Structure of Linear Maps

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

Contents

1 Introduction

A basic course on Linear Algebra is an introduction to the preliminary notions such as vector spaces, linear dependence/independence, basis, linear maps, rank-nullity theorem, and orthonormal basis in an inner product spaces and orthogonal/unitary linear maps. The second phase of linear algebra is the study of structural results such as the decomposition of the vector space w.r.t. a linear map and investigating the possibility of representing the linear map in simple forms. In my article "Structure Theorems of Linear Maps", these topics were developed as a series of exercises with copious hints so as to reach the results as directly and as efficiently as possible. While I still prefer the original article, as it captures the central ideas which are not smothered by too many details, there was a constant need for the detailed treatment by a section of the students. The aim of this article is to give details of the earlier quoted article and I hope that this article serves the needs of the students.

Many of the details were supplied by a set of excellent notes on Sections 2–9 by S. Sundar. He was a student of B.Sc., and a participant of MTTS 2003, 2004 and 2005. He wrote the notes for a series of lectures given by me in MTTS 2005. This article is based to a large extent on his notes. I thank S. Sundar for the preliminary set of notes.

I also like to thank Professor M.I. Jinnah for his insightful comments on this article.

The key ideas involved in the proof of the existence theorems for canonical forms of $A\colon V \to V$ are the following:

(1) Expressing V as a direct sum of invariant subspaces.

(2) A trivial observation: If A and B commute, then $\ker B$ is invariant under A.

(3) Obvious B's that commute with A are given by $p(A)$ where p is a polynomial.

(4) If m_A is the minimal polynomial (that is, the unique monic polynomial p such that $p(A) = 0$), is written as $m_A(X) = p_1(X) \cdots p_k(X)$ where p_j's are relatively prime, then $V = \ker m_A(A) = \ker p_1(A) \oplus \cdots \oplus \ker p_k(A)$. 7/12/06

(5) If m_A splits, say, $m_A(X) = (X - \lambda_1)^{d_1} \cdots (X - \lambda_k)^{d_k}$, then $(A - \lambda_j I)$ acts nilpotently on $\ker(A - \lambda_j)^{d_j}$.

(6) Filippov's proof for the existence of Jordan canonical form can be simplified for nilpotent maps.

2 Warm up

We start with a warm-up. Let V and W be finite dimensional vector spaces over a field F. Let $n = \dim V$ and $m = \dim W$. Let $A\colon V \to W$ be linear.

We want to choose (ordered) bases of V and W in such a way that the matrix of A with respect to these bases is as simple as possible. We shall consider three special cases.

Case 1. Assume that A is onto.

Let $\{w_j : 1 \leq j \leq m\}$ be a basis of W. Since A is onto, we can find $v_j \in V$ such that

$Av_j = w_j$ for $1 \leq j \leq m$. We easily see that $\{v_j : 1 \leq j \leq m\}$ is linearly independent in V.

Reason: Let $c_j \in F$ be scalars such that $\sum_{j=1}^{n} c_j v_j = 0$. We than have

$$\begin{aligned} 0 = A(0) &= A(\sum_{j=1}^{n} c_j v_j) \\ &= \sum_j c_j A v_j \\ &= \sum_j c_j w_j. \end{aligned}$$

Since w_j are linearly independent, we conclude that $c_j = 0$, $1 \leq j \leq m$.

Let $\{u_k : 1 \leq k \leq r\}$ be a basis of $\ker A$. We claim that $\{v_1, \ldots, v_m, u_1, \ldots, u_r\}$ is a linearly independent subset of V.

Reason: Let $a_j, b_k \in F$, $1 \leq j \leq m$, $1 \leq k \leq r$ be scalars such that $\sum_{j=1}^{m} a_j v_j + \sum_{k=1}^{r} b_k u_k = 0$. As earlier, we operate A on both sides to get

$$\begin{aligned} 0 = A(0) &= \sum_{j=1}^{m} a_j A(v_j) + \sum_{k=1}^{r} b_k A(u_k) \\ &= \sum_{j=1}^{m} a_j w_j + \sum_{k=1}^{r} b_k 0, \end{aligned}$$

since $u_k \in \ker A$ for $1 \leq k \leq r$. By linear independence of w_j's we conclude that $a_j = 0$ for $1 \leq j \leq m$. Thus we are left with $\sum_{k=1}^{r} b_k u_k = 0$. Since by choice, u_k's form a basis of $\ker A$, we deduce that $b_k = 0$, $1 \leq k \leq r$. The claim is therefore established.

We now claim that the set $\{v_1, \ldots, v_m, u_1, \ldots, u_r\}$ is a basis of V.

Reason: The set under consideration is linearly independent. So, it is enough to show that the number $m + r$ of elements is $n = \dim V$. But this follows from Rank-Nullity theorem, as $m = \dim W = \dim AV = \operatorname{Im} A$ and $r = \dim \ker A$.

The matrix of A with respect to the ordered bases $\{v_1, \ldots, v_m, u_1, \ldots, u_r\}$ of V and $\{w_1, \ldots, w_m\}$ of W is $(I_{m \times m}, 0_{m \times (n-m)})$ where $I_{m \times m}$ is the identity matrix of size $m \times m$ and $0_{m \times (n-m)}$ is the zero matrix of size $m \times (n-m)$.

Reason: Recall the way the matrix representation is written. Let $T \colon V \to W$ be a linear map and $\{v_j : 1 \leq j \leq n\}$ (respectively, $\{w_k : 1 \leq k \leq m\}$) be an ordered basis of V (respectively, of W). Let $Tv_j = \sum_{k=1}^{m} a_{kj} w_k$. Then the k-th column of the matrix of T with respect to these ordered bases is $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$.

In our case, $Av_j = w_j$ for $1 \leq j \leq m$ so that the first m columns are 'basic column vectors' of size m. Similarly, the s-th column for $s > m$ is $Au_{s-m} = 0$ so that the last $r = n - m$ columns are zero (column) vectors of size m.

Case 2: A is one-one.

Let $\{v_j : 1 \leq j \leq n\}$ be a basis of V. It is easy to show that $\{Av_j : 1 \leq j \leq n\}$ is a linearly independent subset of W.

> Reason: For, if $\sum_{j=1}^n a_j Av_j = 0$, then we have $A\left(\sum_j a_j v_j\right) = 0$, that is, $\sum_j a_j v_j \in \ker A$. Since A is one-one, its kernel is (0) and hence we deduce that $\sum_j a_j v_j = 0$. Since v_j are linearly independent, it follows that $a_j = 0$ for $1 \leq j \leq n$.

Let $w_j = Av_j$, $1 \leq j \leq n$. We extend this linearly independent subset of W to a basis of W, say, $\{w_k : 1 \leq k \leq m\}$. We consider the ordered bases $\{v_j : 1 \leq j \leq n\}$ of V and $\{w_k : 1 \leq k \leq m\}$ of W. With respect to these bases, the matrix of A is $\begin{pmatrix} I_{n\times n} \\ 0_{(m-n)\times n} \end{pmatrix}$.

> Reason: Observe that $Av_j = w_j = 0w_1 + \cdots + 0w_{j-1} + 1.w_j + 0w_{j+1} + \cdots + 0w_n + 0w_{n+1} + \cdots$ so that the j-th column of the matrix of A is the 'standard basic vector' of size m: $(0, \ldots, 0, 1, 0, \ldots, 0)^t$, 1 at the j-th place, where $1 \leq j \leq n$.

Case 3: A is bijective.

In this case, if we start with an ordered basis $\{v_j : 1 \leq j \leq n\}$ of V and set $w_j := Av_j$, $1 \leq j \leq n$, then $\{w_j : 1 \leq j \leq n\}$ is a basis of W. The matrix of A with respect to these ordered bases is $I_{n\times n}$. (Verify!)

Remark 1. The above results are unsatisfactory. If $V = W$, then in each of the cases, we need two bases which need not be the same in order to arrive at a simple matrix representation of A.

Our aim: Given a finite dimensional vector space V over a field F and a linear map $A \colon V \to V$. Find an ordered basis of V so that the matrix of A with respect to this basis takes a simpler form.

3 Direct Sums, Invariant Subspaces and Block Matrices

Definition 2. Let V be a vector space over F. Let W_i, $1 \leq i \leq k$ be vector subspaces of V. We say that V is a *direct sum* of W_i's if the following holds:

1. For any $v \in V$, there exist $w_j \in W_j$, $1 \leq j \leq k$ such that $v = w_1 + \cdots + w_k$,
2. If $v = w_1 + \cdots + w_k = w_1' + \cdots + w_k'$ with $w_j, w_j' \in W_j$ for $1 \leq j \leq k$, then $w_j = w_j'$ for $1 \leq j \leq k$. (Note that this is equivalent to requiring that if $w_1 + \cdots + w_k = 0$ where $w_i \in W_i$, then $w_i = 0$ for $1 \leq i \leq k$.)

We then write $V = \oplus_{j=1}^n W_j$.

Example 3. Let V be a vector space and $B := \{v_i : 1 \leq i \leq n\}$ be a basis. Let $B = S \cup T$ be a partition of B into nonempty subsets. Let $W_1 := \operatorname{span} S$ and $W_2 = \operatorname{span} T$. Then it is easy to verify that $V = W_1 \oplus W_2$.

Example 4. Let $V := M(n, \mathbb{R})$ be the vector space of all square matrices of size n with entries in $\mathbb{R}$. Let W_s (respectively W_a) be the set of all symmetric (respectively skew-symmetric) matrices in V. Then $V = W_s \oplus W_a$. (Verify!)

Remark 5. If $V = W_1 \oplus \cdots \oplus W_k$, and if B_i is a basis of W_i, $1 \leq i \leq k$, then $B := \cup_{i=1}^n B_i$ is a basis of V. (Compare this with Example 3.)

Definition 6. Let $A\colon V \to V$ be linear. A vector subspace W of V is said to be A-invariant (or invariant under A) if $AW \subset W$, that is, $Aw \in W$ for any $w \in W$.

We most often say that W is invariant in place of A-invariant, if there is no possibility of confusion.

Example 7. Let $A = cI\colon V \to V$ for some $c \in F$. Then any subspace W is invariant. Is the converse true?

Example 8. Let the notation be as in Example 4. Consider the linear map $T\colon V \to V$ given by $TA = A + A^t$. Then W_s is invariant under T? Is W_a invariant under T?

Example 9. Let $V = W_1 \oplus W_2$. Let $P_i\colon V \to W_i$ be defined by $P_i(v) = v_i$ where $v = v_1 + v_2$, $i = 1, 2$. Note that P_i is well-defined. One easily shows that P_i is linear and $P_i^2 = P_i$. Also, W_1, W_2 are invariant subspaces of each of P_i.

Example 10. A scalar $\lambda \in F$ is said to be an *eigenvalue* of a linear map $A\colon V \to V$ if there exists a nonzero vector $v \in V$ such that $Av = \lambda v$. For an eigenvalue λ of A, let $V_\lambda := \{x \in V : Ax = \lambda x\}$. Then V_λ is called the eigenspace corresponding to the eigenvalue λ. It is easy to see that V_λ is an invariant vector subspace of A.

Elements of V_λ are called the eigenvectors of A corresponding to the eigenvalue λ.

Note that $V_\lambda = \ker(A - \lambda I)$.

> The key idea in finding out suitable basis of V so that the matrix of $A\colon V \to V$ takes a simple form is to express V as a direct sum of invariant subspaces.

We now explain this. Let $A\colon V \to V$ be a linear map. Assume that there exist invariant subspaces U and W such that $V = U \oplus W$. Let $\{u_1, \ldots, u_r\}$ be an (ordered) basis of U and $\{w_1, \ldots, w_s\}$ a basis of W. Then $\{u_1, \ldots, u_r, w_1, \ldots, w_s\}$ is an ordered basis of V. The matrix of A with respect to this basis is a block matrix of the form $\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}$. Here B is an $r \times r$ matrix and C is of type $s \times s$.

Reason: Let $v_i = u_i$ for $1 \leq i \leq r$ and $v_{r+j} = w_j$ for $1 \leq j \leq s$. Then, for $1 \leq i \leq r$, $Av_i \in U$ since U is A-invariant. Hence $Av_i = \sum_{i=1}^r b_{ji} u_j + \sum_{k=1}^s c_{r+k,k} w_k$ where $c_{r+k,k} = 0$ for $1 \leq k \leq s$. Hence the first r columns (of the matrix of A) will look like $(b_{1i}, \ldots, b_{ri}, 0, \ldots, 0)^t$. Similar considerations show that the last s columns will be of the form $(0, \ldots, 0, c_{r+1,1}, \ldots, c_{r+s,s})^t$.

Note that B (respectively, C) is the matrix of $A|_U$ (respectively $A|_W$) with respect to the basis $\{u_i\}$ (respectively $\{w_j\}$). 6/12/06

How do we generate an invariant subspace? Recall that if G is a group and $a \in G$ is an element, we ask for a subgroup containing a. This is too trivial, since $G \ni a$! So, we refine the question to ask for the smallest subgroup that contains a. (It exists! Why?) If H is any subgroup of G containing a, then $a^n \in H$ for all $n \in \mathbb{Z}$. But we observe that by law of indices, the subset $\{a^n : n \in \mathbb{Z}\}$ is already a subgroup. Hence the smallest subgroup containing a is $\{a^n : n \in \mathbb{Z}\}$. We may adopt this method to our invariant subspace problem too. If W is an invariant subspace containing v, then $A^k v \in V$ for all $k \in \mathbb{Z}_+$. Thus, we are led to consider the linear span of $W := \operatorname{span}\{A^k v : k \in \mathbb{Z}_+\}$. Clearly, W is the smallest invariant subspace containing v.

Since we assume that V is finite dimensional, it follows that the set $\{A^k v : k \in \mathbb{Z}_+\}$ is linearly dependent. Let r be the first integer such that $\{A^k v : 0 \leq k \leq r-1\}$ is linearly independent but not $\{A^k v : 0 \leq k \leq r\}$. If $V := \operatorname{span}\{A^k v : 0 \leq k \leq r-1\}$, so that $r = n$, then we have an ordered basis $v_i := A^{i-1}v$, $0 \leq i \leq n-1$. Assume that $\sum_{k=0}^{n} a_k A^k v = 0$ so that $a_n \neq 0$. Diving the equation by a_n, we may assume that $A^n v = -\sum_{k=0}^{n-1} a_k A^k v$. The matrix of A with respect to this basis is

$$\begin{pmatrix} 0 & \dots & 0 & -a_0 \\ 1 & \dots & 0 & -a_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & -a_{n-1} \end{pmatrix}.$$

Even if $r < n$, we still get a matrix of A restricted to the subspace W in a similar form. If only we could find an invariant subspace W_2 such that $V = W \oplus W_2$, we could write the matrix of A in a simpler form!

In general, it may not be possible. For instance, look at the linear map $A\colon \mathbb{C}^2 \to \mathbb{C}^2$ given by $Ae_1 = 0$ and $Ae_2 = e_1$. Then $W_1 = \operatorname{span}\{e_1\}$ is an invariant subspace (since it is V_0, the eigen subspace corresponding to the eigenvalue 0). We leave it to the reader to show that there exists no invariant subspace W_2 such that $V = W_1 \oplus W_2$.

Reason: Note that if such a subspace W_2 existed, it must of dimension 1. Hence, we may write it as $\mathbb{C}w_2 = (x, y)^t$. The second coordinate $y \neq 0$, since otherwise, $w_2 \in W_1$, a contradiction. But then $A(x, y)^t = A(xe_1 + ye_2) = xAe_1 + yAe_2 = 0 + ye_1 \in W_1$! This shows that W_2 cannot be an invariant subspace.

Definition 11. A linear map $A\colon V \to V$ is said to be *semi-simple* if for any invariant subspace W, there exists an invariant subspace U such that $V = U \oplus W$.

Such a subspace U is called an invariant complement of W.

Thus semi-simple linear maps are most amenable/adapted to our strategy. We shall return to this theme a little later.

A *very useful observation* is the following

Lemma 12. *Let $A, B\colon V \to V$ be linear maps such that $AB = BA$. Then $\ker A$ is invariant under B.*

Proof. Let $K := \ker A$ and $v \in K$. We want to show that $Bv \in K$. This is equivalent to showing that $A(Bv) = 0$. Since A and B commute, we have $A(Bv) = B(Av) = B0 = 0$ and hence the result. $\square$

In particular, under the above assumption, if V_λ is an eigen space of A, then V_λ is B-invariant. Now, how do we find B's that commute with A? Note that all (non-negative) powers A^k of A commute with A and so are their linear combinations. Thus any polynomial in A (with coefficients in F) will commute with A. So, if $p(X)$ is a polynomial over F, then we set

$$p(A) := c_0 I + c_1 A + \cdots + c_m A^m, \text{ where } p(X) = c_0 + c_1 X + \cdots + c_m X^m.$$

Ex. 13. Let $p(X), q(X) \in F[X]$, then pq denotes the multiplication of p and q. It is easy to check that

$$pq(A) = p(A) \circ q(A) = q(A) \circ p(A).$$

Hence in particular, $\ker p(A)$ is invariant under $q(A)$.

Theorem 14. *Let $p(x) \in F[X]$. Assume that $p(X) = p_1(X) \cdots p_k(X)$ where $p_1, \ldots, p_k$ are relatively prime. Let $A\colon V \to V$ be linear.*

(i) *We have*

$$\ker p(A) = \ker p_1(A) \oplus \cdots \oplus \ker p_k(A).$$

Thus $\ker p(A)$ is a direct sum of A-invariant subspaces $p_i(A)$.

(ii) *The natural projections $\pi_i\colon \ker p(A) \to \ker p_i(A)$ is a polynomial in A.*

(iii) *If $W \subset \ker p(A)$ is an A-invariant subspace, then we have*

$$W = \oplus_{i=1}^k \left(W \cap \ker p_i(A)\right).$$

Proof. For $1 \le i \le k$, let $q_i := \prod_{j \ne i} p_j = p/p_i$. Since each p_i is a factor of p, it is clear that $\ker p_i(A) \subset \ker p(A)$ so that $\sum_{i=1}^k \ker p_i(A) \subset \ker p(A)$. To prove the reverse inclusion, we observe that q_i are relatively prime, since p_i are.

> Reason: For if f is a common divisor of q_i's, then f must divide one of the factors, since p_i are relatively prime. Let us assume that f divides p_{j_i} a factor of q_i. Not all of the p_{j_i} $1 \le i \le k$, could be equal since each q_i 'misses' p_i. Thus f is a common divisor of two distinct p_i's, a contradiction.

Hence there exist polynomials f_i such that $f_1 q_1 + \cdots + f_k q_k = 1$. Hence we have $I = f_1(A) q_1(A) + \cdots + f_k(A) q_k(A)$. For any $v \in \ker p(A)$, we have

$$v = I(v) = v_1 + \cdots + v_k, \text{ where } v_i = f_i(A) q_i(A) v, 1 \le i \le k.$$

We claim that $v_i \in \ker p_i(A)$. For,

$$p_i(A) v_i = p_i(A) \circ f_i(A) \circ q_i(A) v = f_i(A) p_i(A) q_i(A) = f_i(A) p(A) v = 0,$$

since $v \in \ker p(A)$. Thus we have shown that $\ker p(A) = \sum_{i=1}^k \ker p_i(A)$.

We now show that the sum is direct. Let $v_1 + \cdots + v_k = 0$ where $v_i \in \ker p_i(A)$. Now, $q_i(A) v_j = 0$ for $j \ne i$. Since p_i and q_i are relatively prime, there exist polynomials f, g such that $f p_i + g q_i = 1$. Hence $f(A) p_i(A) + g(A) q_i(A) = I$ and so,

$$\begin{aligned} v_i = I(v_i) &= f(A)(p_i(A) v_i) + g(A)(q_i(A) v_i) \\ &= f(A)(0) + g(A)(q_i(A)(-\sum_{j \ne i} v_j)) \\ &= 0 + g(A)(0) = 0. \end{aligned}$$

Thus each $v_i = 0$. We conclude that the sum is direct. This completes the proof of (i).

Proof of (ii) is culled from that of (i): $\pi_i = f_i(A)q_i(A)$.

Note that we have a resolution of the identity: $I = \sum_i \pi_i$

To prove (iii), let $w \in W$ and let $w = v_1 + \cdots + v_k$ as in (i). Then $v_i = \pi_i(w) = f_i(A)q_i(A)(w)$. Now $w \in W$ is A-invariant and hence $v_i = f_i(A)q_i(A)w \in W$. Thus each $v_i \in W \cap V_i$. This proves (iii). $\square$

Thus, if we could find a polynomial p such that $\ker p(A) = V$, then we would have expressed V as a direct sum of invariant subspaces. This is certainly possible, at least theoretically. For $\{A^k : k \in \mathbb{Z}_+\}$ is an infinite set in the n^2 dimensional vector space of all linear maps from V to itself. Hence there exists a polynomial $p(X) \in F(x)$ such that $p(A) = 0$. Let $m_A(X)$ be the polynomial of least degree with 1 as the coefficient of the highest degree term. (Why does this make sense?)

Definition 15. The minimal polynomial of $A \colon V \to V$ is the unique monic polynomial (that is, a polynomial in which the coefficient of the highest degree nonzero term is 1) $m_A(X)$ of least degree such that $m_A(A) = 0$.

Since there exist polynomials $p(X) \in F[X]$ such that $p(A) = 0$, the set $\{\deg p : p(X) \in F[X], p(A) = 0\}$ is a nonempty subset of $\mathbb{Z}_+$. Hence there is a least integer in it. Let $p(X) \in F[X]$ be such that $p(A) = 0$ and $\deg p$ is minimal. If we write $p(X) = \sum_{i=0}^k a_i X^i$, then $a_k \neq 0$. So, we may divide p by a_k and assume that the coefficient of X^k is 1. We denote this polynomial by $m_A(X)$.

Now if $q(X) \in F[X]$ is any other polynomial such that $q(A) = 0$, then we claim that m_A divides q.

> Reason: Note that $\deg q \geq \deg m_A$. By division algorithm, we write $q(X) = f(X)m_A(X) + r(X)$ where $0 \leq \deg r < \deg m_A$. If $\deg r > 0$, then we get a contradiction since $r(A) = q(A) - f(A)m_A(A) = 0$ and $\deg r(A) < \deg m_A$. Hence the remainder r is a constant. It has to be zero, since otherwise,
>
> $$0 = q(A) = f(A)m_A(A) + r \cdot I = r \cdot I \neq 0,$$
>
> a contradiction.

In particular, m_A is unique subject to the conditions:
(1) $m_A(A) = 0$, (ii) $\deg m_A \leq \deg p$ for any polynomial p with $p(A) = 0$ and (iii) The coefficient of the top degree (called the leading coefficient) term is 1.

> Reason: If $p \in F[X]$ is any such polynomial, then $p = am_A(X)$. Since the leading coefficient of p is 1, we deduce that $a = 1$.

Ex. 16. Any eigenvalue $\lambda \in F$ of $A \colon V \to V$ is a root of the minimal polynomial of A.

> If short of time, it is suggested that the reader may directly go to Theorem 46 and prove it using Theorem 14.

4 Eigenvalues and Diagonalizable operators

Before we go any further, we analyze the existence of eigen values and eigenvectors. There are two problems with eigenvalues and eigenvectors

Example 17. There may not exist any eigenvalue. For instance, consider the rotation about the origin by $\pi/2$ in the plane $\mathbb{R}^2$. Algebraically, it is given by $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ by setting $A(x,y) = (-y,x)$. Since no line through the origin is invariant, A has no eigenvector. To prove this algebraically, we note that $A^2 = -I$. So, if $\lambda \in \mathbb{R}$ is an eigenvalue, say, with a nonzero eigenvector (x,y), we then get the two equations:

$$\begin{aligned} A^2(x,y) &= A(-y,x) = (-x,-y) = -1(x,y) \\ A^2(x,y) &= A(\lambda(x,y)) = \lambda^2(x,y). \end{aligned}$$

We therefore deduce λ is a real number such that $\lambda^2 = -1$! We thus conclude there exists no eigenvalue.

Example 18. There may not exist 'enough' eigenvectors. What we mean by this is that the set of eigenvectors may not span the given space.

For instance, consider $A\colon \mathbb{C}^2 \to \mathbb{C}^2$ defined by $Ae_1 = 0$ and $Ae_2 = e_1$. Since $A^2e_1 = A0 = 0$ and $A^2e_2 = Ae_1 = 0$, we see that $A^2 = 0$. If $\lambda \in \mathbb{C}$ is an eigenvalue of A and $v \in \mathbb{C}^2$ is a nonzero eigenvector, then $0 = A^2v = \lambda^2 v$ and hence we conclude that $\lambda = 0$. Thus the only eigenvalue of A is zero. Clearly, e_1 is an eigenvector. If $\mathbb{C}^2$ is to be span of eigenvectors of A, then there exits another eigenvector $v = (x,y)^t$ linearly independent of e_1. Hence $y \neq 0$. But, $0 = A(x,y)^t = A(xe_1 + ye_2) = x0 + ye_1$ so that $y = 0$, a contradiction. Thus any eigenvector of A is a scalar multiple of e_1.

We attend to each of these problems now.

Theorem 19. *Let $A\colon V \to V$ be a linear map. Assume that F is algebraically closed, that is, any nonconstant polynomial splits into linear factors. Then A has an eigenvalue.*

Proof. As observed earlier, there exists a polynomial $p(X) = \sum_{k=0}^{d} a_k X^k$ of degree $d \leq n^2$ such that $p(A) = 0$. We may also assume that $a_d \neq 0$. Since F is algebraically closed, there exist $\lambda_1, \ldots, \lambda_d$ such that $p(X) = (X - \lambda_1)\cdots(X - \lambda_d)$. Since $p(A) = 0$, if we fix a nonzero $v \in V$, we have $(A - \lambda_1 I)\cdots(A - \lambda_d I)v = 0$. Let

$$S := \{k : 1 \leq k \leq d \text{ and } (A - \lambda_k)\cdots(A - \lambda_d I)v = 0\}.$$

Clearly, $1 \in S$. Let $r \in S$ be the largest. Then we have

$$\begin{aligned} (A - \lambda_{r+1} I)\cdots(A - \lambda_d I)v &\neq 0 \\ (A - \lambda_r I)\cdots(A - \lambda_d I)v &= 0. \end{aligned}$$

By letting $w := (A - \lambda_{r+1} I)\cdots(A - \lambda_d I)v$, we see that $(A - \lambda_r I)w = 0$. In other words, λ_r is an eigenvalue with w as an eigenvector. $\square$

Remark 20. The above result remains true if we assume that m_A has a linear factor in F. We leave it to the reader to convince himself of this.

Lemma 21. *Let $A\colon V \to V$ be linear. Then nonzero eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Let λ_i, $1 \leq i \leq k$, be distinct eigenvalues of A. Assume that v_i is a nonzero eigenvector with eigenvalue λ_i for $1 \leq i \leq k$. Let $\sum_{i=1}^k a_i v_i = 0$. We claim that each $a_i = 0$. If not, consider $T := (A - \lambda_2 I) \cdots (A - \lambda_k I)$. Then $Tv_j = 0$ for $j \geq 2$. For,

$$Tv_j = (A - \lambda_2 I) \cdots (A - \lambda_{j-1} I)(A - \lambda_{j+1} I) \cdots (A - \lambda_k I)(A - \lambda_j I)v_j = 0.$$

Hence

$$0 = T\left(\sum_{i=1}^k a_i v_i\right) = T(a_1 v_1) = a_1(\lambda_1 - \lambda_2) \cdots (\lambda_1 - \lambda_k)v_1 = 0.$$

Since $\lambda_1 - \lambda_j \neq 0$ for $j > 1$ and $v_1 \neq 0$, we conclude that $a_1 = 0$. Similarly, one shows that $a_i = 0$ for $1 \leq i \leq k$.

If you like a more formal proof, we may proceed as follows. Let $\sum a_i v_i = 0$. Let, if possible, r be the largest integer such that $a_r \neq 0$. Thus, we have $\sum_{i=1}^r a_i v_i = 0$. Consider $T := (A - \lambda_1 I) \cdots (A - \lambda_{r-1} I)$. Then $Tv_i = 0$ $1 \leq i \leq r-1$ and $Tv_r = (\lambda_r - \lambda_1) \cdots (\lambda_r - \lambda_{r-1})$. Applying T to both sides of the equation $\sum_{i=1}^r a_i v_i = 0$, we get $a_r(\lambda_r - \lambda_1) \cdots (\lambda_r - \lambda_{r-1})v_r = 0$. Since $(\lambda_r - \lambda_i) \neq 0$ for $i \neq r$ and $v_r \neq 0$, we conclude that $a_r = 0$. This is a contradiction to our choice of r. This shows that no such r exists or what is the same $a_i = 0$ for $1 \leq i \leq k$. $\square$

Corollary 22. *Let $A\colon V \to V$ be a linear map on an n-dimensional vector space over F. Then A has at most n distinct eigen values.* $\square$

Let F be arbitrary and $A\colon V \to V$. The eigenspaces of A span V iff A is diagonalizable.

Definition 23. We say that a linear map $A\colon V \to V$ is *diagonalizable* if there exists a basis of V with respect to which the matrix of A is diagonal.

The following result, albeit easy, offers a most important sufficient condition for a linear map to be diagonalizable.

Proposition 24. *Let $A\colon V \to V$ has $n = \dim V$ distinct eigenvalues. Then A is diagonalizable.*

Proof. Let λ_k, $1 \leq k \leq n$, be the distinct eigenvalues of A. Then, by very definition, there exist nonzero vectors v_k such that $Av_k = \lambda_k v_k$, $1 \leq k \leq n$. The set $\{v_1, \ldots, v_n\}$ is linearly independent ad hence is a basis of V. Clearly, the matrix of A with respect to this basis is $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. $\square$

Theorem 25. *Let $A\colon V \to V$ be linear. Then the following are equivalent.*

(i) *A is diagonalizable, that is, there exists a basis of V with respect to which the matrix of A is a diagonal matrix.*

(ii) *There exists an A-eigen basis of V, that is, a basis of V consisting of eigenvectors of A.*

(iii) *V is the sum of eigen subspaces of A, that is, $V = \sum_\lambda V_\lambda$ where λ runs through the distinct eigenvalues of A.*

Proof. It is clear that (i) $\iff$ (ii). Also, (ii) $\implies$ (iii) trivial. To show that (iii) $\implies$ (ii), we need only observe that the sum $\sum_\lambda V_\lambda$ is a direct sum by Lemma 21. Now we select a basis of V_λ for each eigenvalue λ of A. Their union is the required eigen-basis. $\square$

Theorem 26. *Let $A\colon V \to V$ be diagonalizable and $W \subset V$ be invariant under A. Then*

(1) *If $V = \oplus V_\lambda$ is the direct sum decomposition into eigenspaces of A, then we have $W = \oplus(W \cap V_\lambda)$.*

(2) *There is an A-invariant complement of W. (In other words, any diagonalizable linear map is semisimple.)*

Proof. Let $w \in W$. Write $w = \sum v_\lambda$ according to the direct sum decomposition $V = \oplus V_\lambda$. We need to show that $v_\lambda \in W$. Since W is invariant under A, it is invariant under $p(A)$ for any polynomial $p(X) \in F[X]$. In particular, $p(A)w = \sum p(\lambda)v_\lambda \in W$ for any polynomial p. Let $\lambda_1, \dots, \lambda_k$ be those eigenvalues for which $v_\lambda \neq 0$. For $1 \leq i \leq k$, consider the polynomial $p_i(X) = \prod_{j\neq i}(X - \lambda_j)$. Then the element $p_i(A)w = \sum p_i(\lambda_j)v_{\lambda_j} = p_i(\lambda_i)v_{\lambda_i} \in W$. Since $p_i(\lambda_i) \neq 0$, we deduce that $v_{\lambda_i} \in W$. As this holds true for all $1 \leq i \leq k$, it follows that $v_\lambda \in W$. This proves (1).

To prove (2), let W'_λ be any complement of $(W \cap V_\lambda)$ in V_λ. (Why does this exist?) Then $W' := \sum W'_\lambda$ is an A-invariant complement of W.

> Reason: Since $A = \lambda I$ on V_λ, by Example 7, any vector subspace of V_λ, in particular, W'_λ is invariant under A. Let $W_\lambda = V_\lambda \cap W$. Then, we have,
>
> $$V = \oplus V_\lambda = \oplus\left(W_\lambda \oplus W'_\lambda\right) = (\oplus W_\lambda) \oplus (\oplus W'_\lambda) = W \oplus W'.$$

$\square$

Theorem 27. *Let $A, B\colon V \to V$ be a pair of commuting diagonalizable linear maps. Then they are simultaneously diagonalizable, that, there exists an ordered basis of V with respect to which each of A, B is represented by a diagonal matrix.*

Proof. Let $V = \oplus V_\lambda(A)$ be the eigenspace decomposition of A and $V = \oplus V_\mu(B)$, that of B. Fix $\lambda \in F$ and $v \in V_\lambda(A)$. We write $v = \sum_\mu v_\mu$ according to the eigenspace decomposition of B. We then have

$$\sum \lambda v_\mu = \lambda v = Av_\lambda = \sum Av_\mu.$$

Since the sums are direct and since the space $V_\mu(B)$ are invariant under B, we conclude that $Av_\mu = \lambda v_\mu$ for all μ. Consequently, we see that $V_\lambda = \oplus_\mu\left(V_\lambda(A) \cap V_\mu(B)\right)$ and hence

$$V = \oplus_{\lambda,\mu}\left(V_\lambda(A) \cap V_\mu(B)\right). \tag{1}$$

If we now choose a basis for each of the nonzero summands $V_\lambda(A) \cap V_\mu(B)$ then each of the vectors in it would be an eigenvector for both A and B. Putting all these together will yield a required basis of V.

One may also argue as follows. Since A and B commute and since $V_\lambda = \ker(A - \lambda I)$, the space V_λ is invariant under B and $V = \oplus_\mu V_\mu$. The result follows now from Theorem 26 (1). $\square$

Remark 28. The result above can be extended to a family of pairwise commuting diagonalizable linear maps. Start with A, B two members of the family, arrive at (1). Pick up an element C of the family and argue as in the proof but with C and the decomposition (1). Due to finite dimensionality, the procedure has to stop at a decomposition of the form $V = \oplus_{i=1}^r V_i$ such that each member of the family acts as a scalar on V_i.

For a more formal proof, argue by induction on the dimension of V.

Ex. 29. Let $A, B\colon V \to V$ be two commuting linear maps. If A and B are diagonalizable, so is $A + B$.

5 Nilpotent Operators

Definition 30. $A\colon V \to V$ is said to be nilpotent if there exists $k \in \mathbb{N}$ such that $A^k = 0$.

Remark 31. Just an idea! Let V be an n dimensional vector space. Suppose that we want to show that $k \leq n$, then we look for a set of k linearly independent elements in V.

Lemma 32. *Let $A\colon V \to V$ be a nilpotent linear map on an n-dimensional vector space V. Then there exists $k \in \mathbb{N}$ such that $A^k = 0$ and $k \leq n$.*

In particular, if A is nilpotent, then $A^n = 0$.

Proof. Since A is nilpotent, there exists $k \in \mathbb{N}$ such that $A^k = 0$. Let m be the smallest positive integer such that $A^m = 0$. Therefore, $A^{m-1} \neq 0$ and hence there exists $v \in V$ such that $A^{m-1}v \neq 0$. We claim that $\{A^k v : 0 \leq k \leq m-1\}$ is linearly independent. If not, there exist $a_k \in F$, $0 \leq k \leq m-1$ such that $\sum_{k=0}^{m-1} a_k A^k v = 0$. Applying A^{m-1} to both sides of this equation, we get

$$\sum_{k=0}^{m-1} a_k A^{k+m-1} v = 0.$$

Since $k + m - 1 \geq m - 1$ for $k \geq 1$, we see that $A^{k+m-1}v = 0$ for $k \geq 1$. Hence the only summand that remains is the term corresponding to $k = 0$. Thus we get $a_0 A^{m-1} v = 0$. Since $A^{m-1}v \neq 0$, we conclude that $a_0 = 0$. We now apply A^{m-2} to conclude that $a_1 = 0$ and so on.

If you want to see a more formal proof, here it is. Let r be the least integer such that $a_r \neq 0$ so that $\sum_{k=r}^{m-1} a_k A^k v = 0$. Applying A^{m-1-k} to both sides of the equation, we get $a_k A^{m-1} v = 0$. Now one proceeds as earlier.

Thus the m vectors $A^k v$, $0 \leq k \leq m-1$ are linearly independent in the n-dimensional vector space V. Hence $m \leq n$. □

Ex. 33. If $A, B\colon V \to V$ are two commuting nilpotent linear maps, then $A + B$ is also nilpotent.

Ex. 34. Let $A\colon V \to V$ be both diagonalizable and nilpotent. Show that $A = 0$.

Proposition 35. *Let $A\colon V \to V$ be nilpotent. Then*

(i) *0 is an eigenvalue of A.*

(ii) *If λ is an eigenvalue of A, then $\lambda = 0$.*

Proof. If $A = 0$, there is nothing to prove. So, we may assume that $A \neq 0$. Let k be the least positive integer such that $A^k = 0$. Then $A^{k-1} \neq 0$. There exists $v \in V$ such that $w := A^{k-1}v \neq 0$. (Necessarily, $v \neq 0$!) Now, we have $Aw = 0$. Since $w \neq 0$, it is an eigenvector with eigenvalue 0. This proves (i).

Let λ be an eigenvalue of A. Let $v \in V$ be such that $Av = \lambda v$. By induction, we see that $A^m v = \lambda^m v$ for $m \in \mathbb{N}$. Since $A^k = 0$ for some $k \in \mathbb{N}$, we see that $0 = A^k v = \lambda^k v$. As $v \neq 0$, we conclude that $\lambda^k = 0$ and hence $\lambda = 0$. This proves (ii). □

Given a nilpotent operator A on V, we can choose a basis of V so that the matrix A is strictly upper triangular.

Definition 36. Let $A = (a_{ij})$ be an $n \times n$ matrix over a field F. The matrix A is said to be *strictly upper triangular* if $a_{ij} = 0$ for $i \geq j$.

Proposition 37. *Let $A\colon V \to V$ be such that there exists a basis of V with respect to which the matrix of A is strictly upper triangular. Then A is nilpotent.*

Proof. This is a straightforward exercise and we urge the reader to prove it on his own.

Let $B := \{v_1, \ldots, v_n\}$ be an ordered basis w.r.t. which the matrix (a_{ij}) of A is strictly upper triangular. Since $Av_i \in \operatorname{span}\{v_j : 1 \leq j \leq i\}$, for $1 \leq i \leq n$, it follows by induction that $A^{i-1}v_i \in Fv_1$ so that $A^i v_i = 0$. Hence we conclude that $A^n v_i = A^{n-i}(A^i v_i) = 0$ for $1 \leq i \leq n$. Consequently, if $v = \sum_{i=1}^n a_i v_i$, then $A^n v = \sum_i a_i A^n v_i = 0$. In other words, A is nilpotent. □ 6/12/06

Proposition 38. *Let $A\colon V \to V$ be linear. Let $\{v_i : 1 \leq i \leq n\}$ be a basis of V with respect to which the matrix of A is strictly upper triangular. Then if we set $V_0 = \{0\}$ and $V_k := \operatorname{span}\{v_i : 1 \leq i \leq k\}$, then we have the following:*
(i) $V_i \subset V_{i+1}$ *for* $0 \leq i \leq n-1$.
(ii) $V_i \neq V_{i+1}$ *for* $0 \leq i \leq n-1$.
(iii $V_0 = \{0\}$ *and* $V_n = V$.
(iv) $AV_i \subseteq V_{i-1}$ *for* $1 \leq i \leq n$.

Proof. This is again straight forward verification and hence the reader should carry out the proofs on his own.

The statements (i)–(iii) are obvious. To prove (iv), we observe that, $Av_1 = 0$ and that for $j \geq 2$

$$Av_j = \sum_{i=1}^{n} a_{ij} v_i = \sum_{i=1}^{j-1} a_{ij} v_i, \text{ since } a_{ij} = 0 \text{ for } i \geq j.$$

It follows that $Av_j \in V_{j-1}$ for $j \geq 1$. If $v \in V_i$, then $v = \sum_{j=1}^{i} a_j v_j$ so that $Av = \sum_{j=1}^{i} a_j A v_j \in V_{i-1}$. □

The finite sequence (V_i) of subspaces is called a *flag*.

The next proposition is a converse of the last one.

Proposition 39. *Let $A: V \to V$ be a linear map on a vector space V over a field of dimension n. Assume that there exists $k \in \mathbb{N}$ and for each i, $0 \le i \le k$, there exists a subspace V_i of V with the following properties:*

(i) $V_i \subset V_{i+1}$ *for* $0 \le i \le k-1$.

(ii) $V_i \neq V_{i+1}$ *for* $0 \le i \le k-1$.

(iii $V_0 = \{0\}$ *and* $V_n = V$.

(iv) $AV_i \subseteq V_{i-1}$ *for* $1 \le i \le k$.

Then there exists a basis for V such that the matrix of A with respect to this basis is strictly upper triangular.

Proof. We can prove this result by induction on n or k. Let us do induction on k for fun!

When $k = 1$, we have $V_0 = \{0\}$ and $V_1 = V$. Since $AV_1 \subset V_0$, we infer that A is the zero operator. Hence any basis of V will do!

Assume that $k \ge 2$ and that the result is true for any linear map $A: V \to V$ as long as the size of the flag is less than k. Let us consider $A: V \to V$ with a flag of size k. Since V_{k-1} is mapped to V_{k-2} by A, the restriction B of A to V_{k-1} has a flag of size $k-1$. Hence there exists a basis $\{v_1, \dots, v_m\}$ of V_{k-1} such that the matrix of B with respect to this basis is strictly upper triangular. Extend this to a basis $\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$ of V. Since $Av_r \in V_{k-1}$ for $r \ge m+1$, we see that the r-th column of the matrix of A is $(a_{1r}, \dots, a_{mr}, 0, \dots, 0)^t$ for $m+1 \le r \le n$. It is now easy to see that the matrix of A is strictly upper triangular. □

Theorem 40. *Let $A: V \to V$ be nilpotent. Then there exists a basis of V with respect to which the matrix of A is strictly upper triangular.*

Proof. Let $k \in \mathbb{N}$ be the least such that $A^k = 0$. Let $V_i := \ker A^i$ for $0 \le i \le k$. Then, the following are obvious:

(i) $V_i \subseteq V_{i+1}$ for $0 \le i \le k-1$.

(ii) $V_0 = \{0\}$ and $V_k = V$.

(iii) $AV_i \subseteq V_{i-1}$ for $1 \le i \le k$.

We now show that V_i is strictly contained in V_{i+1} for $0 \le i \le k-1$. Since $A^{k-1} \neq 0$, there exists $v \in V$ such that $A^{k-1}v \neq 0$. Now the vector $A^{k-1-i} \in V_{i+1}$ but not in V_i for $0 \le i \le k-1$.

> Reason: $A^{i+1}(A^{k-1-i}v) = A^k v = 0$ so that $A^{k-1-i}v \in V_{i+1}$. However, $A^i(A^{k-1-i}v) = A^{k-1}v \neq 0$ so that $A^{k-1-i}v \notin V_i$.

Now the result follows from Proposition 39. □

6 Generalized eigenvectors

Now that we have seen (Theorem 25) that V is a direct sum of eigenspaces of A iff A is diagonalizable, we explore the possibility of extending the concept of eigenvectors. Let us look at the earlier example of $A: \mathbb{C}^2 \to \mathbb{C}^2$ given by $Ae_1 = 0$ and $Ae_2 = e_1$. Even though the only eigenvector of A is e_1, the vector e_2 has the property $A^2 e_2 = 0 \cdot e_2$. Together they form a basis of $\mathbb{C}^2$ with respect to which the map is represented by an upper triangular matrix.

Definition 41. Let $A\colon V \to V$ be linear. Assume that $\lambda \in F$ is an eigenvalue of A. A vector $v \in V$ is said to be a *generalized eigenvector* of A corresponding to the (genuine!) eigenvalue λ if there exists $k \in \mathbb{N}$ such that $(A - \lambda I)^k v = 0$.

The generalized eigensubspace corresponding to the eigenvalue λ is defined by

$$V(\lambda) := \{v \in V : \text{ There exists } k \in \mathbb{N} \text{ such that } (A - \lambda I)^k v = 0\}.$$

Remark 42. Any eigenvector with eigenvalue λ is a generalized eigenvector for λ. In the example preceding the definition, e_2 is a generalized eigenvector which is not an eigenvector corresponding to the eigenvalue 0.

Proposition 43. *Let $A\colon V \to V$ be linear. The the nonzero generalized eigenvectors corresponding to distinct eigenvalues of A are linearly independent.*

Proof. Let λ_j, $1 \leq j \leq m$ be distinct eigenvalues of A. Let $v_j \in V(\lambda_j)$, $1 \leq j \leq m$, be nonzero vectors. Let $\sum_{j=1}^m a_j v_j = 0$ for $a_j \in F$. Let r be the least integer such that $a_r \neq 0$. If $r = m$, then $a_m v_m = 0$ implies that a_m is also zero. Hence all a_j's are zero. So, we may assume that $r < m$.

Let k be the least positive integer such that $(A - \lambda_r)^k v_r = 0$. Note that this means that $(A - \lambda_r)^{k-1} v_r$ is an eigenvector of A with eigenvalue λ_r. We now argue as in Lemma 21.

Since $(A - \lambda_j I)^n v_j = 0$ for $1 \leq j \leq m$ and

$$(A - \lambda_r I)^{k-1} (A - \lambda_{r+1} I)^n \cdots (A - \lambda_m I)^n a_i v_i = 0, \text{ for } i \neq r$$

it follows that

$$a_r (\lambda_r - \lambda_{r+1})^n \cdots (\lambda_r - \lambda_m)^n (A - \lambda_r I)^{k-1} v_r = 0.$$

It follows that $a_r (A - \lambda_r I)^{k-1} v_r = 0$. We conclude that $a_r = 0$. □

Theorem 44. *Let $A\colon V \to V$ be linear and $\lambda \in F$ be an eigenvalue of A. Then the following hold:*

(i) $V(\lambda) = \ker(A - \lambda I)^n$.
(i) *$A - \lambda I$ is nilpotent on $V(\lambda)$.*
(iii) *$V(\lambda)$ is invariant under A.*
(iv) *λ is the only eigenvalue of A restricted to $V(\lambda)$.*

Proof. Let $v \in V(\lambda)$ be nonzero. Let m be the least (positive) integer such that $(A - \lambda I)^m v = 0$. We claim that the set

$$\{(A - \lambda I)^j : 0 \leq j \leq m - 1\}$$

is linearly independent. Let $\sum_{i=0}^{m-1} a_i (A - \lambda I)^i v = 0$. Let us operate $(a - \lambda I)^{m-1}$ on both sides of this equation to get

$$a_0 (A - \lambda I)^{m-1} v + a_1 (A - \lambda I)^m v + \cdots + a_{m-1} (A - \lambda I)^{2m-2} v = 0.$$

Each of the j-th term is zero since we have $(A - \lambda I)^r v$ with $r \geq m$. Thus the equation above becomes $a_0 (A - \lambda I)^{m-1} v = 0$. Since $(A - \lambda I)^{m-1} v \neq 0$, we deduce that $a_0 = 0$. Thus the original equation reduces to $a_1 (A - \lambda I)^m v + \cdots + a_{m-1} (A - \lambda I)^{2m-2} v = 0$ We now apply $(A - \lambda I)^{m-2}$ to both sides and argue as above to conclude that$a_1 = 0$ and so on. Thus all $a_i = 0$ and hence the claim.

A more formal proof: Given that $\sum_{i=0}^{m-1} a_i(A-\lambda I)^i v = 0$, let k be the least index such that $a_k \neq 0$. Therefore we get $A^{m-1-k}\left(\sum_{i=k}^{m-1} a_i(A-\lambda I)^i v\right) = 0$, that is,

$$\sum_{i=k}^{m-1} a_i(A-\lambda I)^{m-1-k+i} v = 0.$$

For $i \geq k+1$, we have $m-1-k+i \geq m$ so that $(A-\lambda I)^{m-1-k+i}v = 0$. Thus the equation reduces to $a_m(A-\lambda I)^{m-1}v = 0$. Since $(A-\lambda I)^{m-1}v \neq 0$, we are forced to conclude that $a_m = 0$, a contradiction to our choice of m.

It follows that $k \leq n$ and hence $(A-\lambda I)^n v = 0$. Thus $V(\lambda) = \ker(A-\lambda I)^n$. This proves (i) as well as (ii). Since $V(\lambda) = \ker(A-\lambda I)^n$ and since A commutes with $(A-\lambda I)^n$, the subspace $V(\lambda)$ is invariant under A. This proves (iii).

Let $\mu \in F$ be an eigenvalue of restriction of A to $V(\lambda)$. Let $0 \neq v \in V(\lambda)$ be such that $Av = \mu v$. We then have

$$(A-\lambda I)^n v = (\mu - \lambda)^n v = 0.$$

Since $v \neq 0$,w e deduce that $\mu = \lambda$. This proves (iv). □

Theorem 45. *Let V be a finite dimensional vector space over a field F. Let $A\colon V \to V$ be a linear map. Assume that the minimal polynomial of A splits over F. Let λ_i, $1 \leq i \leq m$ be the distinct roots of the minimal polynomial of A. Then we have*
(i) *$V = \oplus_{i=1}^m V(\lambda_i)$.*
(ii) *If $\mu \in F$ is an eigen value of A, then $\mu = \lambda_i$ for some $i \in \{1, \ldots, m\}$.*

Proof. We prove (i) by induction on n.

If $n = 1$, then there exists $\lambda \in F$ such that $Av = \lambda v$ for all $v \in V$. Thus, when $n = 1$, $V = \ker(A - \lambda I) = V(\lambda)$ and hence (i) is true in this case.

Let us assume the result to be true for any linear map whose minimal polynomial splits over F on any vector space of dimension less than n. Let $n > 1$.

Let V be an n-dimensional vector space. Assume that $A\colon V \to V$ is a linear map whose minimal polynomial splits over F. Then thanks to Theorem 19, there exists an eigenvalue $\lambda \in F$ of A.

We claim that $V = \ker(A-\lambda I)^n \oplus \operatorname{Im}(A-\lambda I)^n$.

Let $w \in V(\lambda) \cap W$ where we have set $W := \operatorname{Im}(A-\lambda I)^n$. Since $w \in W$, there exists $u \in V$ such that $(A-\lambda I)^n u = w$. Since $w \in \ker(A-\lambda I)^n$, we have

$$0 = (A-\lambda I)^n w = (A-\lambda I)^{2n} u.$$

This means that $u \in V(\lambda)$ and hence $(A-\lambda I)^n u = 0$. But then it follows that $w = (A-\lambda I)^n u = 0$. Thus $V(\lambda) \cap W = \{0\}$. Also, by the rank-nullity theorem, $\dim V = \dim\ker(A-\lambda I)^n + \dim\operatorname{Im}(A-\lambda I)^n = \dim V(\lambda) + \dim W$. Thus we see that $V(\lambda) + W = V$. Therefore the claim is established.

If W is the zero subspace, then $V = V(\lambda)$ and (i) is proved. If not, then $\dim W < \dim V$ (since $\dim V(\lambda) \geq \dim V_\lambda \geq 1$). Since $W = \operatorname{Im}(A-\lambda I)$, it is invariant under A. For, if

$w = (A - \lambda I)^n u$ for some $u \in V$, then

$$Aw = A\,(A - \lambda I)^n\, u = (A - \lambda I)^n Au \in \operatorname{Im}(A - \lambda I)^n.$$

Now, if B denotes the restriction of A to W, then $B\colon W \to W$ with $\dim W < \dim V$. In order to invoke the induction hypothesis, we need to check if the minimal polynomial of B splits over F. But this is trivially so, since the minimal polynomial of B must be a divisor of that of A.

Reason: Let m_A be the minimal polynomial of A on V and m_B that of B. Now, $m_A(B)w = m_A(A)w = 0$ for any $w \in W$. Thus m_A is a polynomial in $F[X]$ such that $m_A(B) = 0$. Since m_B is the minimal polynomial of B, it follows (by division algorithm) that m_B divides m_A. Since m_A splits over F, so does m_B.

Therefore the induction hypothesis is applicable to $B\colon W \to W$. We therefore infer that there exist scalars, say, $\lambda_2, \dots, \lambda_k \in F$ such that $W = \oplus_{i=2}^k W(\lambda_i)$. The following are fairly obvious:

(i) $\lambda_1 \neq \lambda_i$ for $2 \leq i \leq k$.

Reason: For, if not, assume that $\lambda_1 = \lambda_j$ for some $2 \leq j \leq k$. There exists $0 \neq w \in W$ such that $Bw = \lambda_j w$. Since $Bw = Aw$, this means that $(A - \lambda 1)w = 0$ or $w \in \ker(A - \lambda_1)^n$. Since $w \in \ker(A - \lambda I)^n \cap W = \{0\}$, we see that $w = 0$, a contradiction.

(ii) $W(\lambda_j) = V(\lambda_j)$ for $2 \leq j \leq k$. Clearly, $W(\lambda_j) \subset V(\lambda_j)$. Let $v \in V(\lambda_j)$. We write $v = v_1 + w$ with $v_1 \in V(\lambda_1)$ and $w \in W$. Then $w = w_2 + \cdots + w_k$ with $w_i \in W(\lambda_i) \subset V(\lambda_i)$, $2 \leq i \leq k$. Therefore we get

$$v_1 + \cdots + w_{j-1} + (v - w_j) + w_{j+1} + \cdots + w_k = 0.$$

By Proposition 43, it follows that each of the summands is zero, in particular, $v = w_j \in W(\lambda_j)$.

An alternate proof is given below.

Reason: Clearly, $W(\lambda_j) \subset V(\lambda_j)$. Let $v \in V(\lambda_j)$. We write $v = v_1 + w$ with $v_1 \in V(\lambda_1)$ and $w \in W$. We have,

$$0 = (A - \lambda_j I)^n v = (A - \lambda_j I)^n v_1 + (A - \lambda_j I)^n w.$$

Since $V(\lambda_1)$ and W are invariant under A, we see that $(A - \lambda_j I)^n v_1 \in V(\lambda_1)$ and $(A - \lambda_j I)^n w \in W$. Since the sum $V = V(\lambda_1) \oplus W$ is direct,we conclude that $(A - \lambda_j I)^n v_1 = 0$ and $(A - \lambda_j I)^n w = 0$. We claim that $v_1 = 0$.

Reason: Suppose not. Let $r \in \mathbb{N}$ be the least such that $(A - \lambda_j I)^r v_1 = 0$. If we let $v_2 := (A - \lambda_j I)^{r-1} v_1$, then $v_2 \neq 0$. Since $(A - \lambda_j I)v_2 = (A - \lambda_j I)^r v_1 = 0$, the vector v_2 is an eigenvector of A with eigenvalue λ_j. This contradicts the fact that λ_1 is the only eigenvalue of the restriction of A to $V(\lambda_1)$ (see (iv) of Theorem 44). Hence the claim is proved.

The claim follows also from the linear independence of the nonzero generalized eigenvectors corresponding to distinct eigenvalues (Proposition 43).

Therefore, $v \in W$ and hence lies in $W(\lambda_j)$. Consequently, we have shown that $V(\lambda_j) \subset W(\lambda_j)$ for any $2 \le j \le k$. Thus we have established

$$V = V(\lambda_1) \oplus W = \oplus_{j=1}^k V(\lambda_j).$$

This completes the proof of (i).

We now prove (ii). Let $\mu \in F$ be an eigenvalue of A with a nonzero eigenvector u. Let $u = \sum_j v_j$, $v_j \in V(\lambda_j)$, $1 \le j \le k$. Since $u \neq 0$, there exists r such that $v_r \neq 0$. We have

$$0 = (A - \mu I)u = \sum_j (A - \mu I)v_j.$$

Since $(A - \mu I)$ commutes with $(A - \lambda_j I)$ for all j, it leaves $V(\lambda_j)$ invariant. Hence $(A - \lambda_j I)v_j \in V(\lambda_j)$ for all j. Since $V = \oplus V(\lambda_j)$ is a direct sum, we deduce that each of the summands $(A - \mu I)v_j$ in the displayed equation above must be zero. In particular, $(A - \mu I)v_r = 0$, that is, $v_r \in V(\lambda_r)$ is an eigenvector with eigenvalue μ. By Theorem 44-(iv), we conclude that $\mu = \lambda_r$. $\square$

Putting Theorems 44–45 together, we get

Theorem 46 (Structure Theorem for Linear Maps)**.** *Let F be an algebraically closed field and V a finite dimensional vector space over F. Let $A\colon V \to V$ be linear. Assume that $\lambda_j \in F$, $1 \le j \le k$ be all the distinct eigenvalues of A. Then the following are true.*

Give a proof using Thm 14.

(i) $V = \oplus_{j=1}^k V(\lambda_j)$.
(ii) $V(\lambda_j)$, $1 \le j \le k$, *is A-invariant.*
(iii) *The map $(A - \lambda_j I)$ is nilpotent on $V(\lambda_j)$, $1 \le j \le k$.*
(iv) *λ_j is the only eigenvalue of A on $V(\lambda_j)$, $1 \le j \le k$.* $\square$

Proposition 47. *Let V be a finite dimensional vector space over an algebraically closed field F. Let $A\colon V \to V$ be linear. Assume that 0 is the only eigenvalue of A. Then A is nilpotent.*

Proof. By structure theorem, we have $V = V(0)$ and $A = A - 0 \cdot I$ is nilpotent on $V(0)$. $\square$

Remark 48. The above result is true as long as the minimal polynomial of A splits over F, otherwise it is false. For instance, consider the linear map $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ defined by $Ae_1 = e_2$, $Ae_2 = -e_1$ and $Ae_3 = 0$. It is easy to see that 0 is the only eigenvalue of A.

Reason: Proceed as in Examples 17-18.

Since $A^3 e_1 = -e_2$, we see that A is not nilpotent. $\square$

Proposition 49. *Let the notation be as in the structure theorem (Theorem 46). Let m_i be the least positive integer such that $(A - \lambda_i)^{m_i} = 0$ on $V(\lambda_i)$. Then the minimal polynomial of A is given by*

$$m_A(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_k)^{m_k}.$$

Proof. We observe that if λ is a root of m_A with multiplicity r, then $V(\lambda) = \ker(A - \lambda I)^r$.

Reason: Suppose not. Let $v \in V(\lambda)$ be such that $w := (A - \lambda I)^r v \neq 0$. If we let $m_A(X) = q(X)(X - \lambda)^r$, then q and $(X - \lambda)^{n-r}$ are relatively prime. So, there exist polynomials f and g such that $1 = q(X)f(X) + (X - \lambda)^{n-r}g(X)$. We then have

$$\begin{aligned} w &= f(A)q(A)w + g(A)(A - \lambda I)^{n-r}w \\ &= f(A)m_A(A)v + g(A)(A - \lambda I)^n v \\ &= 0 + 0 = 0, \end{aligned}$$

a contradiction. Thus the claim is proved.

The result in an immediate consequence of this claim. For, given $1 \leq k \leq n$, the multiplicity of λ_k as a root of $m_A(X)$ is the same as m_k, according to the claim. $\square$

7 Jordan Canonical Form

Definition 50. Let $A \colon V \to V$ be linear and $\lambda \in F$ be an eigenvalue of A. We say that a finite sequence $v_1, \ldots, v_k$ of nonzero vectors is a *Jordan string* or a *Jordan chain* corresponding to the eigenvalue λif the following holds:

$$\begin{aligned} Av_1 &= \lambda v_1 \\ Av_2 &= v_1 + \lambda v_2 \\ &\vdots \\ Av_k &= v_{k-1} + \lambda v_k. \end{aligned}$$

The integer k is called the length of the Jordan string.

The following are immediate from the definition.

(i) An eigenvector v with eigenvalue λ is a Jordan string of length 1.

(ii) Each vector v_j of the Jordan string lies in the generalized eigenspace $J(\lambda)$. In fact, $(A - \lambda I)^j v_j = 0$.

(iii) The set of vectors in a Jordan string is linearly independent.

Reason: Let $v_1, \ldots, v_k$ be a Jordan string corresponding to the eigenvalue $\lambda \in F$. Assume that $\sum_{i=1}^k a_i v_i = 0$. Let r be the largest integer such that $a_r \neq 0$. Then $r > 1$.

Reason: For, if $r = 1$, then the linear dependence equation above becomes $a_1 v_1 = 0$. Since $v_1 \neq 0$, we are led to conclude $a_1 = 0$, a contradiction.

Then $v_r = \sum_{i=1}^{r-1} -a_r^{-1} a_i v_i$. Let us operate both sides of the equation by $(A - \lambda I)^{r-1}$. We get the following contradiction:

$$v_1 = (A - \lambda I)^{r-1} v_r = \sum_{i=1}^{r-1} -a_r^{-1} a_i (A - \lambda I)^{r-1} v_i = 0,$$

where we have used the observation (ii).

(iv) The length of any Jordan string is at most $n = \dim V$.

(v) If we let $W := \operatorname{span}\{v_1, \ldots, v_k\}$, then W is A-invariant and the matrix of the restriction of A to W is

$$\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & & \\ & & \ddots & & 1 \\ & & & & \lambda \end{pmatrix}.$$

This matrix is called a Jordan block of size k corresponding to $\lambda \in F$.

We now give a proof of the existence of Jordan canonical form for a nilpotent linear map following a method of Filippov. We then make use of the structure theorem to derive the result in the case of a general linear map. It should be noted that Filippov's method yields the general case also without recourse to the structure theorem. See my article on "Jordan Canonical Form" for this approach.

Theorem 51 (Jordan canonical form for nilpotent operators)**.** *Let $A\colon V \to V$ be a nilpotent operator on a finite dimensional vector space over a field F. Assume that the all the roots of the characteristic polynomial lie in F. Then there exists an A-Jordan basis of V.* Where is Jordan basis defined?

Proof. We prove this result by induction on the dimension of V. If $\dim V = 1$, then $A = 0$ and hence any nonzero element is a Jordan basis of V. The result is also true if $A = 0$ and whatever be the dimension of V. So, we now assume that the result is true for all nonzero nilpotent operators on any finite dimensional vector space with dimension less than n where $n > 1$.

Let V be of dimension n and $A\colon V \to V$ be nonzero and nilpotent. Since $\ker A \neq \{0\}$, $\dim \operatorname{Im} A < n$. It is also invariant under A. Thus the restriction of A to $W = \operatorname{Im} A$, which we denote by A again, is a nilpotent operator on W. We can therefore apply the induction hypothesis. We then get a Jordan basis of W, say, $J = J_1 \cup \cdots \cup J_k$ where each J_i is a Jordan string:

$$J_i = \{v_{i1}, \ldots, v_{in_i}\} \text{ with } Av_{i1} = 0 \text{ and } Av_{ij} = v_{ij-1} \text{ for } 2 \leq j \leq n_i.$$

We have, of course, $n_1 + \cdots + n_k = \dim \operatorname{Im} A$.

Suggestion: The reader may assume that there is only one Jordan string during the first reading of the proof below. He may like to understand the proof in a special case, say, $A\colon \mathbb{R}^5 \to \mathbb{R}^5$ given by 6/12/06

$$Ae_1 = 0 = Ae_2, Ae_3 = e_4, Ae_4 = e_5, Ae_5 = 0.$$

By very assumption that J is a basis of $\operatorname{Im} A$, the set $\{v_{i1} : 1 \leq i \leq k\}$ (of the first elements of the Jordan strings J_i) is a linearly independent subset of V and it is a subset of $\ker A$. We extend this set to a basis of $\ker A$, say, $\{v_{11}, \ldots, v_{k1}, z_1, \ldots, z_r\}$. Each last element $v_{in_i} \in J_i$ lies in $\operatorname{Im} A$ and hence we can find $v_{in_i+1} \in V$ such that $Av_{in_i+1} = v_{in_i}$. We now let $B_i := J_i \cup \{v_{in_i+1}\}$ and $B := \cup_{i=1}^k B_i \cup \{z_1, \ldots, z_r\}$. Using rank-nullity theorem, we see that $|B| = n$. We claim that B is linearly independent subset of V. Let

$$(a_{11}v_{11} + \cdots + a_{1n_1+1}v_{1n_1+1}) + \cdots + (a_{k1}v_{k1} + \cdots + a_{1n_k+1}v_{1n_k+1}) + b_1z_1 + \cdots + b_rz_r = 0. \quad (2)$$

We apply A to both sides. Since $z_j, v_{i1} \in \ker A$ for $1 \leq j \leq r$ and $1 \leq i \leq k$, we arrive at the following equation:

$$A\left([a_{12}v_{12} + \cdots + a_{1n_1+1}v_{1n_1+1}] + \cdots + [a_{k2}v_{k2} + \cdots + a_{kn_k+1}v_{1n_k+1}]\right) = 0.$$

Since $Av_{ij} = v_{ij-1}$ for $1 \leq i \leq k$ and $2 \leq j \leq n_i + 1$, we get

$$(a_{12}v_{11} + \cdots + a_{1n_1+1}v_{1n_1}) + \cdots + (a_{k2}v_{k1} + \cdots + a_{kn_k+1}v_{1n_k}) = 0.$$

Since v_{ij}'s that appear in the above equation are linearly independent, we deduce that $a_{ij} = 0$ for $1 \leq i \leq k$ and $2 \leq j \leq n_i + 1$. Thus (2) becomes

$$a_{11}v_{11} + \cdots + a_{k1}v_{k1} + b_1 z_1 + \cdots + b_r z_r = 0.$$

The vectors that appear in the equation above form a basis of $\ker A$ and hence we deduce all the coefficients in the equation are zero. Thus, we have show that all the coefficients in (2) are zero and hence B is linearly independent. □

8 Characteristic Polynomial

Definition 52. The polynomial $\det(A - XI)$ in X is called the *characteristic polynomial* of A. We shall denote it by $p_A(X)$.

Ex. 53. What is the characteristic polynomial of (i) a diagonalizable linear map and (ii) which can be represented as a triangular matrix with respect to some basis? Added on 7/12/06

Proposition 54. *Let $A\colon V \to V$ be linear. Assume that the minimal polynomial of A splits into linear factors, say, $m_A(X) = \prod_{i=1}^k (X - \lambda_i)^{m_i}$ so that $V = \oplus_{i=1}^k V(\lambda_i)$ be the direct sum decomposition of V into generalized eigenspaces. Let $n_i := \dim V(\lambda_i)$. Then $p_A(X) = (X - \lambda_1)^{n_1} \cdots (X - \lambda_k)^{n_k}$.*

In particular, $p_A(A) = 0$. (Cayley-Hamilton Theorem)

Proof. Using the standard notation, for any $\alpha \in F$, the eigen values of $\alpha I - A$ are $\alpha - \lambda_j$ with multiplicities n_j, $1 \leq j \leq k$. Hence the determinant of $\alpha I - A$ is the product of its eigenvalues, that is, $(\alpha - \lambda_1)^{n_1} \cdots (\alpha - \lambda_k)^{n_k}$. Since this true for all $\alpha \in F$, we infer that $\det(XI - A) = (X - \lambda_1)^{n_1} \cdots (X - \lambda_k)^{n_k}$.

Since $m_A(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_k)^{m_k}$, it follows that m_A divides p_A and in particular, $p_A(A) = 0$. □

If $\lambda \in F$ is an eigenvalue of $A\colon V \to V$, then $\dim V_\lambda$ is called the *geometric multiplicity* of λ and $\dim V(\lambda)$ is called the *algebraic multiplicity* of λ. The latter is thanks to the fact that the multiplicity of λ as a root of the characteristic polynomial is $\dim V(\lambda)$.

Example 55. Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be the nilpotent map $Ae_i = e_{i+1}$, $1 \leq i \leq n$ and $Ae_n = 0$. Then the characteristic polynomial (which is also the minimal polynomial) is $X^n = 0$. The eigenvalue $\lambda = 0$ has 1 as its geometric multiplicity while n as its algebraic multiplicity. Added on 7/12/06

The following gives a characterization of diagonalizability of A in terms of its characteristic and minimal polynomials.

Theorem 56. *Let $A\colon V \to V$ be linear. Then the following are equivalent.*

(i) *A is diagonalizable.*

(ii) *The characteristic polynomial of A can be written as a product of linear factors in F, say, $(X-\lambda_1)^{n_1}\cdots(X-\lambda_k)^{n_k}$ with $n_i = \dim V_{\lambda_i} = \dim\ker(A-\lambda_i)$.* (One says that the characteristic polynomial splits over F and the algebraic multiplicity of any eigenvalue is its geometric multiplicity.)

(iii) *If the distinct eigenvalues of A are $\lambda_1,\dots,\lambda_k$, then the minimal polynomial of A is $(X-\lambda_1)\cdots(X-\lambda_k)$.*

Proof. (i) $\implies$ (ii): Easy.

(ii) $\implies$ (iii): Since the characteristic polynomial $p_A(X)$ splits over F, we can write it as

$$p_A(X) = (X-\lambda_1)^{n_1}\cdots(X-\lambda_k)^{n_k}.$$

Since the algebraic and geometric multiplicities are equal, for each $1 \le i \le k$, there exist n_i linearly independent eigenvectors with eigenvalue λ_i. Hence the operator $(A-\lambda_1 I)\cdots(A-\lambda_k I)$ kills all the vectors in V. Hence the minimal polynomial of A must be a divisor of $(X-\lambda_1)\cdots(X-\lambda_k)$. No term such as $(X-\lambda_i)$ can be absent in the minimal polynomial of A. For example,

$$(A-\lambda_1 I)\cdots(A-\lambda_{i-1}I)(A-\lambda_{i+1}I)\cdots(A-\lambda_k I)$$

(note the absence of the term $(A-\lambda_i I)$ in the product above) cannot kill V_{λ_i}.

(iii) $\implies$ (i): Follows from the structure theorem. $\square$

Suppose that the minimal polynomial of A splits over F but we cannot say anything about the geometric and algebraic multiplicities. Then the next result (Theorem 58) deals with this.

Definition 57. A *flag* in an n-dimensional vector space V is a sequence $(V_i)_{i=0}^n$ of vector subspaces such that $\dim V_i = i$ for $0 \le i \le n$ and $V_i \subset V_{i+1}$ for $0 \le i \le n-1$.

We say that a linear map $A\colon V \to V$ *stabilizes* the flag if $AV_i \subset V_i$ for $0 \le i \le n$,

Theorem 58. *Let $A\colon V \to V$ be a linear map on an n-dimensional vector space over F. Then the following conditions are equivalent.*

(1) *$V = \sum V(\lambda)$, that is, V is the sum of generalized eigenspaces of A.*

(2) *There is a Jordan basis of V for A with respect to which the matrix of A is of the form diag$(J_{n_1}(\lambda_1),\dots,J_{n_k}(\lambda_k))$.*

(3) *There is a basis of V with respect to which the matrix of A is upper triangular.*

(4) *A stabilizes a flag of V.*

(5) *A has n eigenvalues in F (counted with multiplicity).*

(6) *The characteristic polynomial of A splits into linear factors in F.*

Proof. Note that the sum in (1) is, in fact, a direct sum. Since $(A-\lambda_i I)$ is nilpotent on $V(\lambda_i)$, there exists a basis of $V(\lambda_i)$ with respect to which the matrix of $A-\lambda_i I$ (restricted to $V(\lambda_i)$) is $J_{n_i}(0)$. Hence the matrix of the restriction A to $V(\lambda_i)$ is $J_{n_i}(\lambda_i)$. Putting all these bases together yields a basis of V with respect to which the matrix of A is as stated. Thus (1) $\implies$ (2).

The implications (2) $\implies$ (3) $\iff$ (4) $\implies$ (5) $\implies$ (6) are either obvious or easy consequences of some of our earlier results.

If the characteristic polynomial of A splits into linear factors, say,

$$p_A(X) = (X - \lambda_1)^{n_1} \cdots (X - \lambda_k)^{n_k},$$

then by Theorem 14, we have

$$V = \ker(A - \lambda_1 I)^{n_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{n_k} = V(\lambda_1) \oplus \cdots \oplus V(\lambda_k).$$

$\square$

Theorem 59. *Let $A\colon V \to V$ have n eigenvalues (counted with multiplicity) in F. Then there exists a unique decomposition $A = A_D + A_N$ where A_D is diagonalizable and A_N is nilpotent. Moreover, A_D and A_N are polynomials in A and hence they commute with each other.*

The decomposition $A = A_D + A_N$ is called the (additive) Jordan decomposition of A.

Proof. The idea is quite simple. If we choose a Jordan basis of V for A and if the matrix of A with respect to this basis is $\operatorname{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_k))$, then A_D is the map corresponding to the diagonal part of this matrix and A_N is the one corresponding to $\operatorname{diag}(J_{n_1}(0), \ldots, J_{n_k}(0))$.

In abstract terms, $A_D := \lambda_1\pi_1 + \cdots + \lambda_k\pi_k$ where $\pi_i\colon V \to V(\lambda_i)$ is the canonical projection. As observed in (ii) of Theorem 14, the maps π are polynomials in A and hence so is A_D. Hence the same is true about $A_N := A - A_D$. Also, we have

$$A_N = A - A_D = A \circ \sum_i \pi_i - \sum_i \lambda_i \pi_i = \sum_{i=1}^{k} (A - \lambda_i I)\pi_i.$$

As a consequence, we conclude that $A_N = (A - \lambda_i I)$ on $V(\lambda_i)$ and hence is nilpotent on $V(\lambda_i)$. Therefore, A_N is nilpotent on V.

We now prove the uniqueness. Let $A = D + N$ be another such decomposition with D diagonalizable etc. Then

$$NA = N(D + N) = ND + N^2 = DN + N^2 = (D + N)S = AN.$$

Similarly,we show that A and D commute with each other. Since A_D and A_N are polynomials in A, they also commute with D and N. Therefore, $A_D - D$ is diagonalizable (Theorem 27, or more precisely, Ex. 29) and $A_N - N$ is nilpotent. We see that $A_D - D = N - A_N$ is both diagonalizable and nilpotent and hence is the zero operator (Ex. 34). $\square$

Remark 60. Note that in the proof above, we assumed that all the roots of the minimal polynomial m_A lie in F. This is the case when F is algebraically closed. However, Cayley-Hamilton theorem is true over any field. The general version could be deduced from this special case which we shall not go into.

Ex. 61. Let $A\colon V \to V$ be linear. Assume that the minimal polynomial $m_A(X) \in F[X]$ is irreducible. Show that A is semisimple. *Hint:* If we set $K := \{f(A) : f \in F[X]\}$, then K is a field. Consider V as a vector space over K in an obvious way. The K subspaces of the K-vector space V are precisely the A-invariant subspaces of the F-vector space V!

Use this to give an example of a semisimple map which is not diagonalizable.

Remark 62. The following are some of the important features of the Jordan canonical form of a linear map and they are very useful in determining the Jordan canonical form.

(i) The sum of the sizes of the blocks involving a fixed eigenvalue equals the algebraic multiplicity of the eigenvalue, that is, the multiplicity of the the eigenvalue as a root of the characteristic polynomial.
(ii) The number of blocks involving an eigenvalue equals its geometric multiplicity, that is, the dimension of the corresponding eigenspace .
(iii) The largest block involving an eigenvalue equals the multiplicity of the eigenvalue as a root of the minimal polynomial.

Let J be a Jordan canonical form of A. Then A and J are similar. Hence their characteristic polynomials are the same. Statement (i) follows if we observe that the eigenvalues of a Jordan block $J_j(\lambda)$ is λ with algebraic multiplicity k.

Statement (ii) follows from the observation that the eigenvalue λ of similar matrices (or linear maps) have the same geometric multiplicity and the fact that any Jordan block $J_k(\lambda)$ has one dimensional eigenspace.

Statement (iii) follows from the observations: (a) the map $T := J_k(\lambda) - \lambda I_{k\times k}$ is nilpotent with index k, that is, $T^k = 0$ but $T^{k-1} \neq 0$ and (b) if $J = \operatorname{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k))$, then its minimal polynomial is the product of the minimal polynomials of $J_{n_i}(\lambda_i)$.

Theorem 63 (Uniqueness of the Jordan Form)**.** *The Jordan form is unique apart from a permutation of the Jordan blocks.*

Proof. Let us assume that A is similar to two Jordan forms J_1 and J_2. Then there is some eigenvalue λ of A such that the corresponding blocks in J_1 and J_2 differ. As observed in the above remark (Property (ii), more precisely), the number of blocks corresponding to λ in J_1 and J_2 will be the geometric multiplicity , say, k of λ. Let $m_1 \geq m_2 \geq \cdots \geq m_k$ be the sizes of the blocks of J_1 corresponding to the eigenvalue λ. Let $m_1 \geq m_2 \geq \cdots \geq m_k$ be the sizes of the blocks in J_2. It follows that there exists some $1 \leq j \leq k$ such that $m - i = n_i$ for all $1 \leq i \leq j-1$ but $m_j \neq n_j$. Assume without loss of generality that $n_j > m_j$. But then $(J_1 - \lambda I)^{m_j} = 0$ but $(J_2 - \lambda I)^{m_j} \neq 0$. This is absurd since J_1 and J_2 are similar. □

9 Similarity

We assume that the field F is algebraically closed. We say that two matrices $A, B \in M(n, F)$ are similar if there exists an invertible matrix T such that $TAT^{-1} = B$. Another way of looking at this is via group actions. Let $G := GL(n, F)$ denote the group of invertible matrices in $M(n, F)$. Then G acts on $M(n, F)$ via conjugation: $GL(n, F) \times M(n, F) \to M(n, F)$ given by $(T, A) \mapsto TAT^{-1}$. Then A and B are similar iff they lie in the same orbit.

Jordan canonical form gives a distinguished representative of the orbit of A under this action. In particular, A and B are similar iff they have the 'same' Jordan canonical form but for the permutation of the Jordan blocks. We leave the details for the reader to ponder upon!

10 Exercises

The following are some of the standard exercises based on the material of this article.

Ex. 64. The characteristic polynomial of A is $(X-1)^3(X-2)^2$ and its minimal polynomial is $(X-1)^2(X-2)$. What is its Jordan form?

Ex. 65. The characteristic polynomial of A is $(X-1)^3(X-2)^2$. Write down all possible Jordan forms of A.

Ex. 66. Find all possible Jordan forms of an 8×8 matrix whose characteristic polynomial is $(X-1)^4(X-2)^4$ and the minimal polynomial $(X-1)^2(X-2)^2$ if the geometric multiplicity of the eigenvalue $\lambda = 1$ is three.

Ex. 67. Show that any square matrix A is similar to its transpose. *Hint:* If A is similar to J what is A^T similar to?

Ex. 68. Show that there is no $A \in M(3,\mathbb{R})$ whose minimal polynomial is x^2+1 but there is $B \in M(2,\mathbb{R})$ as well as $C \in M(3,\mathbb{C})$ whose minimal polynomial is X^2+1.

Ex. 69. Let $A\colon V \to V$ be such that $A^2 = A$. Discuss whether or not there exists an eigen-basis of V.

Ex. 70. Let $A^{k+1} = A$ for some $k \in \mathbb{N}$. Show that A is diagonalizable. *Hint:* Observe that

$$(J_n(\lambda))^{k+1} = (\operatorname{diag}_n(\lambda) + J_n(0))^{k+1}.$$

Ex. 71. Let $V = U \oplus W$. Let $P_W\colon V \to W$ be the canonical projection and $R_w\colon V \to V$ be the reflection with respect to W: $R_W(w+u) = w-u$. Compute the minimal polynomials of P_W and R_W.

Ex. 72. Let $A\colon V \to V$ be of rank 1. Then $AV = Fv_0$ for some $v_0 \in V$. Show that $A^2 = \lambda A$ where $Av_0 = \lambda v_0$.

Does there exists an eigen-basis of V?

Ex. 73. Are the following matrices diagonalizable? (a) $J_n(\lambda)$, (b) a nilpotent matrix and (c) $A \in M(n,\mathbb{C})$ such that $A^k = I$ for some $k \in \mathbb{N}$.

Ex. 74. Let $A \in M(3,\mathbb{C})$. Assume that the characteristic and minimal polynomials of A are known. Show that there exists only one possible Jordan form. Is it still true of we replace $\mathbb{C}$ by $\mathbb{R}$ or if we replace 3 by 4?

Ex. 75. Consider the two matrices

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Show that their characteristic polynomial is $(X-1)^4$ and the minimal polynomial is $(X-1)^2$, but they do not have the same Jordan form. (Question: What are the Jordan forms of the given matrices?) Thus for two matrices to be similar it is necessary but not sufficient that they have the same characteristic and the same minimal polynomial.

Ex. 76. Show that if $A \in M(n, \mathbb{C})$ is such that $A^n = I$, then A is a diagonalizable.

Ex. 77. Prove that if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A and if $p(X)$ is a polynomial, then $p(\lambda_i)$, $1 \leq i \leq n$, are the eigenvalues of $p(A)$.

Ex. 78. If $A := \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$ show that $A^{50} = 2^{50} \begin{pmatrix} -24 & 25 \\ -25 & 26 \end{pmatrix}$.

Ex. 79. What are all the possible canonical forms of matrices in $M(2, \mathbb{C})$? It is a good exercise to arrive at this directly with "bare hands".

One should also do some numerical examples such as finding the Jordan canonical form of a few matrices. I refer the reader to my article on "Jordan Canonical Form' for examples and exercises of this kind.

11 Inner Product Spaces

Let V be an inner product space over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Let $\operatorname{End} V$ denote the set of all linear maps from V to itself.

Definition 80. Let $A\colon V \to V$ be linear. Then the map $A^*\colon V \to V$ is defined by the equation

$$\langle y, A^*x\rangle = \langle Ay, x\rangle \text{ for all } x, y \in V.$$

Reason: Fix $x \in V$. Consider the map $f_x\colon y \mapsto \langle Ay, x\rangle$. It is linear. Hence (by Riesz representation theorem), there exists a unique vector $v \in V$ such that $f_x(y) = \langle y, v\rangle$. We let A^*x stand for this vector v.

We claim that $A^*\colon V \to V$ given by $x \mapsto A^*x$ is linear.

Reason: Fix $x, y \in V$. For $z \in V$, we have

$$\begin{aligned} \langle z, A^*(x+y)\rangle &= \langle Az, x+y\rangle \\ &= \langle Az, x\rangle + \langle Az, y\rangle \\ &= \langle z, A^*x\rangle + \langle z, A^*y\rangle . \end{aligned}$$

Hence, for all $z \in V$, we have

$$\langle z, A^*(x+y) - A^*x - A^*y\rangle = 0.$$

Taking $z = A^*(x+y) - A^*x - A^*y$, we find that

$$\langle A^*(x+y) - A*x - A^y, A^*(x+y) - A*x - A^y\rangle = 0.$$

We conclude that $A^*(x+y) - A^*x - A^*y = 0$.

Similarly, if $\lambda \in \mathbb{F}$, then for all $z \in V$,

$$\langle z, A^*(\lambda x)\rangle = \langle Az, \lambda x\rangle = \overline{\lambda}\,\langle Az, x\rangle = \overline{\lambda}\,\langle z, A^*x\rangle = \langle z, \lambda A^*x\rangle .$$

As earlier, we conclude that $A^*(\lambda x) = \lambda A^*(x)$.

Note that the proof above remains valid even if $\mathbb{F} = \mathbb{R}$.

The map A^* is called the adjoint of A.

Lemma 81. *The map $A \mapsto A^*$ from $\operatorname{End} V$ to itself has the following properties.*
(i) $(A^*)^* = A$ *for any* $A \in \operatorname{End} V$.
(ii) $(A+B)^* = A^* + B^*$ *for any* $A, B \in \operatorname{End} V$.
(iii) $(\lambda A)^* = \overline{\lambda} A^*$ *for any* $A \in \operatorname{End} V$ *and* $\lambda \in \mathbb{F}$.
(iv) $(AB)^* = B^*A^*$ *for any* $A, B \in \operatorname{End} V$.

Proof. The proofs are routine verifications. We shall prove (i) as a sample and leave the rest to the reader.

For, $x, y \in V$, we have

$$\langle Ay, x\rangle = \langle y, A^*x\rangle = \overline{\langle A^*x, y\rangle} = \overline{\langle x, (A^*)^*y\rangle} = \langle (A^*)^*y, x\rangle . \tag{3}$$

Thus, $Ay - (A^*)^*y = 0$ for all $y \in V$. $\square$

Definition 82. Let $A\colon V \to V$ be linear. We say that V is *self-adjoint* if $A = A^*$.

It is customary to call a self-adjoint map A on a complex (respectively, real) inner product space as *hermitian* (respectively, *symmetric*).

If $AA^* = A^*A$, then A is said to be *normal*. Note that any self-adjoint map is normal.

Ex. 83. $A\colon V \to V$ is self-adjoint iff $\langle x, Ay\rangle = \langle AX, y\rangle$ for all $x, y \in V$.

Ex. 84. If A is self-adjoint, and $\lambda \in \mathbb{R}$, then so is λA.

Lemma 85. *Let $A\colon V \to V$ be normal. Then* $\ker A = \ker A^*$.

Proof. Let $x \in \ker A$. Then, we have

$$\begin{aligned} 0 = \langle Ax, Ax\rangle &= \langle x, A^*Ax\rangle \\ &= \langle x, AA^*x\rangle \\ &= \langle A^*x, A^*x\rangle . \end{aligned}$$

Thus, $Ax = 0$ iff $A^*x = 0$. ☐

Lemma 86. *Let $A\colon V \to V$ be normal. Then any generalized eigenvector of A is an eigenvector.*

Proof. We claim that $\ker A^k = \ker A$. We prove this by induction on k. Assume the result for k. Let $x \in V$ be such that $A^{k+1}x = 0$. Then, $A^k(x) \in \ker A = \ker A^*$. That is, $A^k(x) \in \ker A^*$. Therefore we have A period inserted. 9 replaced by (

$$0 = \left\langle A^*(A^k(x)), A^{k-1}x\right\rangle = \left\langle A^k(x), A^k(x)\right\rangle .$$

Hence $A^k(x) = 0$ which implies $Ax = 0$ by induction hypothesis.

Now, let λ be an eigenvalue of A and $v \in V(\lambda)$. Note that if A is normal, so is $A - \lambda I$.

> Reason: For, $(A - \lambda I)^* = (A^* - \overline{\lambda} I)$ and clearly, $(A - \lambda I)$ and $(A^* - \overline{\lambda} I)$ commute with each other.

The result now follows from the claim. ☐

Lemma 87. *If λ is an eigenvalue of A, then $\overline{\lambda}$ is an eigenvalue of A^*.*

Proof. Let $x \in V$ be such that $Ax = \lambda x$. The result follows from the following:

$$\begin{aligned} \left\langle (A^* - \overline{\lambda} I)x, (A^* - \overline{\lambda} I)x\right\rangle &= \langle x, (A - \lambda I)(A^* - \lambda I)x\rangle \\ &= \left\langle x, (A^* - \overline{\lambda} I)(A - \lambda I)x\right\rangle \\ &= \left\langle x, (A^* - \overline{\lambda} I)(0)\right\rangle . \\ &= 0. \end{aligned}$$

That is, $(A^* - \overline{\lambda} I)x = 0$. (Note that this result shows that if the field is $\mathbb{R}$, then $\lambda \in \mathbb{R}$ is an eigenvalue of A iff it is an eigenvalue of A^*.) ☐

Lemma 88. *Let $A \colon V \to V$ be normal. Then nonzero eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof. Let $Ax = \lambda x$ and $Ay = \mu y$, with $\lambda \neq \mu$. We are required to show that $\langle x, y \rangle = 0$. Note that if $x \neq 0$ and $y \neq 0$, then $x \neq y$

> Reason: For, otherwise, $Ax = \lambda x = \mu x$ so that $(\lambda - \mu)x = 0$. Since $\lambda - \mu \neq 0$, we conclude that $x = 0$, a contradiction.

We now compute:

$$\begin{aligned}
(\lambda - \mu) \langle x, y \rangle &= \langle (\lambda - \mu)x, y \rangle \\
&= \langle (A - \mu I)x, y \rangle \\
&= \langle x, (A^* - \overline{\mu})y \rangle \\
&= 0,
\end{aligned}$$

since y is an eigenvector of A^* with eigenvalue $\overline{\mu}$. Thus, $(\lambda - \mu) \langle x, y \rangle = 0$. Since $\lambda - \mu \neq 0$, we arrive at the result. □

Lemma 89. *Let $\lambda \in \mathbb{F}$ be an eigenvalue of a normal linear map $A \colon V \to V$. Then the orthogonal complement $V_\lambda^\perp$ of the eigenspace V_λ is invariant under A.*

Proof. Let $u \in V_\lambda^\perp$. Then we need to show that $Au \in V_\lambda^\perp$, that is, we must show that $\langle Au, v \rangle = 0$ for $v \in V_\lambda$. We have

$$\langle Au, v \rangle = \langle u, A^* v \rangle = \left\langle u, \overline{\lambda} v \right\rangle = \lambda \langle u, v \rangle = 0.$$

□

Theorem 90 (Spectral Theorem for Normal Linear Maps)**.** *Let $A \colon V \to V$ be a linear map on a finite dimensional inner product space V over $\mathbb{C}$. Then A is normal iff there exists an orthonormal eigen-basis, that is, an orthonormal basis of V consisting of eigenvectors of A.*

Proof. The proof is by induction on the dimension n of V. When $n = 1$, the result is clear, since any linear map is a multiplication by a scalar. Therefore, any nonzero vector will be constitute an eigen-basis.

Assume that $n > 1$. Since $\mathbb{C}$ is algebraically closed, there exists an eigenvalue $\lambda \in \mathbb{C}$. The eigen-subspace V_λ is nonzero, say, of dimension k. Hence its orthogonal complement $V_\lambda^\perp$ has dimension strictly less than n. By the last lemma, $V_\lambda^\perp$ is invariant under A, by induction hypothesis, $V_\lambda^\perp$ has an orthonormal eigen-basis of A restricted to $V_\lambda^\perp$. Let it be $\{v_{k+1}, \ldots, v_n\}$. Let $\{v_1, \ldots, v_k\}$ be an orthonormal basis of V_λ. Then clearly, $\{v_j : 1 \leq j \leq n\}$ is an orthonormal eigen-basis of A. □

Lemma 91. *If A is self-adjoint, then any eigenvalue of A is real.*

Proof. Let λ be an eigenvalue of A with an eigenvector u of unit norm. It suffices to show that $\overline{\lambda} = \lambda$. We have

$$\lambda = \lambda \langle u, u \rangle = \langle \lambda u, u \rangle = \langle Au, u \rangle = \langle u, A^* u \rangle = \langle u, Au \rangle = \langle u, \lambda u \rangle = \overline{\lambda} \langle u, u \rangle = \overline{\lambda}.$$

□

The next theorem gives us a result which does not impose any condition on the linear map.

Theorem 92. *Let $A \colon V \to V$ be any linear map on a finite dimensional inner product space over $\mathbb{C}$. Then there exists an orthonormal basis with respect to which the matrix of A is upper triangular.*

Proof. Let λ be an eigenvalue of A and v a unit vector such that $Av = \lambda v$. Let $W := (\mathbb{C}v)^{\perp}$. Consider $B \colon W \to W$ defined by

$$Bw := Aw - \langle Aw, v \rangle v.$$

Thus, B is the map A followed by the orthogonal projection from V onto W. Clearly, $\dim W = \dim V - 1$. By induction hypothesis, we may assume that there exists an ON basis of W with respect to which B is upper triangular. Let $\{w_1, \ldots, w_{n-1}\}$ be such a basis. In particular, we have

$$Bw_i \in \operatorname{span}\{w_1, \ldots, w_i\}.$$

Then $\{v_1 := v, v_2 := w_1, \ldots, v_n := w_{n-1}\}$ is an ON basis of V. We note that

$$\begin{aligned}
Av_1 &= \lambda v_1 \\
Av_2 &= Bw_1 + \langle Aw_1, w_1 \rangle v_1 \in \operatorname{span}\{v_1, v_2\} \\
&\vdots \\
Av_i &= Bw_{i-1} + \langle Aw_i, v \rangle v \in \operatorname{span}\{v_1, w_1, \ldots, w_{i-1}\}.
\end{aligned}$$

Thus, $Av_i \in \operatorname{span}\{v_1, \ldots, v_i\}$ for $1 \le i \le n$. Hence the matrix of A with respect to this basis is upper triangular. □

Proposition 93. *Let $A \colon V \to V$ be self-adjoint. Assume that $\langle Ax, x \rangle = 0$ for all $x \in V$. Then $A = 0$.*

Proof. This is clear from the spectral theorem for normal operators. Let $\{v_i : 1 \le i \le n\}$ be an ON eigen-basis of V. Then $\langle Av_i, v_i \rangle = 0$ for all $1 \le i \le n$. Hence

$$0 = \langle Av_i, v_i \rangle = \lambda_i \langle v_i, v_i \rangle = \lambda_i.$$

Thus, $\lambda_i = 0$ for all i and hence

$$Ax = A\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i A v_i = 0.$$

We also offer a direct proof.

$$\langle A(x+y), x+y\rangle = \langle Ax, x\rangle + \langle Ax, y\rangle + \langle x, Ay\rangle + \langle Ay, y\rangle = 0 + 2\langle Ax, y\rangle + 0.$$

If we fix $x \in V$, we see that $\langle Ax, y\rangle = 0$ for all $y \in V$. Hence $Ax = 0$. Since $x \in V$ is arbitrary, the result follows. □

12 Unitary and Orthogonal Linear Maps

Theorem 94. *Let $A\colon V \to V$ be a linear map on an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then the following are equivalent:*
(a) *$\langle Ax, Ay\rangle = \langle x, y\rangle$ for all $x, y \in V$.*
(b) *$\|Ax\| = \|x\|$ for all $x \in V$.*
(c) *$AA^* = A^*A = I$.*
(d) *A takes an ON basis to an ON basis.*

Proof. This is a standard result and the reader should have already learnt. So, the proof may be skipped.

(a) $\implies$ (b): If we take $x = y$ in (a), then we get

$$\|x\|^2 = \langle x, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2.$$

Taking (non-negative) square roots, (b) follows.

(b) $\implies$ (c): Observe the following:

$$\langle x, x\rangle = \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle.$$

Hence, we see that $\langle x, (A^*A - I)x\rangle = 0$. Since $(A^*A)^* = A^*A^{**} = A^*A$, we infer that A^*A is self-adjoint. Hence, $A^*A - I$ is self-adjoint. It follows from Proposition 93 that $A^*A - I = 0$. Similarly, one shows that $AA^* = I$.

(c) $\implies$ (d): Assume that $AA^* = I = A^*A$. Let $\{v_i : 1 \le i \le n\}$ be an ON basis. Then we have

$$\langle v_i, v_j\rangle = \langle A^*Av_i, v_j\rangle = \langle Av_i, Av_j\rangle.$$

Thus $\{Av_i : 1 \le i \le n\}$ is an ON set which has $\dim V$ elements. Hence it is an ON basis.

(d) $\implies$ (a): Let $\{v_i : 1 \le i \le n\}$ be an ON basis. Note that if $x = \sum_i x_i v_i$ and $y = \sum_i y_i v_i$, then (in the case of $\mathbb{F} = \mathbb{C}$, the case of $\mathbb{F} = \mathbb{R}$ being similar)

$$\langle x, y\rangle = \left\langle \sum_i x_i v_i, \sum_j y_j v_j \right\rangle = \sum_{i,j} x_i \overline{y_j} \langle v_i, v_j\rangle = \sum_i x_i \overline{y_i}.$$

Now, we compute $\langle Ax, Ay\rangle$:

$$\langle Ax, Ay\rangle = \left\langle \sum_i x_i Av_i, \sum_j y_j Av_j \right\rangle = \sum_{i,j} x_i \overline{y_j} \langle Av_i, Av_j\rangle = \sum_i x_i \overline{y_i}. = \langle x, y\rangle.$$

□

Definition 95. If $\mathbb{F} = \mathbb{C}$ and if $A\colon V \to V$ is a linear map which has any and hence all the properties of the theorem, then A is called a *unitary* map.

If $\mathbb{F} = \mathbb{R}$ and if $A\colon V \to V$ is a linear map which has any and hence all the properties of the theorem, then A is called a *orthogonal* map.

Lemma 96. *Let $A\colon V \to V$ be unitary (respectively, orthogonal). Then any eigenvalue of A is of unit modulus.*

Proof. Easy. Let v be a unit eigenvector corresponding to an eigenvalue λ.

$$1 = \langle v, v \rangle = \langle Av, Av \rangle = \langle \lambda v, \lambda v \rangle = \lambda\overline{\lambda} \langle v, v \rangle = \lambda\overline{\lambda}.$$

□

Theorem 97 (Spectral Theorem for Unitary Maps)**.** *Let V be an inner product space over $\mathbb{C}$. Let $A\colon V \to V$ be a unitary map. Then there exists an orthonormal basis of V with respect to which the matrix of A is a diagonal matrix of the form* $\operatorname{diag}(e^{it_1}, \dots, e^{it_n})$.

Proof. The proof is by induction on the dimension n of V. Let $n = 1$. Then any linear map is of the form $x \mapsto \lambda x$. Hence if we take v to be any unit vector, then v is an eigenvector with eigenvalue λ. By the last lemma, $\lambda = 1$ so that $\lambda = e^{it}$ for some $t \in \mathbb{R}$. Thus the map is $x \mapsto e^{it} x$.

$n > 1$. Let us assume result for all complex inner product spaces of dimension less than $n.n > 1$. Let V be a complex inner product space of dimension n and let $A\colon V \to V$ be a unitary map. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A wit a unit eigenvector v. Then $W := (\mathbb{C}v)^\perp$ is a vector subspace of dimension $n - 1$.

Reason: Consider the map $x \mapsto \langle x, v \rangle$. This is a linear map, it is nonzero since $v \mapsto 1$ and its kernel is precisely W. The claim now follows from the rank-nullity theorem.

We claim that W is invariant under A.

Reason: Let $w \in W$. We need to show that $Aw \in W$, that is, $\langle Aw, v \rangle = 0$. Consider the following:

$$0 = \langle w, v \rangle = \langle Aw, Av \rangle = \langle Aw, \lambda v \rangle = \overline{\lambda} \langle Aw, v \rangle .$$

Since $\overline{\lambda} \neq 0$, the claim follows.

Let W be equipped with the induced inner product. Then B, the restriction of A to W is a unitary operator. By induction hypothesis, there exists an ON basis $\{v_2, \dots, v_n\}$ such that $Av_j = e^{it_j} v_j$. Clearly, $\{v_1 := v, v_2, \dots, v_n\}$ is an ON basis as required. □

To prove an analogous spectral theorem for orthogonal matrix, we need a few preliminary results. It is worthwhile to revisit Example 17 at this juncture.

We shall assume for the rest of the section that V denotes a finite dimensional real inner product space. The crucial algebraic fact which we need is the following

Lemma 98. *Let $p(X) \in \mathbb{R}[X]$. Then $p(X)$ is a product of real polynomials of degree 1 or 2.*

Proof. We shall assume the fundamental theorem of algebra. Let $p(X) := X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0 \in \mathbb{R}[X]$. Let $\lambda \in \mathbb{C}$ be a root of p. Then $\overline{\lambda}$ is a root of p.

> Reason:
> $$p(\overline{\lambda}) = \overline{\lambda}^n + a_{n-1}\overline{\lambda}^{n-1} + \cdots + a_1\overline{\lambda} + a_0 = \overline{p(\lambda)} = 0.$$

Hence non-real complex roots occur in conjugate pairs. Hence

$$X^2 + bX + c := (X - \overline{\lambda})(X - \lambda) = X^2 - 2\mathrm{Re}\,\lambda X + |\lambda|^2$$

is a divisor of p in $\mathbb{C}[X]$. Note that $b^2 - 4c < 0$.

> Reason:
> $$b^2 - 4c = 4(\mathrm{Re}\,\lambda)^2 - 4\,|\lambda|^2 = 4[(\mathrm{Re}\,\lambda)^2 - |\lambda|^2] < 0.$$

If we write $p(X) = (X^2 + bX + c)q(X)$ with $q(X) \in \mathbb{C}[X]$, we claim that $q(X) \in \mathbb{R}[X]$.

> Reason: Note that $q(X) = p(X)/(X^2 + bX + c)$ since $X^2 + bX + c \neq 0$, as it has no real roots. For any $t \in \mathbb{R}$, we have $q(t) \in \mathbb{R}$ since $p(t), t^2 + bt + c \in \mathbb{R}$. So, $\mathrm{Im}\,q(t) = \mathrm{Im}\,(a_{n-1})t^{n-1} + \cdots + \mathrm{Im}\,(a_1)t + \mathrm{Im}\,(a_0) = 0$ for all $t \in \mathbb{R}$. But then the real polynomial $\mathrm{Im}\,q$ has infinitely many roots and hence must be identically zero. We therefore conclude that the coefficients of q lie in $\mathbb{R}$.

We now apply induction hypothesis to q and get the result. □

Proposition 99. *Let $A\colon V \to V$ be linear. Then there exists an A-invariant subspace $W \subseteq V$ with $\dim W$ equal to 1 or 2.*

Proof. We mimic the proof of Theorem 19.

Let $p(X) \in \mathbb{R}[X]$ be a nonzero monic polynomial such that $p(A) = 0$. We write $p(X) = p_1(X)\cdots p_k(X)$ where $\deg p_j \leq 2$ for $1 \leq j \leq k$. Let $v \in V$ be a nonzero vector such that $p(A)v = 0$. Arguing as in Theorem 19, we find that there exists a maximum i such that $p_1(A) \circ \cdots \circ p_i(A)(w) = 0$ where $p_{i+1}(A) \circ \cdots \circ p_k(A)v \neq 0$.

If $p_i(X)$ is of degree 1, then $p_i(X) = X - \lambda_i$ for some $\lambda_i \in \mathbb{R}$. Then λ_i is an eigenvalue with eigenvector w. If $p_i(X)$ is of degree 2, say, $p_i(X) = X^2 + bX + c$, then $(A^2 + bA + cI)w = 0$. If we take, $W := \mathrm{span}\{w, Aw\}$, then W is invariant under A.

> Reason: We see that $A(Aw) = A^2 w = -bAw - cw \in \mathrm{span}\{w, Aw\}$.

□

Proposition 100. *Let $A\colon V \to V$ be self-adjoint. Let $b, c \in \mathbb{R}$ be such that $b^2 - 4c < 0$. Then $A^2 + bA + cI$ is invertible.*

Proof. The idea is to show that

$$\left\langle (A^2 + bA + cI)v, v \right\rangle > 0 \text{ for any nonzero } v \in V. \tag{4}$$

The inequality clearly implies that A is one-one and hence onto. We compute, for $v \neq 0$,

$$\begin{aligned}
\left\langle (A^2 + bA + cI)v, v \right\rangle &= \left\langle A^2 v, v \right\rangle + b \left\langle Av, v \right\rangle + c \left\langle v, v \right\rangle \\
&= \left\langle Av, Av \right\rangle + b \left\langle Av, v \right\rangle + c \left\langle v, v \right\rangle \\
&\geq \| Av \|^2 - |b| \, \| Av \| \, \| v \| + c \, \| v \|^2 \\
&= \left(\| Av \| - \frac{|b| \, \| v \|}{2} \right)^2 + \left(c - \frac{b^2}{4} \right) \| v \|^2 \\
&> 0.
\end{aligned}$$

(We have used Cauchy-Schwarz inequality above.) ☐

Lemma 101. *Let $A \colon V \to V$ be self-adjoint. Then A has a real eigenvalue.*

Proof. The proof is very similar to that of Theorem 19. If $p(X) \in \mathbb{R}[X]$ is a monic polynomial such that $p(A) = 0$, then we write

$$p(X) = p_1(X) \cdots p_r(X)(X - \lambda_1) \cdots (X - \lambda_k),$$

where $p_j(X)$ are second degree polynomials with non-real roots. Hence if we write $p_j(X) = X^2 + b_j X + c_j$, then $b_j^2 - 4c_j < 0$. Hence the fact that $p(A) = 0$ implies that

$$p_1(A) \cdots p_r(A)(A - \lambda_1 I) \cdots (A - \lambda_k) = 0.$$

Since $p_j(A)$ are invertible by the last lemma, we deduce that $(A - \lambda_1 I) \cdots (A - \lambda_k) = 0$. So, one of $A - \lambda_j I$ must be singular.

If no such linear factors occur, then $p(A)$ is invertible in view of the last lemma and hence $p(A) \neq 0$, a contradiction. ☐

Proceeding as in the proof of Theorem 90, we arrive at

Theorem 102 (Spectral Theorem for Self-adjoint Maps)**.** *Let $A \colon V \to V$ be self-adjoint. Then there exists an ON basis of V consisting of eigenvectors of A.* ☐

Reason:

Ex. 103. Find the matrix (w.r.t. the standard basis) of an orthogonal map of $\mathbb{R}^2$ with the Euclidean inner product. *Hint:* Note that $\{Ae_1, Ae_2\}$ is an orthonormal basis of $\mathbb{R}^2$ and that any vector of unit norm can be written as $\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ for some $\theta \in \mathbb{R}$. Hence A is either of the form $k(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ or of the form $r(\theta) := \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ for some $\theta \in \mathbb{R}$. $k(\theta)$ (resp. $r(\theta)$) is called a rotation (resp. a reflection).

Ex. 104. If $A \colon V \to V$ is orthogonal and λ is an eigen value of A, then $\lambda = \pm 1$.

Ex. 105. Let $T\colon V \to V$ be orthogonal. Let $A := T + T^{-1} = T + T^*$. Then A is symmetric and let $V := \oplus_i V_i$ be the orthogonal decomposition of A into distinct eigen spaces. Then

(a) T leaves each V_i invariant.

(b) If V_λ is an eigen space of A with eigen value λ, then we have $T^2 - \lambda T + I = 0$ on V_λ.

(c) If $\lambda = \pm 2$, then T acts as $\pm I$ on V_λ.

(d) If $\lambda \neq \pm 2$, then $W := \mathbb{R}v + \mathbb{R}(Tv)$ is a two-dimensional subspace such that $TW \subset W$. Also, if $V_\lambda = W \oplus W^\perp$, then $TW^\perp \subset W^\perp$. Hence V_λ is orthogonal direct sum of two dimensional vector subspaces invariant under T.

(e) If $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is orthogonal and satisfies $T^2 + \lambda T + I = 0$ for some $\lambda \neq \pm 2$, then T is a rotation.

Ex. 106 (Spectral Theorem for Orthogonal Operators)**.** Let T be orthogonal. Then there exists an orthonormal basis of V with respect to which T can be represented as follows:

$$T = \begin{pmatrix} \pm 1 & & & & & \\ & \ddots & & & & \\ & & \pm 1 & & & \\ & & & \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} & & \\ & & & & \ddots & \\ & & & & & \begin{pmatrix} \cos\theta_r & -\sin\theta_r \\ \sin\theta_r & \cos\theta_r \end{pmatrix} \end{pmatrix}.$$

That is, T is the block matrix

$$T = \operatorname{diag}\left(\pm 1, \cdots, \pm 1, k(\theta_1), \cdots, k(\theta_r)\right).$$

Hint: Ex. 105 and Ex. 103.

The last couple of results are valid for inner product spaces over $\mathbb{R}$ or $\mathbb{C}$. Let V be an inner product space over $\mathbb{R}$ or $\mathbb{C}$ and $A\colon V \to V$ be linear.

Definition 107. A is said to be *positive* if (i) A is self-adjoint and (ii) $\langle Ax, x\rangle \geq 0$ for all $x \in V$.

Ex. 108. Show that the eigen values of a positive operator A are nonnegative and that there exists a unique operator S such that S is positive and $S^2 = A$. The operator S is called the positive square root of A.

Ex. 109 (Polar Decomposition for Invertible Maps)**.** Let $A\colon V \to V$ be nonsingular. Then there exists a unique decomposition $A = PU$ where U is unitary (or orthogonal) and P is positive. (This decomposition is called the polar decomposition of A.) *Hint:* Think of complex numbers. The map AA^* is positive and let S be its positive square root. Then $U := S^{-1}A$ may do the job. But why does S^{-1} exist?

Ex. 110 (Polar Decomposition)**.** Let $A\colon V \to V$ be any linear map. Then there exists a unitary (orthogonal) map U and a positive map P such that $A = PU$. *Hint:* Let $S := \sqrt{AA^*}$. Let $W := SV$. Define $U_1\colon W \to V$ by setting $U_1(Sv) := Av$. Observe that $\dim W^\perp = \dim(AV)^\perp$. Define a unitary map $U_2\colon W^\perp \to (AV)^\perp$.