the definition and write out the determinant of A to get $4 \cdot 6 = 24$ terms; 12 positive and 12 negative.

Example 1.0.2. Evaluate the following determinant in $\mathbb{Z}/5$;

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & 0 & 1 & 0 \end{pmatrix} &= 1 \det \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} - 2 \det \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + 3 \det \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} - 0 \\ &= 1(-1) - 2(1 + 2 - 2 - 1) + 3(2) \\ &= 5 \equiv 0 \pmod{5} \end{aligned}$$

The development of the theory of the determinant function has a long history (read about it).

An alternative definition to the cofactor expansion for determinant is

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \quad (1.2)$$

where $\mathfrak{S}_n$ is the set of all bijective maps $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ (i.e. *permutations* of $[n] = \{1, 2, \dots, n\}$) and sgn , the *signature*, is a map from $\mathfrak{S}_n$ to $\{\pm 1\}$ which is the unique morphism on the group $\mathfrak{S}_n$ satisfying that signature of the identity is 1 and $\text{sgn}(12) = -1$. Furthermore, $\text{sgn}(\sigma_1\sigma_2) = \text{sgn}(\sigma_1) \cdot \text{sgn}(\sigma_2)$ for any $\sigma_1, \sigma_2 \in \mathfrak{S}_n$.

Explicit determination of signature by example

Let $n = 9$, there are $9!$ elements in $\mathfrak{S}_9$. Examples, in *disjoint cycle notation*, are

$\sigma_1 = (1\ 2\ 3\ 4)(5)(6)(7)(8)(9)$ and $\sigma_2 = (1\ 9)(2\ 8)(3\ 7)(4\ 6\ 5)$ where we assign positive signature to cycles of odd length and negative otherwise. Now,

$$\text{sgn}(\sigma_1) = \text{sgn}(1\ 2\ 3\ 4) \text{sgn}(5) \text{sgn}(6) \text{sgn}(7) \text{sgn}(8) \text{sgn}(9) = -1$$

and

$$\text{sgn}(\sigma_2) = \text{sgn}(1\ 9) \text{sgn}(2\ 8) \text{sgn}(3\ 7) \text{sgn}(4\ 6\ 5) = (-1)(-1)(-1)(1) = -1.$$

In our alternate definition of determinant for a 9×9 matrix, of the $9!$ terms is

$$+a_{17}a_{21}a_{35}a_{42}a_{56}a_{68}a_{79}a_{84}a_{93}$$

where the signature was determined by $(1\ 7\ 9\ 3\ 5\ 6\ 8\ 4\ 2)$.

1.1 Properties of determinants

Theorem 1.1.1 (4.3 from book). *The determinant is a linear function in each row when the remaining rows are held fixed.*

Sketch: This is seen immediately by induction using the cofactor expansion (equation (1.1)) where we have swapped the desired row into the first and taken note of the effect of determinants under elementary row operations (see soon 1.2). $\square$

Lemma 1.1.2. *If A has its i^{th} row equal to the standard basis vector e_k , then*

$$\det A = (-1)^{i+k} \det \tilde{A}_{ik}.$$

Theorem 1.1.3 (4.4 from book). *Cofactor expansion can be performed along any row. That is*

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det \tilde{A}_{ij}.$$

Theorem 1.1.4. $\det A = \det A^T$.

Proof. Let $A = (a_{ij})$ and $B = A^T$ so that $a_{ij} = b_{ji}$. A term $\text{sgn } \sigma a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ in the determinant of A will be equal to a corresponding term in $\det B$ given by

$$\text{sgn}(\sigma^{-1}) b_{\sigma(1)1} \cdots b_{\sigma(n)n} = \text{sgn } \sigma^{-1} b_{1\sigma^{-1}(1)} \cdots b_{n\sigma^{-1}(n)}.$$

Thus $\det A = \det B$ once we see that $\text{sgn } \sigma = \text{sgn } \sigma^{-1}$ which is easy to observe since their disjoint cycle decompositions are of the same shape. $\square$

Corollary 1.1.5. *Every theorem about determinant stated about rows of A remains true for the columns as well.*

1.2 Effect of row operations on evaluation of determinant

1. If B is obtained from A by the ERO $R_i \mapsto cR_i$, then $\det B = c \cdot \det A$.
2. If B is obtained from A by interchanging rows, then $\det B = -\det A$.
3. If B is obtained from A by adding to a row, a scalar multiple of another row, then $\det B = \det A$.

Example 1.2.1. Evaluate $\det \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \\ 0 & 1 & 3 & 1 \\ 1 & 1 & 2 & 1 \end{pmatrix}$ in $\mathbb{Z}/5$.

Solution: First, subtracting the second column from the fourth we have

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \\ 0 & 1 & 3 & 1 \\ 1 & 1 & 2 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 4 & -1 \\ 0 & 1 & 3 & 0 \\ 1 & 1 & 2 & 0 \end{pmatrix}$$

which does not affect the determinant. So evaluating (by expansion along the fourth column)

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 4 & -1 \\ 0 & 1 & 3 & 0 \\ 1 & 1 & 2 & 0 \end{pmatrix} &= (-1)^{1+4} 2 \det \begin{pmatrix} 1 & 3 & 4 \\ 0 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix} + (-1)^{2+4} (-1) \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix} \\ &= -2 \det \begin{pmatrix} 0 & 2 & 2 \\ 0 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix} - \det \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix} \\ &= -2(6 - 2) - (3 - 1) = -10 \equiv 0 \pmod{5}. \end{aligned}$$

Theorem 1.2.2. For an upper-triangular¹ matrix A ,

$$\det A = \prod_{i=1}^n a_{ii}.$$

Proof. A matrix A can be upper triangularized using only the third ERO listed above and this operation is known to have no effect on the determinant of A . $\square$

Corollary 1.2.3. A is singular if and only if $\det A \neq 0$.

Theorem 1.2.4. Determinant is multiplicative. That is

$$\det(AB) = \det(A) \cdot \det(B).$$

Proof. It suffices to show this identity holds for elementary matrices and diagonal A (i.e. the Row Echelon form of A). This is because every A is expressible as $A = E_1 E_2 \cdots E_k \cdot C$ where E_i 's are elementary and C is a RREF² of A . Thus,

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k \cdot CB) \\ &= \prod_{i=1}^k \det(E_i) \det C \det B \\ &= \det(A) \cdot \det(B) \end{aligned}$$

Now, for each of the three types of elementary matrices the identity $\det(EB) = \det E \cdot \det B$ is easily verified to hold. For example if E denotes scalar multiplication of a row by λ , then both sides of this equation are seen to have determinant equal to $\lambda \det B$.

Finally, the RREF of A will either be I_n or have a row consisting entirely of 0's. When $C = I_n$ then $\det(CB) = \det C \cdot \det B$ is clear and when C has a last row consisting of 0's, so does the product CB . Hence $0 = \det(CB) = \det C = \det C \cdot \det B$ as required. $\square$

Background motivation for $\det AB = \det A \det B$

Let $A, B \in M_3(\mathbb{R})$. They can be conceived as linear maps $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. The absolute value of the determinant is called the *magnification power*. That is, the region in $\mathbb{R}^3$ having volume V will have a new volume, under the image of A , given by $|\det A| \cdot V$.

¹i.e. $a_{ij} = 0$ for $i > j$

²Row-Reduced-Echelon-Form

1.3 Cramer's rule

The equation $A\vec{x} = \vec{b}$ where $A \in M_n$ is invertible can be solved by

$$x_i = \frac{\det [\vec{a}_1 | \cdots | \vec{a}_{i-1} | \vec{b} | \vec{a}_{i+1} | \cdots | \vec{a}_n]}{\det A}$$

where $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\vec{a}_i$ denote the columns of A .

Definition 1.3.1. Let $A \in M_n$ and let $c_{ij} = (-1)^{i+j} \det \tilde{A}_{ij}$. These are called the (i, j) -*cofactors* of A and the transpose of the matrix formed by the cofactors is called the *adjugate* of A , denoted $\text{adj}(A)$. That is $\text{adj } A = (c_{ij})^T$.

Theorem 1.3.2.

$$A \cdot \text{adj } A = \det A \cdot I_n.$$

Proof. Examining the (i, j) -entry of this product reveals,

$$(A \cdot \text{adj } A)_{ij} = \sum_{k=1}^n a_{ik} c_{kj} = \sum_{k=1}^n a_{ik} (-1)^{k+j} \det \tilde{A}_{jk}.$$

When $i = j$ this is exactly the cofactor expansion for the determinant of A along the i^{th} row. However, when $i \neq j$, this is equivalent to the cofactor expansion along the j^{th} row of a matrix constructed from A by replacing the j^{th} row with the i^{th} row of A . Hence, this matrix has two identical rows so is singular. That is to say, $(A \cdot \text{adj } A)_{ij} = \det A \cdot \delta_{ij}$ as claimed. $\square$

Corollary 1.3.3. If A is invertible then,

$$A^{-1} = \frac{1}{\det A} \text{adj } A.$$

Characterization of the determinant function

The determinant is the unique function $f : M_n(\mathbb{F}) \rightarrow \mathbb{F}$ with the following properties;

1. *multi linearity*; It is linear in each of the n rows. That is, for all $X, Y \in \mathbb{F}^n, \lambda \in \mathbb{F}$

$$f(R_1, R_2, \dots, R_{i-1}, \lambda X + Y, R_{i+1}, \dots, R_n) = \lambda f(R_1, \dots, X, \dots, R_n) + f(R_1, \dots, Y, \dots, R_n).$$

2. *Alternating*; If two adjacent rows are equal, then $f = 0$.

3. $f(I_n) = 1$.

and these three properties characterize the determinant function.

Chapter 2

Eigen values & eigen vectors

Throughout this chapter we will be concerned with the structure of a linear operator¹ $T : V \rightarrow V$.

Definition 2.0.4. An *eigenvalue* of a linear operator $T : V \rightarrow V$ is a scalar $\lambda \in \mathbb{C}$ such that there exists non-zero *eigenvector* $v \in V$ satisfying that

$$Tv = \lambda v.$$

Example 2.0.5. If $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denotes the counter clockwise rotation about the origin by 30° , there is no real eigen vector and hence no eigen value. However if $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ denotes the same rotation about the z -axis then the z -axis is an eigen vector with eigen value 1 and there are no others.

Theorem 2.0.6 (For operators on finite dimensional vector spaces). *Let A in $M_n(\mathbb{F})$ (representing an operator). A scalar $\lambda \in \mathbb{F}$ is an eigen value for A if and only if $\det(A - \lambda I_n) = 0$.*

Proof. λ is an eigen value for A if and only if there exists non-zero $x \in \mathbb{F}^n$ such that $Ax = \lambda x$ which is equivalent to $(A - \lambda \cdot I_n)x = 0$. This is equivalent to the fact that $\ker A - \lambda \cdot I_n \neq \{0\}$ so that $A - \lambda \cdot I_n$ is not invertible which, according to Lemma 1.2.3, means $\det(A - \lambda \cdot I_n) = 0$. $\square$

Definition 2.0.7. The *characteristic polynomial* of a matrix $A \in M_n$ is defined as

$$\chi_A(\lambda) = \det(A - \lambda \text{id}_n).$$

Note now that the previous theorem can be restated as "the eigen values of A are the roots of its characteristic polynomial". Observe that the characteristic polynomial of an $n \times n$ matrix is a degree n polynomial with leading coefficient $(-1)^n$.

Example 2.0.8. Find all eigen values for the linear map $T : (\mathbb{Z}/5)^2 \rightarrow (\mathbb{Z}/5)^2; (x, y) \mapsto (2x + y, 3x - y)$.

Solution: Consider the matrix representation of T with respect to the standard basis. That is $A = [T]_{std} = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$. The characteristic polynomial is then

$$\det \begin{pmatrix} 2 - \lambda & 1 \\ 3 & -1 - \lambda \end{pmatrix} = (2 - \lambda)(-1 - \lambda) - 3 = \lambda^2 - \lambda - 5 \equiv \lambda^2 - \lambda = \lambda(\lambda - 1).$$

So the eigen values of T are 0, 1.

In order to be more general, it will be important to define everything in terms of linear operators rather than matrices.

Definition 2.0.9. The *characteristic polynomial* for an operator T on finite dimensional V is defined by $\chi_T(x) = \det(A - x \cdot I_n)$ for any matrix $A = [T]_\beta^\beta$ where β is any basis of V . In order for this to be well-defined, we must show that $\det(A_1 - \lambda I_n) = \det(A_2 - \lambda I_n)$ for $A_i = [T]_{\beta_i}^{\beta_i}$.

¹a linear transformation from a vector space to itself

That is, if $A_1 \sim A_2$ (are similar), there there exists Q , a matrix representation of the identity mapping between two different bases for V , such that $A_2 = QA_1Q^{-1}$. Having this connection between A_i 's through Q and that determinant is multiplicative implies

$$\begin{aligned}\det(A_2 - \lambda I_n) &= \det(QA_1Q^{-1} - \lambda I_n) \\ &= \det(QA_1Q^{-1} - \lambda QQ^{-1}) \\ &= \det Q(A_1 - \lambda)Q^{-1} \\ &= \det Q \det(A_1 - \lambda) \det Q^{-1} \\ &= \det(A_1 - \lambda) \det QQ^{-1} \\ &= \det(A_1 - \lambda)\end{aligned}$$

Corollary 2.0.10. *An operator T on a finite dimensional vector space V can have at most $\dim V$ distinct eigen values.*

Example 2.0.11 (Counter example for the above statement in infinite dimensions). Let V be the space of smooth function on the real line (i.e. infinitely differentiable $f : \mathbb{R} \rightarrow \mathbb{R}$). Let $T : V \rightarrow V$ be the operation of differentiation (once). Given a scalar λ , we see that $Tf = \lambda f$ is satisfied by $f(x) = e^{\lambda x}$. Hence λ is an eigen value for every $\lambda \in \mathbb{R}$.

Definition 2.0.12. Let T be an operator and λ an eigen value of T . Consider $E_\lambda(T) = \{v \in V : Tv = \lambda v\}$ whose non-zero vectors are the eigen vectors of T having eigen value λ . This is a subspace² of V and is called an *eigen space* for T (associated with the value λ).

2.1 Diagonalization of a matrix (or an operator)

Definition 2.1.1. A matrix $A \in M_n$ is *diagonalizable* if it is similar to a diagonal matrix. That is, if there exists a nonsingular (change of basis) matrix P such that $P^{-1}AP$ is a diagonal matrix.

More generally, as we will soon see, an operator $T : \mathcal{V} \rightarrow \mathcal{V}$ is said to be *diagonalizable* if there exists a basis of $\mathcal{V}$ consisting entirely of eigenvectors for T .

Proposition 2.1.2. *A necessary condition for a matrix A to be diagonalizable is that its characteristic polynomial factors (completely) into linear factors.*

Proof. Suppose A is diagonalizable, then A is similar to a diagonal matrix $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ so that $\chi_A(\lambda) = \chi_D(\lambda)$. Now $\chi_D(\lambda) = \det \text{diag}(\lambda_1 - \lambda, \dots, \lambda_n - \lambda) = \prod_{i=1}^n (\lambda_i - \lambda)$ (a complete factorization into linear factors). $\square$

Example 2.1.3. If $A \in M_2(\mathbb{R})$ and $\chi_A(\lambda) = \lambda^2 + 1$ then A is not diagonalizable (over $\mathbb{R}$) (e.g. $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$).

Are there other necessary conditions for A to be diagonalizable? Let A be diagonalizable, then this is equivalent to the statement that there exists an ordered basis β for the operator $T : \mathbb{F}^n \rightarrow \mathbb{F}^n$ (given as left multiplication by A) so that $[A]_\beta^\beta = D = \text{diag}(\lambda_1, \dots, \lambda_n)$. If $\beta = \{b_1, b_2, \dots, b_n\}$ this requires that $Ab_i = \lambda_i b_i$ for each i . That is β is a basis whose members are all eigen vectors of A . Thus, whether A is diagonalizable or not depends on whether it admits sufficiently many (n) linearly independent eigen vectors.

Algorithm for diagonalization

The following simple recipe can be applied to determine all aspects about the diagonalizability of any matrix.

Let A be any $n \times n$ matrix.

1. Find the *eigenvalues* $\lambda_1, \dots, \lambda_n$ of A . These are the roots of the *characteristic polynomial*

$$\chi_A(\lambda) = \det(A - \lambda \cdot I_n)$$

²since $E_\lambda = \{v \in V : (T - \lambda \cdot I)v = 0\} = \ker(T - \lambda \cdot I)$ and kernels are subspaces

2. For each of the distinct, eigen values, find the *eigenspaces* E_{λ_i} (for each λ_i). These are the null spaces

$$E_{\lambda_i} = \ker(A - \lambda_i \cdot I_n).$$

At this point, you will be able to determine whether or not A is diagonalizable. That is,

- If you find n linearly independent (i.e. a basis of) eigenvectors for A , then A is diagonalizable,
- otherwise, A is not diagonalizable.

3. With our basis of eigenvectors $v_1, \dots, v_n$ we can construct the *change of basis matrix* as

$$P = [v_1 | v_2 | \dots | v_n]$$

whose columns are the eigenvectors of A .

Note: the order in which you place the columns of P determines the order in which the eigenvalues will appear on the diagonalized version of A

4. Finally, we will obtain the diagonalized version of A as

$$D_A = P^{-1}AP.$$

Example 2.1.4. Determine whether $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ is diagonalizable.

Solution: First, the characteristic polynomial is

$$\begin{aligned} \chi_A(\lambda) &= \det \begin{pmatrix} 1-\lambda & 2 & 0 \\ 2 & 3-\lambda & 0 \\ 0 & 0 & 4-\lambda \end{pmatrix} = [(1-\lambda)(3-\lambda) - 4](4-\lambda) \\ &= (\lambda^2 - 4\lambda - 1)(4-\lambda) \\ &= (\lambda - (2 + \sqrt{5}))(\lambda - (2 - \sqrt{5}))(4-\lambda) \end{aligned}$$

So the eigen values are distinct³!

To demonstrate the entire algorithm, compute the eigen spaces;

$$E_{2+\sqrt{5}} = \ker \begin{pmatrix} 1-2-\sqrt{5} & 2 & 0 \\ 2 & 3-2-\sqrt{5} & 0 \\ 0 & 0 & 4-2-\sqrt{5} \end{pmatrix} = \text{span}\left\{\left(1, \frac{1+\sqrt{5}}{2}, 0\right)\right\}$$

Similarly,

$$E_{2-\sqrt{5}} = \text{span}\left\{\left(1, \frac{1-\sqrt{5}}{2}, 0\right)\right\} \text{ and } E_4 = \text{span}\{(0, 0, 1)\}.$$

Now, to put A in diagonal form, write $P = \begin{pmatrix} 1 & 1 & 0 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$ then

$$P^{-1}AP = \text{diag}(2 + \sqrt{5}, 2 - \sqrt{5}, 4).$$

2.2 Geometric and algebraic multiplicity

For $A \in M_n(\mathbb{F})$ with eigen value λ_0 . The *geometric multiplicity* of λ_0 is defined as $\dim(E_{\lambda_0})$. The *algebraic multiplicity* of λ_0 is defined as the highest power of $(\lambda - \lambda_0)$ appearing in the factorization of the characteristic polynomial of A (note that this is always ≥ 1).

Proposition 2.2.1. *The geometric multiplicity less than or equal to the algebraic multiplicity.*

³Note: this automatically implies diagonalizability since eigen spaces are always at least 1 dimensional and eigen vectors for distinct eigen values are independent

Proof. Let $\{v_1, \dots, v_r\}$ be a basis for E_{λ_0} . This is an independent family in $\mathbb{F}^n$ and can be extended to a basis, say $\{v_1, \dots, v_r, u_{r+1}, \dots, u_n\}$. Under this new basis A is represented by

$$A = \left(\begin{array}{ccc|c} \lambda_0 & \dots & 0 & B \\ & \ddots & & \\ 0 & & \lambda_0 & \\ \hline 0 & \dots & 0 & C \end{array} \right)$$

where B and C are unknown. If we compute the characteristic polynomial using the above "partially diagonalized" matrix, we see that

$$\chi_A(\lambda) = \det \underbrace{\text{diag}(\lambda - \lambda_0, \dots, \lambda - \lambda_0)}_{r \text{ times}} \cdot \chi_C(\lambda) = (\lambda - \lambda_0)^r \cdot \chi_C(\lambda)$$

and $\chi_C(\lambda)$ may or may not contain additional factors of $(\lambda - \lambda_0)$. Hence implying that the algebraic multiplicity is at least as big as the geometric multiplicity. $\square$

With all that we have seen so far, another condition for the diagonalizability of A is that the geometric multiplicity of each distinct eigen value coincides with the algebraic multiplicity.

2.3 Primary decomposition theorem

Let us recall some material from Linear algebra I regarding the theory of subspaces; If $W_1, W_2 \leq V$ are subspaces then $W_1 + W_2 = \{w_1 + w_2 : w_i \in W_i\}$ is again a subspace. Furthermore, if $W_1 \cap W_2 = \{0\}$ we say that the sum is *direct* and write $W_1 \oplus W_2$.

The following three statements are equivalent;

- (i) $W_1 \cap W_2 = \{0\}$,
- (ii) if β_i is a basis for W_i , then $\beta_1 \cup \beta_2$ is a basis for $W_1 + W_2$
- (iii) If $v_1, \dots, v_n$ is an independent family in W_1 and $u_1, \dots, u_m$ is the same for W_2 , then $v_1, \dots, v_n, u_1, \dots, u_m$ is independent in V .

Now,

Theorem 2.3.1. *If V is an $\mathbb{F}$ -vector space and $T : V \rightarrow V$ is an operator. Let $p(x), q(x) \in F[x]$ be polynomials which are coprime (i.e. $\gcd(p, q) = 1$) then*

$$\ker[p(T) \cdot q(T)] = \ker p(T) \oplus \ker q(T).$$

Example 2.3.2. Let V be the space of smooth functions from $\mathbb{R}$ to $\mathbb{R}$ and let D denote the differentiation operator.

$$\ker D = \{f : f(x) = c\} - \text{the constant functions}$$

and

$$\ker(D + 5) = \{f : f' - 5f = 0\} = \{f(t) = ke^{-5t}\}.$$

Now, our job is to **solve the differential equation** $f'' - 5f' + 6 = 0$.

Solution: The solution set $\{f : f'' - 5f' + 6 = 0\}$ is equal to

$$\ker(D^2 - 5D + 6) = \ker(D - 3)(D - 2) = \ker(D - 3) \oplus \ker(D - 2) = \{ke^{2t}\} \oplus \{ke^{3t}\} = \{f : f = ae^{2t} + be^{3t}\}.$$

Another application of this is to proof the following

Lemma 2.3.3. *If $\lambda_1, \dots, \lambda_s$ are distinct eigen values of $A \in M_n$ and if β_i is a basis for E_{λ_i} then the union $\bigcup_{i=1}^s \beta_i$ is always independent.*

Proof. Indeed, using PDT, $E_{\lambda_i} = \ker(A - \lambda_i)$ and $(\lambda - \lambda_1), \dots, (\lambda - \lambda_s)$ are all coprime (as the λ_i 's are all distinct) so $E_{\lambda_1} + E_{\lambda_2} + \dots + E_{\lambda_s}$ is a direct sum. $\square$

2.4 Annihilating polynomials, Minimal polynomials and partial diagonalization

Definition 2.4.1. A polynomial of positive degree $p(x) \in F[x]$ is an *annihilating polynomial* for $T : V \rightarrow V$ (over F) if $p(T) = 0$.

Note that if p annihilates T , then so will any polynomial that is divisible by p . The most interesting annihilating polynomials are those of lowest possible degree.

Definition 2.4.2. If an annihilating polynomial exists at all, we call the monic one of lowest possible degree the *minimal polynomial* for T .

Lemma 2.4.3. If m_T is the minimal polynomial for T , then multiples of m_T generate all possible annihilating polynomials of T .

Proof. Suppose $p(x)$ annihilates T . By the division algorithm, we may write $p(x) = m_T(x) \cdot q(x) + r(x)$ where the degree of r is strictly less than that of m_T . However, now we have $r(T) = p(T) - m_T(T)q(T) = 0$. Since m_T is annihilating polynomial for T of minimal possible degree, then we must have that $r(x) = 0$ and hence $p(x)$ is a multiple of m_T . $\square$

Lemma 2.4.4. If V is a finite dimensional vector space, then an annihilating polynomial always exists. Hence minimal polynomials always exist for finite dimensional vector spaces.

Proof. $\{1, T, T^2, T^3, \dots\}$ is an infinite sequence of operators on V . If $\dim V = n < \infty$ then the space of all operators is again finite dimensional ($\dim \text{Hom}(V, V) = n^2$). Hence, our sequence of operators has to be dependant. Thus, there exists a non-zero combination $\sum_{i=1}^N a_i T^i = 0$ and hence $p(x) = \sum_{i=1}^N a_i x^i$ is annihilating. $\square$

2.5 Cayley-Hamilton theorem

Theorem 2.5.1 (Cayley-Hamilton theorem). If $T : V \rightarrow V$ is an operator with $\dim V = n < \infty$ then the characteristic polynomial is an annihilating polynomial for T .

Theorem 2.5.2 (Diagonalization theorem). $A \in M_n(\mathbb{F})$ is diagonalizable if and only if its minimal polynomial splits with no repeated factors.

Definition 2.5.3. Let $T : V \rightarrow V$ be operator. A subspace $W \leq V$ is called *T -invariant* if $TW \subset W$.

Example 2.5.4. If $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is rotation about the z -axis by 30° then the invariant subspace are the z -axis, the xy -plane and $\{0\}$.

Let T be a linear operator on V . If $V = W_1 \oplus W_2$ and W_i are T -invariant, let β_i be bases for W_i . Then $\beta = \beta_1 \cup \beta_2$ is a basis for V and $[T]_\beta$ is block diagonal of the form $\left(\begin{array}{c|c} [T]_{W_1} \beta_1 & 0 \\ \hline 0 & [T]_{W_2} \beta_2 \end{array} \right)$.

Theorem 2.5.5. Let T be a linear operator. Then $\ker p(T)$ is T -invariant for any polynomial $p(x) \in \mathbb{F}[x]$.

Proof. To show that $T \ker p(T) \subset \ker p(T)$. That is $Tv \in \ker p(T)$ for all $v \in \ker p(T)$. Let $v \in \ker p(T)$ and evaluate

$$p(T)(Tv) = p(T)T(v) = Tp(T)(v) = T(0) = 0$$

as required. $\square$

Theorem 2.5.6. Let T be an operator (or matrix) on finite dimensional V . Let $\chi_T(x) = \prod_{i=1}^m q_i(x)^{r_i}$ where the q_i 's are irreducible factors. Then, by Cayley-Hamilton and Primary decomposition theorem.

$$V = \ker \chi_T(T) = \ker \left(\bigoplus_{i=1}^m q_i(T)^{r_i} \right) = \bigoplus_{i=1}^m \ker q_i(T)^{r_i}$$

is a decomposition of V into a direct sum of T -invariant subspaces.

Example 2.5.7. Let $A \in M_5(\mathbb{R})$ and suppose that $\chi_A(x) = -(x-1)^3(x^2+1)$. Then A is not diagonalizable, but we see that A is similar to a block diagonal matrix $B = \left(\begin{array}{c|c} 3 \times 3 & 0 \\ \hline 0 & 2 \times 2 \end{array} \right)$. That is, there exists P such that $B = P^{-1}AP$. Furthermore, we can see that the 2×2 corner is not diagonal because the characteristic polynomial $x^2 + 1$ does not factor over $\mathbb{R}$. The 3×3 corner has characteristic polynomial $-(x-1)^3$, thus has minimal polynomial $(x-1), (x-1)^2, (x-1)^3$. This corner is diagonalizable if and only if its minimal polynomial is $(x-1)$ and would be I_3 .

2.6 Canonical decomposition

Let T be an operator on a finite dimensional V (or that V is infinite dimensional and T has an annihilating monic polynomial), then V can be decomposed as the direct sum of T -invariant subspaces $V = \bigoplus_{i \in I} W_i$ so that none of the W_i 's can be decomposed further (i.e. the W_i 's are *irreducible T -invariant subspaces*).

Furthermore, each W_i is finite dimensional, cyclic and $T|_{W_i}$ has characteristic polynomial equal to its minimal polynomial (up to a constant factor). The decomposition is unique in the sense that the number of W_i 's having a common minimal polynomial is unique.

Definition 2.6.1. Two operators $T_1, T_2 : V \rightarrow V$ are similar if there exists a $P : V \rightarrow V$ such that $T_2 = P^{-1}T_1P$.

Two operators with annihilating polynomial are similar if and only if, in their canonical decomposition, the above counts agree.

2.7 Cyclic subspaces

For $T : V \rightarrow V$, let $v \neq 0$ in V be fixed. Consider $W = \text{span}\{v, Tv, T^2v, \dots\} \leq V$ (called the *Orbit* of v under T).

Proposition 2.7.1. W is T -invariant.

Proposition 2.7.2. If T has annihilating polynomial $p(x) = \sum_{i=0}^n c_i x^i$ with $c_n = 1$ then $\dim W \leq n$.

Proof. Notice that $\sum_{i=0}^n c_i T^i = 0$ on V . Thus, for any fixed v , $\sum_{i=0}^n c_i T^i v = 0$ where $c_n \neq 0$. Hence, the list $\{v, Tv, \dots, T^n v\}$ is dependent and $T^n v \in \text{span}\{v, Tv, \dots, T^{n-1}v\}$ implying that $\dim W \leq n$. $\square$

Proposition 2.7.3. For non zero $v \in V$, assuming T has an annihilating polynomial, the dimension of W is equal to the first k such that $T^k v \in \text{span}\{v, Tv, \dots, T^{k-1}v\}$. A basis for W is then $\beta = \{v, Tv, \dots, T^{k-1}v\}$.

T -invariant subspaces generated by the use of orbits are called *T -cyclic* subspaces of T . For finite dimensional T -cyclic subspaces with basis chosen as $\beta = \{v, Tv, \dots, T^{k-1}v\}$ the matrix representation for $[Tv]_\beta$ is of the form $\begin{pmatrix} 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & -a_2 \\ 0 & 0 & 1 & -a_3 \end{pmatrix}$ in the case, say $k = 4$, because $T^4 v = -\sum_{i=0}^3 a_i T^i v$. This is called the *companion matrix representation* of T and the characteristic polynomial of $T|_W$ is $(-1)^k (x^k + \sum_{i=0}^{k-1} a_i x^i)$.

2.8 Another application of PDT

If a sequence is defined by a recursive relation and initial values. Say $a_1 = 5$ and $a_{k+1} = \lambda a_k$. In general then then $a_n = 5\lambda^{n-1}$.

Example 2.8.1. Find an explicit formula for the *Fibonacci sequence* defined by $a_1 = a_2 = 1$ and $a_n = a_{n-1} + a_{n-2}$.

Solution: Consider the space V of all real sequences $\{a_n\}_{n=1}^\infty$ and let T be the *left unilateral shift operator* $T(\{a_n\}) = \{a_2, a_3, a_4, \dots\}$. Check that T is an operator (linear).

Next point of interest is satisfying the recursive formula $a_n = a_{n-1} + a_{n-2}$. Observe that

$$\begin{aligned}\{a_n\} &= \{a_1, a_2, a_3, \dots\} \\ T\{a_n\} &= \{a_2, a_3, a_4, \dots\} \\ T^2\{a_n\} &= \{a_3, a_4, a_5, \dots\}\end{aligned}$$

So that $T^2\{a_n\} = T\{a_n\} + \{a_n\}$ or $T^2 - T - 1 = 0$ on the subspace W of sequences satisfying the recursion rule. That is $W = \ker(T^2 - T - 1)$. By Primary decomposition theorem, because $x^2 - x - 1$ has roots $\lambda_{\pm} = \frac{1 \pm \sqrt{5}}{2}$, we get a coprime factorization and

$$W = \ker(T - \lambda_+)(T - \lambda_-) = \ker(T - \lambda_+) \oplus \ker(T - \lambda_-).$$

Now, $\{b_n\} \in \ker(T - \lambda_+)$ is equivalent to $T\{b_n\} = \lambda_+\{b_n\}$ hence $b_k = \lambda_+ b_{k-1}$ or $b_n = \lambda_+^{n-1} b_1$. Similarly $\{c_n\} \in \ker(T - \lambda_-)$ if and only if $c_n = \lambda_-^{n-1} c_1$.

Thus $\{a_n\} \in W$ is the sum $\{a_n\} = \{b_n\} + \{c_n\} = b_1 \lambda_+^{n-1} + c_1 \lambda_-^{n-1}$. Now, given initial conditions $1 = a_1 = b_1 + c_1$ and $1 = a_2 = \lambda_+ b_1 + \lambda_- c_1$, it remains to solve

$$\begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} b_1 \\ c_1 \end{pmatrix} = \frac{1}{\lambda_- - \lambda_+} \begin{pmatrix} \lambda_- & -1 \\ -1 & -\lambda_+ \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Chapter 3

Inner product spaces (revisited)

The following theory is intended for real or complex vector spaces only (not over a general field). Assuming that inner product spaces have already been covered in Linear algebra 1, briefly and *inner product* on a $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ vector space V is a non-degenerate, skew-symmetric, bilinear map $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$. Skew-symmetric is also known as ‘conjugate symmetric’ which is not quite symmetric for complex vector space but is for real ones.

Recall, a few results

Theorem 3.0.2 (Gram-Schmidt orthonormalization). *If V is a finite dimensional inner product space, then there exists an orthonormal basis¹ for V .*

Theorem 3.0.3. *Every finite dimensional inner product space V of dimension n is isomorphic to $\mathbb{F}^n$. In fact, fixing an orthonormal basis β for V then the map $v \in V \leftrightarrow [v]_\beta \in \mathbb{F}^n$ is an isometric isomorphism.*

Just to recall some notation, we show that the inner product from the previous theorem is preserved. Let $u, v \in V$ and write our orthonormal basis $\beta = \{v_1, \dots, v_n\}$ so that $u = \sum a_i v_i$ and $v = \sum b_i v_i$. Now, since β is orthonormal (i.e. $\langle v_i, v_j \rangle = \delta_{ij}$) then

$$\langle u, v \rangle = \left\langle \sum_i a_i v_i, \sum_j b_j v_j \right\rangle = \sum_{i,j} a_i \bar{b}_j \langle v_i, v_j \rangle = \sum_i a_i \bar{b}_i = \langle [u]_\beta, [v]_\beta \rangle$$

3.1 The question of orthogonal diagonalization

Let V be a finite dimensional inner product space and $T : V \rightarrow V$ a linear operator on V . T is said to be *orthogonally diagonalizable* if there exists an orthonormal basis of eigen vectors for T .

Theorem 3.1.1. *T is orthogonally diagonalizable if and only if it is diagonalizable and eigen spaces for distinct eigen values are perpendicular. Note that the Gram-Schmidt process will provide orthonormal bases within each of the eigen spaces.*

Example 3.1.2. Consider $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by $T(1,0) = (2,0)$ and $T(1,1) = (3,3)$. Is it orthogonally diagonalizable?

Solution: Eigen values are 2 and 3. Eigen spaces are $E_2 = \text{span}\{(1,0)\}$ and $E_3 = \text{span}\{(1,1)\}$ which are not perpendicular. Hence T is not orthogonally diagonalizable.

3.2 The adjoint of an operator

Let V be an inner product space and T an operator on V . If there exists another linear operator $T^* : V \rightarrow V$ satisfying

$$\langle Tu, v \rangle = \langle u, T^*v \rangle, \forall u, v \in V$$

then it is unique and it called the *adjoint* of T .

¹A basis β is orthonormal if $v_i \perp v_j, \forall i \neq j$ and $\|v_i\| = 1, \forall i$

Proposition 3.2.1. *For finite dimensional vector spaces, adjoints exist.*

Proof. By the isomorphism principal V can be replaced with $\mathbb{F}^n$ (with standard inner product) and T with an $n \times n$ matrix A . The proposition is then equivalent to the claim that every $n \times n$ matrix comes with an adjoint and the definition of adjoint for a matrix A on $\mathbb{F}^n$ is $\langle Au, v \rangle = \langle u, A^*v \rangle$. Notice, however, that here this formula, given by matrix multiplication, is equivalent to

$$\bar{v}^T Au = \overline{A^*v}^T u$$

and we intend to find a matrix A^* satisfying this. Expanding the right hand side says

$$\bar{v}^T Au = \bar{v}^T \overline{A^*}^T u$$

implying that $A = \overline{A^*}^T$ which implies $A^* = \overline{A}^T$ which is uniquely determined. $\square$

Definition 3.2.2. A linear operator T on an inner product space with adjoint T^* is called *self-adjoint* (Hermitian) if $T = T^*$. It is called *normal* if $TT^* = T^*T$.

Clearly, self-adjoint implies normal, and in the simple case of one-by-ones $[i]$ it is easily seen that normal does not imply self-adjoint.

Definition 3.2.3. T is *inner product preserving* if $\langle Tu, Tv \rangle = \langle u, v \rangle$ for all $u, v \in V$.

Inner product preserving T with an adjoint is equivalent to $\langle u, T^*Tv \rangle = \langle u, v \rangle$ and hence $T^*Tv - v = 0, \forall v \in V$. That is $T^*T = 1$

Proposition 3.2.4. T is inner product preserving if and only if $T^*T = I_n$ (on finite dimensional V).

Proposition 3.2.5. On finite dimensional vector spaces, T is inner product preserving if and only if T maps orthonormal bases to orthonormal bases.

Proof. Let $\beta = \{v, \dots, v_n\}$ be an orthonormal basis for V . Then we claim that $\{Tv_1, \dots, Tv_n\}$ is also an orthonormal basis. This is seen by $\langle Tv_i, Tv_j \rangle = \langle v_i, v_j \rangle = \delta_{ij}$.

Conversely, if T maps one orthonormal basis to another, then we can verify (exercise) that T preserves inner product. $\square$

Definition 3.2.6. A real matrix $P \in M_n(\mathbb{R})$ satisfying $P^*P = I_n$ (i.e. $P^T P = I_n$) is called *orthogonal* (rather than inner product preserving). A complex matrix $P \in M_n(\mathbb{C})$ satisfying $P^*P = I_n$ is called *unitary*.

Remark 3.2.7. An operator T (on f.d. V) is orthogonally diagonalizable if and only if it admits an orthonormal basis of eigen vectors (this is the definition). A real matrix $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable if and only if there exists orthogonal P such that $P^{-1}AP = D$. Similarly, $A \in M_n(\mathbb{C})$ is *unitarily diagonalizable* if there exists unitary U such that $U^*AU = D$.

3.3 Spectral theorem

Theorem 3.3.1 (Spectral theorem). *A real matrix $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable if and only if A is symmetric (i.e. self-adjoint). Equivalently, an operator T on a finite dimensional real vector space is diagonalizable if and only if it is self-adjoint.*

A complex matrix $A \in M_n(\mathbb{C})$ is unitarily diagonalizable if and only if A^* commutes with A (i.e. A is normal). Equivalently, an operator T on a finite dimensional complex vector space is unitarily diagonalizable if and only if T is normal.

Proof. Assuming that $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable so that there exists orthogonal P such that $P^{-1}AP = D$. Then $A = PDP^{-1}$ and $A^T = (PDP^{-1})^T = (P^{-1})^T D^T P^T = PDP^{-1} = A$. Similar proof works for unitarily diagonalizable $A \in M_n(\mathbb{C})$.

Conversely, we will need the following 3 lemmas;

Lemma 3.3.2. *If T is an operator with adjoint. If $W \leq V$ is T -invariant, then $W^\perp$ is T^* -invariant.*

proof: Let $W \leq V$ be T -invariant and show that $T^*(W^\perp) \subset W^\perp$. Let $v \in W^\perp$ and for any $w \in W$ consider $\langle w, T^*v \rangle = \langle Tw, v \rangle = 0$ hence $T^*v \in W^\perp$.

Lemma 3.3.3. *Suppose $T : V \rightarrow V$ is normal. If E_λ is an eigen space for T then E_λ is both T and T^* invariant.*

proof: Every eigen space of T is T -invariant (no inner product necessary). To show it is also T^* , recall $E_\lambda = \ker(T - \lambda)$. Now $T^*(E_\lambda) \subset \ker(T - \lambda)$ is equivalent to $((T - \lambda)(T^*(E_\lambda))) = 0$ which is to say that $(T - \lambda)(T^*(v)) = 0$ for all $v \in E_\lambda$. Since T is normal, this means $T^*(T - \lambda)(v) = 0$ which is true since $v \in \ker(T - \lambda)$.

Corollary 3.3.4. *If T is normal and E_λ is an eigen space for T , then both E_λ and $E_\lambda^\perp$ are both T and T^* invariant.*

Lemma 3.3.5. *For $\mathbb{C}$; Let T be normal on a f.d. V . Then V is the direct sum of eigen spaces which are mutually orthogonal.*

proof: By induction on the dimension n of V . If $\dim V = 1$ statement is trivial. Assuming result holds for spaces of dimension lower than n . T must have at least one eigen value, say λ , $\mathbb{C}$ (fundamental theorem of algebra). Now $V = E_\lambda \oplus E_\lambda^\perp$. Moreover E_λ and $E_\lambda^\perp$ are both T and T^* invariant (since T is normal). Now $E_\lambda^\perp$ has dimension strictly less than n as $\dim E_\lambda \geq 1$. Considering now, the restriction of T, T^* to $E_\lambda^\perp$, the induction hypothesis allows us to continue to decompose $E_\lambda^\perp$ into a direct sum of mutually orthogonal subspaces.

Now, the proof for $\mathbb{R}$ is almost identical except one will require the existence of a real eigen value λ which is shown using that $T^* = T$. Indeed, consider $A \in M_n(\mathbb{R})$ as if it were complex and find a complex eigen value $\lambda \in \mathbb{C}$ (which exists). Now, to show that λ is actually real: Suppose $Tv = \lambda v$ for $v \neq 0$ (i.e. $v \in E_\lambda(T)$). Now,

$$\begin{aligned}\langle Tv, v \rangle &= \langle v, T^*v \rangle \\ \lambda \langle v, v \rangle &= \langle v, \lambda v \rangle \\ \lambda \|v\|^2 &= \bar{\lambda} \|v\|^2 \\ \lambda &= \bar{\lambda} \quad \Rightarrow \lambda \in \mathbb{R}.\end{aligned}$$

□

3.4 Application of spectral theorem; Quadratic forms

Definition 3.4.1. A real quadratic form Q is a function of n real variables $x_1, \dots, x_n$ into $\mathbb{R}$ having the special form

$$Q(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

Example 3.4.2. $Q(x, y) = x^2 + 4xy - 5y^2$ for $x, y \in \mathbb{R}$ is a quadratic form in 2-variables.

The constants a_{ij} used in writing a quadratic form are not unique (i.e. the above example can be equivalently expressed as $Q(x, y) = x^2 + 2xy + 2yx - 5y^2$). However, if it is written with symmetric $a_{ij} = a_{ji}$ then it is unique. Then every quadratic form can be expressed uniquely by symmetric $A = (a_{ij})$.

Example 3.4.3. $Q(x, y, z) = 4x^2 - 7yz + 9xy + \sqrt{3}y^2$ has the unique symmetric expression

$$Q(x, y, z) = 4x^2 - \frac{7}{2}yz - \frac{7}{2}zy + \frac{9}{2}xy + \frac{9}{2}yx + \sqrt{3}y^2.$$

Definition 3.4.4. Let Q be the quadric form corresponding to the real symmetric matrix $A \in M_n(\mathbb{R})$. Then the function $Q(\vec{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j$ has the compact expression $Q(\vec{x}) = \vec{x}^T A \vec{x}$. This is called the symmetric matrix representation of Q .

The spectral theorem asserts that for any real symmetric matrix A , there exists orthogonal P such that $P^{-1}AP$ is diagonal. With this, one may simplify quadric forms as follows;

$$Q(x) = x^T A x = x^T (PDP^{-1})x = (x^T P^{-1})D(Px) = (Px)^T D(Px)$$

since $P^{-1} = P^T$. This implies that our quadratic form in the new variable $y = Px$ has no off-diagonal terms. That is, if $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ then $Q(y) = \sum_{i=1}^n \lambda_i y_i^2$.

Example 3.4.5. Sketch the curve $2x^2 - 4xy + 5y^2 = 36$ (i.e. the level curve $Q(x, y) = 36$) in the plane.

Solution: First, the symmetric matrix for Q is $A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$. Its eigen values are $\lambda_1 = 1, \lambda_2 = 6$ and the corresponding eigen spaces are $E_{\lambda_1} = \text{span}\{(2, 1)\}, E_{\lambda_2} = \text{span}\{(-1, 2)\}$ respectively. By Gram-Schmidt, an orthonormal basis of eigen vectors is $\left\{ \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\}$ so let $P = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$ which is orthogonal.

Our new variables are now given by $y = Px$ and in these coordinates $Q(y_1, y_2) = y_1^2 + 6y_2^2$ so the curve (after simplification) becomes $1 = \left(\frac{y_1}{6}\right)^2 + \left(\frac{y_2}{\sqrt{6}}\right)^2$ which is a rotated ellipse with minor and major axes specified by the coefficients. The principal axes are furthermore governed by our basis of eigen vectors encoded in P .

Q: What is the area of this ellipse?

A: $6\sqrt{6}\pi$.

It is because of this application of the spectral theorem that it is also called the *Principal Axis Theorem* (in the $\mathbb{R}$ case).

3.5 Projections

Theorem 3.5.1 (Projection form of Spectral theorem). *Let $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$ be normal and let $E_{\lambda_1}, E_{\lambda_2}, \dots, E_{\lambda_s}$ be the eigen spaces for T (where $\lambda_1, \dots, \lambda_s$ refer to distinct eigen values in $\mathbb{C}$). Then the spaces are mutually orthogonal and $\mathbb{C}^n = \bigoplus_{i=1}^s E_{\lambda_i}$. Hence, furthermore $E_{\lambda_i}^\perp = \bigoplus_{j \neq i}^s E_{\lambda_j}$.*

With this, each vector $v \in \mathbb{C}^n$ can be written uniquely as $v = v_1 + v_2 + \dots + v_s$ with $v_i \in E_{\lambda_i}$. For each fixed i let $\pi_i : \mathbb{C}^n \rightarrow \mathbb{C}^n$ be defined as $\pi_i(v) = v_i$ so then the range of π_i is E_{λ_i} , $\pi_i^2 = \pi_i$ and $\pi_i \circ \pi_j = 0$. This says that π_i 's are projections² with disjoint ranges.

Claim: $\pi_i^* = \pi_i$.

Indeed, we wish to confirm that $\pi_i u = \pi_i^* u$ for each $u \in \mathbb{C}^n$. This is equivalent to $\langle \pi_i u, v \rangle = \langle \pi_i^* u, v \rangle = \langle u, \pi_i v \rangle$ for all $u, v \in \mathbb{C}^n$. Let $u = u_1 + u_2 + \dots + u_s$ and $v = v_1 + \dots + v_s$ where $u_i, v_i \in E_{\lambda_i}$. We must confirm that $\langle u, v_1 + \dots + v_s \rangle = \langle u_1 + \dots + u_s, v_i \rangle$. This is equivalent to the true statement $\langle u_i, v_i \rangle = \langle u_i, v_i \rangle$ since $\langle u_i, v_j \rangle = 0$ for all $i \neq j$.

In short, one says that the projections $\{\pi_i\}_{i=1}^s$ are disjoint and self adjoint. Furthermore, $v = v_1 + \dots + v_s = \pi_1 v + \dots + \pi_s v$ for every $v \in \mathbb{C}^n$ so that $I = \sum_{i=1}^s \pi_i$ (the projections partition the identity). From $v = v_1 + \dots + v_s$ we get $Tv = Tv_1 + \dots + Tv_s = \lambda_1 v_1 + \dots + \lambda_s v_s$ so that $T = \sum_{i=1}^s \lambda_i \pi_i$ and this is the spectral decomposition of T . The set $\{\lambda_1, \dots, \lambda_s\}$ is called the *spectrum* of T .

Also, $T^* = \left(\sum_{i=1}^s \lambda_i \pi_i\right)^* = \sum_{i=1}^s \bar{\lambda}_i \pi_i^* = \sum_{i=1}^s \bar{\lambda}_i \pi_i$.

Conversely, if $\{\pi_i\}_{i=1}^s$ is a family of self-adjoint projections partitioning the identity then $T = \sum_{i=1}^s \lambda_i \pi_i$ yields a normal operator whose adjoint is $\sum_{i=1}^s \bar{\lambda}_i \pi_i$.

Corollary 3.5.2. *An operator T on $\mathbb{C}^n$ is normal if and only if T^* is a polynomial in T . That is, there exists a $p(x) \in \mathbb{C}[x]$ such that $T^* = p(T)$.*

Proof. Suppose first that $T^* = p(T)$ for some p . Then T commutes with T^* (since it commutes with all polynomials in T) so is normal.

Conversely. Suppose T is normal. By the spectral theorem, $T = \sum_{i=1}^s \lambda_i \pi_i$ with λ_i 's distinct and $\{\pi_i\}_{i=1}^s$ is a family of disjoint, self-adjoint projections partitioning the identity and $T^* = \sum_{i=1}^s \bar{\lambda}_i \pi_i$.

²A linear operator whose square is itself is called a projection

Now, provided one believes³, there exists a polynomial $p(x)$ such that $p(\lambda_i) = \bar{\lambda}_i$ for each i . With that, $p(T)$ will give T^* (as required) which is explained as follows: We have $T = \sum_{i=1}^s \lambda_i \pi_i$ and

$$T^2 = \left(\sum_{i=1}^s \lambda_i \pi_i \right) \left(\sum_{i=1}^s \lambda_i \pi_i \right) = \sum_{i,j=1}^s \lambda_i \lambda_j \pi_i \pi_j = \sum_{i=1}^s \lambda_i^2 \pi_i^2 = \sum_{i=1}^s \lambda_i^2 \pi_i$$

so then $T^k = \sum_{i=1}^s \lambda_i^k \pi_i$. Now, $T^2 + T^3 = \sum_{i=1}^s (\lambda_i^2 + \lambda_i^3) \pi_i$ and this generalizes to say

$$p(T) = \sum_{i=1}^s p(\lambda_i) \pi_i = \sum_{i=1}^s \bar{\lambda}_i \pi_i = T^*.$$

□

3.6 Singular value decomposition

Consider the linear map $A : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ given by the matrix $\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$. The product A^*A is then symmetric so the spectral decomposition asserts that it is orthogonally diagonalizable. Note that $A^*A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$, its eigen values (in descending order) are $\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$. Each eigen space

E_{λ_i} is of dimension 1 and their bases are $v_1 = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, v_2 = \frac{1}{3} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$ and $v_3 = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$ respectively where the vectors have already been normalized.

So, $V = (v_1 | v_2 | v_3)$ is a matrix for which $V^* A^* A V = \text{diag}(360, 90, 0) =: \Sigma$. Notice that $A^* A v_i = \lambda_i v_i$ for each i , so

$$\langle A v_i, A v_i \rangle = \langle A^* A v_i, v_i \rangle = \lambda_i \langle v_i, v_i \rangle = \lambda_i$$

implying that $\|A v_i\| = \sqrt{\lambda_i}$. Similarly,

$$\langle A v_i, A v_j \rangle = \dots = \lambda_i \langle v_i, v_j \rangle = 0.$$

Thus, $\{A v_1, A v_2, A v_3\}$ is a list of orthogonal vectors with each vector having length $\sqrt{\lambda_i}$.

Theorem 3.6.1 (Singular value decomposition - for matrices). *Let $A \in M_{m \times n}(\mathbb{C})$ represent a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$ of rank r . Then there exists a ‘diagonal’ matrix $\Sigma \in M_{m \times n}(\mathbb{C})$ whose diagonal*

entries appear in descending order (e.g. $\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & \dots & & 0 \end{pmatrix}$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ and $\sigma_{r+1} = 0$.)

and unitary $U \in M_m(\mathbb{C}), V \in M_n(\mathbb{C})$ such that $A = U \Sigma V^*$.

Furthermore, the matrix Σ is unique. The values $\sigma_1, \dots, \sigma_r$ are called the singular values of A and writing $A = U \Sigma V^*$ is called the singular value decomposition of A

Proof. Consider $A^*A \in M_n(\mathbb{C})$ which is clearly self-adjoint. In particular it is normal and so, by the spectral theorem, is unitarily diagonalizable. That is, there exists an orthonormal basis $\{v_1, \dots, v_n\}$ for $\mathbb{C}^n$ and $(A^*A)v_i = \lambda_i v_i$ for some scalars $\lambda_i \in \mathbb{C}$ (the eigen values of A^*A). Let $V = [v_1 | v_2 | \dots | v_n]$ then V is unitary. Consider $\{A v_1, \dots, A v_n\} \subset \mathbb{C}^m$ which span the range of A . These vectors are, furthermore, mutually orthogonal with lengths given by $\|A v_i\| = \sqrt{\lambda_i}$ (shown above). Let $\sigma_i = \sqrt{\lambda_i}$ so that $\|A v_i\| = \sigma_i$ and, without loss of generality, arrange the order of $\{v_1, \dots, v_n\}$ so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$. Now $\{A v_1, \dots, A v_r\}$ is a basis for the range of A and normalizing to $u_i := \frac{1}{\sigma_i} A v_i$ for $i = 1, \dots, r$ gives an orthonormal basis for the range space. Extending this basis to an orthonormal basis $\{u_1, u_2, \dots, u_r, u_{r+1}, \dots, u_m\}$ for $\mathbb{C}^m$ and let $U = (u_1 | \dots | u_m)$. This U is then unitary and with this we have $A v_i = \sigma_i u_i$ for each i , which reads as $AV = U \Sigma$ (i.e. $A = U \Sigma V^*$) as desired. □

³I would like a demonstration

Chapter 4

Canonical decompositions of linear operators

Lemma 4.0.2. *Let $T : V \rightarrow V$ be an operator and $p(x)$ an annihilating polynomial for T . If $p(x)$ is factored into a product $p(x)p_1^{\gamma_1} \cdots p_s^{\gamma_s}$ where each p_i is irreducible over $\mathbb{F}$, then*

$$V = \bigoplus \ker p_i^{\gamma_i}(T).$$

This is the primary decomposition theorem with $W_i = \ker p_i^{\gamma_i}(T)$ and all are T -invariant. Next, one asks the question; Can these W 's be decomposed further into T -invariant subspaces?

Example 4.0.3. Consider $A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 2 \end{pmatrix}$ representing $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. An annihilating polynomial for A is $p() = -(x-1)(x-2)^2$ (By Cayley-Hamilton) so the above lemma implies $\mathbb{R}^3 = \underbrace{\ker(T-1)}_{W_1} \oplus \underbrace{\ker(T-2)^2}_{W_2}$.

Now, could W_1 or W_2 be decomposed further? For W_1 ; $\dim W_1 = 1$ so there no meaningful further decomposition is possible. However, $\dim W_2 = 2$ and it is decomposable further into T -invariant subspaces if and only if $W_2 = W_3 \oplus W_4$ with $\dim W_3 = 1 = \dim W_4$ if and only if $W_2 = W_3 \oplus W_4$ where each 1-dimensional T -invariant subspace W_3, W_4 carries eigen vectors if an only if $T|_{W_2}$ is diagonalizable if and only if T is diagonalizable if and only if $(x-1)(x-2)$ annihilates T . To check this, compute $(A-1)(A-2) = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}$ which is not zero. Hence W_2 has not further decomposition.

Lemma 4.0.4. *Let $T : V \rightarrow V$ be annihilated by p which is irreducible. Then $V = \bigoplus_{i \in I} W_i$ where each W_i is T -cyclic of dimension equal to the degree of p . This decomposition is final canonical.*

Example 4.0.5. Are there operators T on $\mathbb{R}$ over $\mathbb{Q}$ such that $T^2v = -v$ for all $v \in \mathbb{R}$? This is equivalent to the question “are there operators on $\mathbb{R}$ over $\mathbb{Q}$ which are annihilated by $p(x) = x^2 + 1$?”

To give an example, let $\mathbb{R} = \bigoplus_{i \in I} W_i$ where each W_i is of dimension 2. Let β be a basis for $\mathbb{R}$ over $\mathbb{Q}$ and notice that β is infinite. Let β be separated as a union of disjoint subsets of size 2, say $\beta = \bigcup_{i \in I} s_i$ where $s_i = \{b_1^i, b_2^i\}$ and declare that $W_i = \text{span } s_i$. On W_i , define T by $T(b_1^i) = b_2^i$ and $T(b_2^i) = -b_1^i$ and quickly check that $T^1 = -1$ on W_i . Thus $x^2 + 1$ annihilates $T|_{W_i}$ for each i and so annihilates T on $\bigoplus_{i \in I} W_i$.

Lemma 4.0.6. *If T is an operator on V and is annihilated by an irreducible polynomial $p(x)$ then $V = \bigoplus_{i \in I} W_i$ where each W_i is cyclic of dimension equal to the degree of $p(x)$.*

This lemma will be strengthened slightly as follows; If T is an operator on V annihilated by an irreducible $p(x)$, let $W \leq V$ be a (given) T -invariant subspace. Then there exists a family of T -cyclic subspaces $\{W_i : i \in I\}$ where each W_i is of dimension equal to the degree of p so that $V = W \oplus (\bigoplus_{i \in I} W_i)$. Note that the special case $W = \{0\}$ gives the original lemma.

Proof. Consider the collection of possible families $\{W_j : j \in J\}$ of T -cyclic subspaces where each W_j has dimension equal to the degree of p and with a direct sum $\bigoplus_{j \in J} W_j$ disjoint from W (i.e. intersecting W only at $\{0\}$). We would like to argue that there exists a large enough family such that the sum $W \oplus (\bigoplus_{i \in I} W_i)$ is the full V .

By Zorn's lemma (equivalent to the axiom of choice and/or the transfinite induction principle), there exists a maximal family $\{W_i : i \in I\}$ satisfying all of the above conditions. Maximal meaning that this is a collection which cannot be enriched by a collection which contains it and is truly larger. We need only argue now that this maximal collection satisfies the result of our lemma. Suppose on the contrary that $W \oplus (\bigoplus_{i \in I} W_i) \subsetneq V$.

We will contradict that this collection is maximal by constructing one extra W_{j_0} . To get this, pick $v_0 \in V \setminus W \oplus (\bigoplus_{i \in I} W_i)$ and for convenience let us now denote $W \oplus (\bigoplus_{i \in I} W_i)$ by U . Let W_{j_0} be the T -cyclic subspace generated by v_0 (i.e. $W_{j_0} = \text{span}\{v_0, Tv_0, T^2v_0, \dots, T^{k-1}v_0\}$ where k is the degree of $p(x)$).

Then we claim that $\dim W_{j_0} = k$. Indeed, if $\dim W_{j_0} = l < k$, consider $T|_{W_{j_0}}$. Its characteristic polynomial will be of degree no more than l and $m(x)$ divides every polynomial which annihilates $T|_{W_{j_0}}$. Clearly the irreducible annihilates $T|_{W_{j_0}}$ so $m(x)$ divides $p(x)$ which contradicts that p is irreducible as the degree of m is strictly less than the degree of p .

Next we see that W_{j_0} is disjoint from U . Notice that $W_{j_0} \cap U \leq W_{j_0}$ and that W_{j_0} is T -invariant too. We claim that in W_{j_0} , the only T invariant subspaces are itself and $\{0\}$. Indeed, let $S \leq W_{j_0}$ be a proper T -invariant subspace (i.e. S is not equal to $\{0\}$ or W_{j_0}). Then the minimal polynomial of $T|_S$ will be of degree less than k and its minimal polynomial divides $p(x)$ which is not possible.

Therefore, as $v_0 \notin U$, $W_{j_0} \cap U = \{0\}$ follows. So $\{W_i : i \in I\} \cup W_{j_0}$ is a larger family with all of the desired properties (contradicting the maximality). $\square$

Note that, like W_{j_0} , every W_i cannot have proper T -invariant subspaces, so there are no further T -invariant decompositions.

Example 4.0.7. Suppose that $A \in M_n(\mathbb{F})$ and the minimal polynomial is $m(x) = x^2 + x + 1$. Assuming that $\mathbb{F}$ is a field where $x^2 + x + 1$ is irreducible (e.g. $\mathbb{F} = \mathbb{R}$ and certainly not $\mathbb{C}$). What can be said about A ?

Solution: We can see from the lemma that $\mathbb{F}^n = \bigoplus_{i=1}^{n/2} W_i$ where each W_i has dimension 2. A basis for each W_i is, say, $\{v_i, Av_i\}$. That is $Av_i = 0v_1 + Av_1$ and $A(Av_i) = -Av_i - v_i$ (from the polynomial). Furthermore, then, $A|_{W_i} \simeq \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$ (the companion matrix). Hence, according to this, a canonical decomposition of A is the block diagonal matrix $\text{diag} \left(\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \dots, \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \right)$.

Lemma 4.0.8. If $T : V \rightarrow V$ is annihilated by $q(x)^r$ where $q(x)$ is irreducible (wlog, monic) and $r \geq 1$, then $V = \bigoplus_{i \in I} W_i$ where each W_i is T -invariant of dimension equal to the degree of q^{r_i} with $1 \leq r_i \leq r$. Furthermore, T acts cyclically on W_i (i.e. W_i is T -cyclic) with characteristic polynomial equal to $q(x)^{r_i}$ (up to a sign).

sketch for $r = 2$. Consider $N = q(T)$ along with T . So (since we have assumed $r = 2$) $N^2 = 0$. We have $\text{Range } N \cap \ker N \leq \ker N \leq \ker N^2 = V$.

Algorithm for decomposition: First note that the kernel and range of any polynomial in T is T invariant. This implies that $\text{Range } N \cap \ker N$ is also T -invariant. On this intersection, N annihilates T , so by Lemma 4.0.4 this region decomposes into $\bigoplus W_i$ with each W_i T -invariant of dimension the degree of q . Furthermore on each W_i T acts cyclically with characteristic polynomial q .

Suppose, for demonstration, that $q(x) = x^2 + bx + c$ so that $T^2 = N - bT - c$. Pick $v \in V \setminus \ker N$. Then $T^2v = Nv - bTv - cv \in \text{span}\{v, Tv, Nv\}$ and

$$T^3v = T(Nv - bTv - cv) = TNv - \underbrace{b(Nv - bTv - cv) - cTv}_{\in \text{span}\{v, Tv, Nv\}}.$$

So then, it is clear that $\{v, Tv, T^2v, T^3v, \dots, T^k v\} \in \text{span}\{v, Tv, Nv, TNv\}$ and this forms a basis for the decomposition of $V \setminus \ker N$ which has dimension equal to the degree of q^2 . $\square$

4.1 Jordan and rational normal forms

The canonical decomposition theorem asserts that for each $T : V \rightarrow V$ there is a canonical (fins) decomposition $V = \bigoplus_{i \in I} W_i$ where each W_i is T -cyclic with characteristic polynomial q^{r_i} where q is irreducible. Note that the assumption that T has an annihilating polynomial $p(x)$ here is implicit and $q^{r_i} | p$.

Consider one such W on which T acts cyclically with characteristic polynomial $q(x)^r$. If we use the cyclic basis $\{v, Tv, T^2v, \dots, T^k v\}$ where $k = \deg(q^r) - 1$, then the matrix representation of $T|_W$ is the companion matrix.

Example 4.1.1. If $q(x) = x^2 + 2x + 5$ and $r = 2$, then $q^2(x) = x^4 + 4x^3 + 14x^2 + 20x + 25$, the companion matrix is then

$$C(q^2(x)) = \begin{pmatrix} 0 & 0 & 0 & -25 \\ 1 & 0 & 0 & -20 \\ 0 & 1 & 0 & -14 \\ 0 & 0 & 1 & -4 \end{pmatrix}.$$

When each W_i is equipped with a cyclic basis, we call the decomposition a *Rational canonical form*.

Example 4.1.2. If $A \in M_5(\mathbb{R})$ has characteristic polynomial $p(x) = -(x-3)(x^2+2x+5)$ and if its canonical decomposition is $W_1 \oplus W_2$ where $W_1 = \ker(T-3)$, $W_2 = \ker(T^2+2T+5)^2$ then $A \simeq \text{diag}(C(x-3), C(q(x)^2))$. The notation is $C(x-3), \oplus C(q(x)^2)$. This is the rational normal form of A .

If, instead of the cyclic bases, we use

$$\{v, Tv, T^2v, \dots, T^m v; Nv, TNv, \dots, T^m N; N^2v, TN^2v, \dots, T^m N^2v; \dots; T^m N^{r-1}v\}$$

where $m = \deg q(x) - 1$ and $N = q(T)$. When $T|_W$ is represented as a matrix using the above basis we call the matrix a Jordan block and the canonical decomposition is called the *Jordan normal form*.

Example 4.1.3. Let $q(x) = x^2 + 2x + 5$ and $r = 2$. then the basis is found using the vectors in the following diagram;

$$\begin{array}{ccc} v & \xrightarrow{T} & Tv \\ N \downarrow & & \downarrow \\ Nv & \longrightarrow & TNv \end{array}$$

We have, with $N = T^2 + 2T + 5$, that

$$T(Tv) = T^2v = Nv - 2Tv - 5v,$$

$$TNv$$

and

$$T(TNv) = N^2v - 2TNv - 5Nv = -2TNv - 5Nv.$$

Hence, the matrix representation becomes

$$\left(\begin{array}{cc|cc} 0 & 5 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ \hline 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -2 \end{array} \right) = \left(\begin{array}{c|c} C(q(x)) & 0 \\ \hline N_2 & C(q(x)) \end{array} \right) \sim C(q^2(x)).$$

Chapter 5

Review Lectures

Similarity

Let V be a vector space of dimension 4 and let S, T be linear operators on V with the following descriptions;

For S , there is a basis $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ on which S acts according to the rule $\alpha_1 \mapsto \alpha_2, \alpha_2 \mapsto \alpha_4, \alpha_3 \mapsto \alpha_1 + \alpha_2$ and $\alpha_4 \mapsto \alpha_1 - \alpha_2$.

For T , there is a basis $\beta = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ on which T acts by $\beta_1 \mapsto \beta_2, \beta_2 \mapsto \beta_4, \beta_3 \mapsto \beta_1 + \beta_2$ and $\beta_4 \mapsto \beta_1 - \beta_2$.

Show that $S \sim T$.

Solution: Need to find an invertible map $P : V \rightarrow V$ such that $P^{-1}SP = T$. Declare P by, the linear bijection $\alpha_i \leftrightarrow \beta_i$ between bases.

Now suppose we change the definition of T so that $\beta_2 \mapsto \beta_1 + \beta_4$. Are they still similar?

Solution: Certainly T is still similar to an operator T_2 whose description is $\alpha_1 \mapsto \alpha_2, \alpha_2 \mapsto \alpha_1 + \alpha_4, \alpha_3 \mapsto \alpha_1 + \alpha_2$ and $\alpha_4 \mapsto \alpha_1 - \alpha_2$. Since similarity is transitive, then it suffices to determine similarity of the following two matrices;

$$[S]_{\alpha} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, [T_2]_{\alpha} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

To check; determinants, characteristic polynomial, Jordan or rational canonical form.

Rank Nullity Theorem

Recall, for a linear map $T : \mathbb{F}^m \rightarrow \mathbb{F}^n$ that $\text{Rank } T + \dim \ker T = m$. In terms of decomposition, let $\{u_1, \dots, u_k\}$ be a basis for $\ker T$ (i.e. $k = \dim \ker T$) and let $\{u_1, \dots, u_k, u_{k+1}, \dots, u_m\}$ be an extension to a basis for $\mathbb{F}^m$. Then we can prove that $\{Tu_1, \dots, Tu_m\}$ spans the range of T and $\{Tu_{k+1}, \dots, Tu_m\}$ is independent. Hence, forming a basis for the range of T .

We can extend $v_1 = Tu_{k+1}, \dots, v_{m-k} = Tu_m$ to a basis for $\mathbb{F}^n$ so that $\mathbb{F}^n = \ker T \oplus \text{span}\{u_{k+1}, \dots, u_m\}$ and $\mathbb{F}^n = T\mathbb{F}^m \oplus \{v_{m-k+1}, \dots, v_n\}$ and

$$[T]_{\substack{\{v_{m-k}, \dots, v_1, v_n, \dots, v_{k+1}\} \\ \{u_m, u_{m-1}, \dots, u_{k+1}, \dots, u_1\}}} = \begin{pmatrix} I_k & 0_{m-k,k} \\ 0_{n-k,k} & 0_{m-k,n-k} \end{pmatrix}.$$

In terms of matrices, this asserts that there exists an invertible P and Q such that $P^{-1}[T]Q$ (where P and Q denote change of coordinate matrices between corresponding bases).

Differential equations

If D denotes the differential operator on an appropriate space of functions, find $\ker(D - 1)^2$.

Solution: To start, let $y = e^{\lambda t}z(t)$ so that

$$Dy = e^{\lambda t}Dz(t) + \lambda e^{\lambda t}z(t) = e^{\lambda t}z(t)(D + \lambda),$$

$$D^2y = \dots = e^{\lambda t}(D + \lambda)^2z(t)$$

and, in particular,

$$D^k y = e^{\lambda t}(D + \lambda)^k z(t).$$

Now, using this in the case $k = 2$ means

$$e^t D^2(e^{-t}z) = (D - 1)^2 z$$

so that $z \in \ker(D - 1)^2$ if and only if $e^t D^2(e^{-t}z) = 0$ which is equivalent to $D^2(e^{-t}z) = 0$ or $e^{-t}z = at + b$. In general, that is

$$\ker(D - \lambda)^n = \left\{ e^{\lambda t} \cdot \sum_{k=0}^n a_k t^k \right\}.$$

For a particular solution to, say, $(D^3 - 2D^2 + D)y = t^2 + 1$. We need,

$$\begin{aligned} y(t) &= \frac{t^2 + 1}{D^3 - 2D^2 + D} \\ &= \frac{t^2 + 1}{D(D - 1)^2} \\ &= \frac{1}{(D - 1)^2} \left[\frac{1}{D}(t^2 + 1) \right] \\ &= \frac{1}{(D - 1)^2} \left[\int (t^2 + 1) \right] \\ &= \frac{1}{(D - 1)^2} \left[\frac{x^3}{3} + x \right] \\ &= (1 + 2D + 3D^2 + 4D^3) \left[\frac{x^3}{3} + x \right] \\ &= \frac{x^3}{3} + 2x^2 + 7x + 10 \end{aligned}$$

Where we have computed the inverse $\frac{1}{(D-1)^2}$ by long division...

Jordan forms

Let A be 3×3 real matrix with characteristic polynomial $(-\lambda - 1)(\lambda - 2)^2$. What are the possible Jordan normal forms of A ?

Based on this information alone, the Jordan form of A is either

$$\mathcal{J}(\lambda - 1) \oplus \mathcal{J}((\lambda - 2)^2) = [1] \oplus \begin{pmatrix} C(\lambda - 2) & 1 \\ 0 & C(\lambda - 2) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

or

$$\mathcal{J}(\lambda - 1) \oplus \mathcal{J}(\lambda - 2) \oplus \mathcal{J}(\lambda - 2) = \text{diag}(1, 2, 2)$$