NOTES II FOR 130A

JACOB STERBENZ

Abstract. Here are some notes on the Jordan canonical form as it was covered in class.

Contents

1. Polynomials 1
2. The Minimal Polynomial and the Primary Decomposition 3
3. When the Characteristic Polynomial Factors 4
4. The Structure of Nilpotent Operators 5
5. The Real and Complex Jordan Forms 7

1. Polynomials

As will become apparent shortly, the theory of invariant subspaces of a linear transformation $T \in \mathcal{L}(V)$ turns on the issue of factoring polynomials into prime powers. Accordingly we begin here with a basic review of polynomial arithmetic.

First some notation. Let $\mathbb{F}$ denote a field of numbers. For our purposes one can take this to be either real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. However, everything we say here is equally valid for any field (e.g. rational numbers $\mathbb{Q}$). We denote by:

$$\mathbb{F}[x] = \{ p(x) \mid p(x) = \sum_{i=0}^{d} a_i x^i, \text{where } a_i \in \mathbb{F} \} ,$$

for various values of $d$. For any particular $p \in \mathbb{F}[x]$ we define:

$$\deg(p) = \min\{ d \mid p(x) = \sum_{i=0}^{d} a_i x^i \text{ and } a_d \neq 0 \} .$$

Note that this is just the usual notion of degree of a polynomial (the highest power of $x$ you see in $p(x)$). We also say that $p \in \mathbb{F}[x]$ is monic if $\deg(p) = d$ and $a_d = 1$. This is simply a normalization condition that is useful in the statement of several results to follow.

Perhaps the most basic result about polynomials in one variable is the following:

**Theorem 1.1** (Division with remainder). Let $p, q \in \mathbb{F}[x]$ be any two polynomials. Then there exists $g, r \in \mathbb{F}[x]$ such that $q = pg + r$ where $\deg(r) < \deg(p)$.

**Proof.** This follows from long division of polynomials just as one learns in high school algebra. For starters we may assume $\deg(p) \leq \deg(q)$ because otherwise we can simply choose $g = 0$ and $r = q$. Then we have:

$$q(x) = \sum_{i=0}^{d_1} a_i x^i, \quad p(x) = \sum_{i=0}^{d_2} b_i x^i,$$

with $d_1 \geq d_2$, and $a_{d_1} \neq 0$ and $b_{d_2} \neq 0$. Let $g_1 = \frac{a_{d_1}}{b_{d_2}} x^{d_1 - d_2}$, and then set $q_1 = q - g_1 p$. We have $\deg(q_1) < \deg(q)$. If $\deg(q_1) < \deg(p)$ we are done by setting $r = q_1$. Otherwise repeat this process to produce $q_2 = q_1 - g_2 p$ with $\deg(q_2) < \deg(q_1)$. Again check if $\deg(q_2) < \deg(p)$ or not, and divide again if this condition has not yet been achieved. Eventually this process will produce some $q_k$ with $\deg(q_k) < \deg(p)$ and $q_k = q - (g_1 + g_2 + \ldots + g_k)p$. Finally set $g = \sum_{i=1}^{k} g_i$ and $r = q_k$ are we are done. \qed
For collections of polynomials we have the following importation notion:

**Definition 1.2.** A collection of polynomials $I \subseteq \mathbb{F}[x]$ is called an “ideal” if:

i) For any $p, q \in I$ we have that $p + q \in I$.

ii) For any $p \in I$ and any other $q \in \mathbb{F}[x]$ we have $pq \in I$.

An example of an ideal would be the set $I = \{ p \in \mathbb{F}[x] \mid a_0 = 0 \}$. In other words this ideal consists of all polynomials of the form $xg(x)$ for some (not fixed) $g \in \mathbb{F}[x]$. Division with remainder shows that in some sense this is the general picture for polynomial ideals:

**Theorem 1.3** (Generation of ideals). Let $I \subseteq \mathbb{F}[x]$ be any nonzero ideal. Then there exists a unique monic $p \in \mathbb{F}[x]$ such that $I = \{ q \in \mathbb{F}[x] \mid q = pg \text{ for some } g \in \mathbb{F}[x] \}$. In this case we write $I = (p)$ and say “$I$ is the ideal generated by $p$.”

**Proof.** Let $d$ be the minimum degree over all nonzero polynomials in $I$, and let $p \in I$ be a monic polynomial with $\deg(p) = d$. Let $q \in I$ be anything else. Then we have $q = gp + r$ for some $\deg(r) < \deg(p)$ polynomial. But $q - gp \in I$ by rules i) and ii) for ideals, so $r \in I$ as well. If $r \neq 0$ then it would be a nonzero polynomial in $I$ with degree less that $p$, a contradiction. Thus $r = 0$ and we have $q = pg$ as desired.

Note that uniqueness follows because if $\deg(p) = \deg(q)$ and $q = gp$ for some $g \in \mathbb{F}[x]$, we must have $g = a$ for some $a \in \mathbb{F}$ (this is simply because $\deg(gp) = \deg(g) + \deg(p)$). Thus $p$ is the only monic polynomial that can generate $I$.

**Definition 1.4.** Next, given two polynomials $p, q \in \mathbb{F}[x]$ we say “$p$ divides $q$” of there exists a $g \in \mathbb{F}[x]$ with $q = pg$. In this case we write $p\mid q$.

We say a polynomial $p$ is “prime” if $\deg(p) \geq 1$ and if $q\mid p$ for some $q \in \mathbb{F}[x]$ we must have either $p = aq$ for some $a \in \mathbb{F}$ or $\deg(q) = 0$ (in the latter case $q = a_0$ for some $a_0 \in \mathbb{F}$).

Given a (finite) collection of polynomials $p_i \in \mathbb{F}[x]$ we say “$p_i$ are relatively prime” if $q\mid p_i$ for all $i$ implies that $\deg(q) = 0$. In other words the only common factors of all the $p_i$ are constants.

It is important to understand that the notion of being prime really depends on $\mathbb{F}$. The polynomial $p = x^2 + 1$ is prime when $\mathbb{F} = \mathbb{R}$ but not when $\mathbb{F} = \mathbb{C}$. The fundamental theorem of algebra states that the only prime polynomials if $\mathbb{C}[x]$ are those of the form $p = a(x - \lambda)$. On the other hand the only prime polynomials in $\mathbb{R}[x]$ are of the form $p = a(x - \lambda)$ or $p(x) = ax^2 + bx + c$ where $b^2 - 4ac < 0$. This is because if a polynomial has real coefficients then its roots are either real or come in complex conjugate pairs. Also, note that a collection of non-constant polynomials $p_i \in \mathbb{C}[x]$ are relatively prime iff they have no root in common. The same is true for $p_i \in \mathbb{R}[x]$ as long as one takes into account complex roots (this is necessary because $p_1 = x^2 + 1$ and $p_2 = x^4 + 2x^2 + 1$ have no real roots in common but they are also not relatively prime in $\mathbb{R}[x]$).

From the previous result on generation of ideals we now have the following theorem which will be our main tool in the next Section:

**Theorem 1.5.** Let $q_i \in \mathbb{F}[x]$ be a finite collection of nonzero relatively prime polynomials. Then there exists $g_i \in \mathbb{F}[x]$ such that $\sum_i q_i g_i = 1$.

**Proof.** First define an ideal that is generated by the $q_i$:

$$I = \{ q \in \mathbb{F}[x] \mid q = \sum_i g_i q_i \text{ for some } g_i \in \mathbb{F}[x] \}.$$  

One can check immediately that this collection of polynomials satisfies i) and ii) above, and since the $q_i$ are nonzero $I$ itself is also nonzero. Thus there exists a unique monic $p \in \mathbb{F}[x]$ with $I = (p)$. In particular $p\mid q_i$ for all $i$, and because the $q_i$ are relatively prime this means $\deg(p) = 0$. The only monic polynomial of degree 0 is $p = 1$. Thus $1 \in I$, and so by the construction of $I$ there must exists $g_i \in \mathbb{F}[x]$ with $\sum_i g_i q_i = 1$.

We end with a prime factorization theorem for polynomials which mirrors what one knows about integers:

**Theorem 1.6.** Let $q \in \mathbb{F}[x]$. Then one can write uniquely $q = a \prod_{i=1}^k p_i^{r_i}$, where $a \in \mathbb{F}$, where $p_i \in \mathbb{F}[x]$ are distinct prime monic polynomials, and $r_i \in \mathbb{N}$. If $\mathbb{F} = \mathbb{C}$ then each $\deg(p_i) = 1$, and if $\mathbb{F} = \mathbb{R}$ then each $1 \leq \deg(p_i) \leq 2$.  

2
Proof. The existence of some prime factorization of \( q \in \mathbb{F}[x] \) follows by inductively splitting \( q \) into smaller factors. For example either \( q \) is already prime or \( q = q_1q_2 \) where each \( 1 \leq \deg(q_1) < \deg(q) \). Then if at least one of \( q_1, q_2 \) is not prime we split further and continue this process until all factors are prime.

The main issue then is the uniqueness of the prime factorization. This is a consequence of the following claim: If \( p \) is prime and \( p|q_1q_2 \), then either \( p|q_1 \) or \( p|q_2 \). The claim follows because if \( p, q_1 \) are relatively prime then one can find \( f, g \in \mathbb{F}[x] \) with \( fp + gq_1 = 1 \). Then \( fq_2p + gq_1q_2 = q_2 \). But \( p|q_1q_2 \) implies \( q_1q_2 = gp \) so rearranging things we have \((fq_2 + g)p = q_2\). In other words \( p|q_2 \).

By repeatedly applying the previous fact to both sides of the following equation we see that if:

\[
q = a \prod_{i=1}^{k} p_i^{r_i} = a' \prod_{i=1}^{k'} (p_i')^{r_i'} ,
\]
then we must have \( a = a' \), \( k = k' \), \( r_i = r_i' \), and \( p_i = p_i' \). \( \Box \)

2. The Minimal Polynomial and the Primary Decomposition

Now let \( V \) be a vector space over the real or complex numbers \( \mathbb{F} \). Let \( T \in \mathcal{L}(V) \). The primary purpose of this section is to associate a certain polynomial \( m_T \in \mathbb{F}[x] \) with \( T \) in such a way that the prime factorization of \( m_T \) gives a great deal of information about the invariant subspaces of \( T \). Recall that finding the invariant subspaces of \( T \) is the key step to finding all of the solutions to \( x = Tx \).

To set things up we first define an ideal in \( \mathbb{F}[x] \) associated with \( T \). This is:

\[
I(T) = \{ q \in \mathbb{F}[x] \mid q(T) = 0 \} .
\]
Here we mean that if \( q(x) = \sum_{i=0}^{d} a_i x^i \) then we define \( q(T) = \sum_{i=0}^{d} a_i T^i \) with the proviso that \( T^0 = I \) for any \( T \). Note that if \( q(T) = 0 \) simply means \( q(T)x = 0 \) for all \( x \in V \). It is not hard to check that \( I(T) \) is indeed an ideal (basically zero plus zero is zero, and zero times anything is zero).

A little bit less obvious is that \( I(T) \) is not the zero ideal. To see this note that if \( x \in V \) then the collection of vectors \( \{ x, Tx, T^2x, \ldots, T^nx \} \), where \( n = \dim(V) \), must be linearly dependent. Thus there exists \( q \in \mathbb{F}[x] \) such that \( q(T)x = 0 \). Constructing such \( q_i \) for a basis \( B = \{ x_1, \ldots, x_n \} \) and then setting \( q = \prod q_i \) gives a nonzero polynomial with \( q(T)x = 0 \) for all \( x \in V \).

Definition 2.1. From Theorem 1.3 in the previous section we must have \( I(T) = (m_T) \) for some unique monic \( m_T \in \mathbb{F}[x] \). In other words there exists a polynomial \( m_T \) with \( m_T|q \) for all other \( q \in \mathbb{F}[x] \) with \( q(T) = 0 \). We call \( m_T \) the “minimal polynomial of \( T \)”.

Using the prime factorization of \( m_T \) we have the fundamental result:

Theorem 2.2 (Primary Decomposition Theorem). Let \( T \in \mathcal{L}(V) \) and \( m_T \) its minimal polynomial. Then if \( m_T = \prod_{i=1}^{k} p_i^{r_i} \) is the prime factorization of \( m_T \) there exists a \( T \) invariant direct sum decomposition \( V = \bigoplus_{i=1}^{k} V_i \) where for each \( i = 1, \ldots, k \) one has \( m_T|_{V_i} = p_i^{r_i} \) and moreover each \( V_i = \ker(p_i^{r_i}(T)) \).

Proof. We’ll break the proof down into a number of steps.

Step 1: (Construction of \( V_i \)) This is the key part of the proof. Let \( m_T = \prod_{i=1}^{k} p_i^{r_i} \) be the factorization of \( m_T \) into distinct prime powers and for each \( i = 1, \ldots, k \) define:

\[
q_i = \prod_{j \neq i} p_j^{r_j} = m_T \frac{p_i^{r_i}}{p_i^{r_i}} .
\]
Then by construction the \( q_i \) are relatively prime so by Theorem 1.5 there exists \( g_i \in \mathbb{F}[x] \) with \( \sum_i g_i q_i = 1 \). Using these polynomials define \( E_i \in \mathcal{L}(V) \) by the formulas \( E_i = g_i(T)q_i(T) \).

The first thing that is immediate is that \( [E_i, T] = E_i T - T E_i = 0 \), and also \( \sum_i E_i = I \). Now define:

\[
V_i = \text{ran}(E_i) = \{ y \in V \mid y = E_i x \text{ for some } x \in V \} .
\]
Note that \( \text{span}(V_1, \ldots, V_k) = V \) because \( x = \sum E_i x \) for each \( x \in V \). Therefore to show \( V = \bigoplus_{i=1}^{k} V_i \) we only need \( V_i \cap V_j = \{ 0 \} \) for all \( i \neq j \). This will follow if \( E_i E_j = 0 \) for all \( i \neq j \). But \( m_T|_{V_i} g_i q_j q_j \) whenever \( i \neq j \) because \( g_i q_j \) for \( i \neq j \) contains every prime factor \( p_i^{r_i} \) for all \( l = 1, \ldots, k \). Thus \( E_i E_j = g_i(T)q_i(T)g_j(T)q_j(T) = 0 \).
To complete this part of the proof we just need to show the $V_i$ are $T$ invariant. By $E_i E_j = 0$ we have $x = E_i x$ for all $x \in V_i$. Thus if $x \in V_i$ we have $T x = T E_i x = E_i T x \in \text{ran}(E_i) = V_i$.

**Step 2:** (Show that $V_i = \ker(p_i'(T))$) First let $x \in V_i$. Then $x = g_i(T)q_i(T)x$, so multiplying both sides by $p_i'(T)$ and using $p_i'(T)q_i(T) = m_i(T) = 0$ we have $p_i'(T)x = 0$. Thus $V_i \subseteq \ker(p_i'(T))$.

On the other hand suppose $p_i'(T)x = 0$. Since $x = \sum x_j$ where each $x_j \in V_j$, and since $T(V_j) \subseteq V_j$ we must also have $p_i'(T)x_j = 0$. But the $V_j$ are linearly independent so this implies $p_i'(T)x_j = 0$ for each $j$. Now for $j \neq i$ we have $p_i'(T)g_jq_j$, so there exists $\bar{g} \in \mathbb{F}[x]$ with $\bar{g}p_i'(T) = g_jq_j$. Multiplying the previous equation by $\bar{g}(T)$ gives $E_i x_j = 0$ for all $i \neq j$. In other words $x_j = 0$ for all $i \neq j$. Thus $x = x_i$ so we have shown $\ker(p_i'(T)) \subseteq V_i$.

**Step 3:** (Show that the minimal polynomial of $T|_{V_i}$ is $p_i'$) Since $V_i = \ker(p_i'(T))$ we have $p_i'(T_{|V_i}) = 0$. Thus $m_{T|_{V_i}} | p_i'$, so we must have $m_{T|_{V_i}} = p_i'$ for some $r' \leq r_i$. Now define $\tilde{m} = p_i' q_i = p_i' \prod_{j \neq i} p_j' y$. Since each $p_j'(T) V_i = \{0\}$ and $p_i'(T) V_i = \{0\}$ we have $\tilde{m}(T) = 0$ on all of $V$. Thus $m_{T|V_i}$ which implies $r_i \leq r'$ and we are done. 

\[\square\]

3. When the Characteristic Polynomial Factors

We now bring in the connection between the minimal polynomial $m_T$ and the characteristic polynomial $p_T$. Recall that the latter is defined by the equation $p_T(x) = \det(xI - T)$, and $\lambda \in \mathbb{F}$ is an eigenvalue of $T$ iff $p_T(\lambda) = 0$. First we give an indication that $m_T$ is in fact closely related to $p_T$:

**Lemma 3.1.** Given $T \in \mathcal{L}(T)$ the polynomials $m_T$ and $p_T$ have exactly the same roots. Moreover, $m_T$ factors over $\mathbb{F}$ iff $p_T$ factors over $\mathbb{F}$.

**Proof.** First of all let $p \in I(T)$ be any polynomial such that $p(T) = 0$. Then if $T x = \lambda x$ for some $x \neq 0$ a direct computation shows $p(T)x = p(\lambda)x$. Thus $p(\lambda) = 0$. Therefore the roots of $p_T$ are roots of $m_T$ as well.

The other direction is a bit more work. Let $\lambda \in \mathbb{F}$ be a root of $m_T$. Then by prime factorization $m_T = (x-\lambda)^r q(x)$ where $(x-\lambda)$ and $q$ are relatively prime. Let $V = V_\lambda \oplus V_q$ be the corresponding invariant decomposition of $V$ according to Theorem 2.2. Then the minimal polynomial of $T|_{V_\lambda}$ is exactly $(x-\lambda)^r$. In other words $(T|_{V_\lambda} - \lambda I)^r = 0$ but $(T|_{V_q} - \lambda I)^{-1} \neq 0$. Let $y = (T|_{V_\lambda} - \lambda I)^{-1} x$ with $y \neq 0$. Then $(T|_{V_\lambda} - \lambda I)y = 0$ so $y \in V_\lambda \subseteq V$ is an eigenvector for $T$ (on all of $V$).

Finally, the part about factorization is trivial if $\mathbb{F} = \mathbb{C}$ (in this case every polynomial factors). On the other hand if $V$ is a real vector space, by extending $T$ to real operator on the complexification $V^\mathbb{C}$, we just need to see that the minimal polynomial of the extension is the same as the original. Since $\overline{p(T)} = \overline{p(T)}$ for any $p \in \mathbb{C}[x]$, we see that if $p(T) = 0$ then $\Re(p)(T) = \Im(p)(T) = 0$, where $\Re(p)$ and $\Im(p)$ are the real polynomials with $p = \Re(p) + i\Im(p)$. This implies that the minimal polynomial of a real operator is real, and so the minimal polynomials of $T$ on both $V$ and $V^\mathbb{C}$ are the same. 

\[\square\]

Before stating our main result we need one more idea. First a definition:

**Definition 3.2.** An operator $N \in \mathcal{L}(V)$ is called “nilpotent” if $N^k = 0$ for some $k \geq 1$. If $N$ is nilpotent we call $k_0 = \min\{k \mid N^k = 0\}$ the “order of $N$”.

Now:

**Lemma 3.3.** Let $N \in \mathcal{L}(V)$ be nilpotent. Then the order of $N$ is $\leq \dim(V)$. Moreover for any $\lambda \in \mathbb{F}$ with $\lambda \neq 0$ one has $\lambda I + N$ is invertible.

**Proof.** Let $k_0$ be the order of $N$, and choose some $x \in V$ with $y = N^{k_0-1} x \neq 0$. Consider the collection of vectors $\{x, Nx, N^2x, \ldots, N^{k_0-1}x\}$, and suppose $\sum_{i=0}^{k_0-1} a_i N^i x = 0$ for some $a_i \in \mathbb{F}$. Applying $N^{k_0-1}$ to both sides gives $a_0 y = 0$. Hence $a_0 = 0$. Applying $N^{k_0-2}$ gives $a_1 y = 0$, so $a_1 = 0$. Continuing in this way applying $N^{k_0-j}$ for $j = 3, \ldots, k_0$ (in order) gives $a_{j-1} = 0$. This shows that $\{x, Nx, N^2x, \ldots, N^{k_0-1}x\}$ is a linearly independent collection of $k_0$ vectors, and so we must have $k_0 \leq \dim(V)$.

Finally, the part about the inverse follows from the direct calculation $(\lambda I + N)^{-1} = \frac{1}{\lambda} \sum_{i=0}^{k_0-1} (-\frac{1}{\lambda} N)^i$. 

\[\square\]

We can now state the Main Theorem of these notes:
Theorem 3.4 (Abstract Jordan Form). Let $V$ be a vector space over $\mathbb{F}$ and let $T \in \mathcal{L}(V)$. Suppose that the characteristic polynomial of $T$ factors over $\mathbb{F}$, that is $p_T(x) = \prod_{i=1}^k (x - \lambda_i)^{d_i}$ for some $\lambda_i \in \mathbb{F}$. Then there is a $T$ invariant direct sum decomposition $V = \bigoplus_{i=1}^k V_i$ where $V_i = \text{ker}((T - \lambda_i I)^{d_i})$. Moreover $V_i = \text{ker}((T - \lambda_i I)^r)$ are exactly the subspaces associated with the factorization of the minimal polynomial $m_T(x) = \prod_{i=1}^k (x - \lambda_i)^{r_i}$, and we have the relation $r_i \leq d_i = \dim(V_i)$. In particular $m_T|_{V_i}$.

As a consequence we can write $T = S + N$ where $S$ is diagonalizable with eigenvalues $\lambda_i$, and $N$ is nilpotent with $[S,N] = 0$. Moreover there is only such pair $S,N$ which yields this decomposition. Therefore, we have that $T$ itself is diagonalizable iff $N = 0$, which is equivalent to $r_i = 1$ for every prime factor of the minimal polynomial.

Proof. Assuming that $p_T(x) = \prod_{i=1}^k (x - \lambda_i)^{d_i}$, for some $\lambda_i \in \mathbb{F}$ and $d_i \in \mathbb{N}$ implies by Lemma 3.1 that $m_T(x) = \prod_{i=1}^k (x - \lambda_i)^{r_i}$ for the same $\lambda_i$ but possibly different $r_i$. Now let $V = \bigoplus_{i=1}^k V_i$ be the direct sum decomposition according to Theorem 2.2, so that by construction $V_i = \text{ker}((T - \lambda_i I)^{r_i})$. Since the $V_i$ are $T$ invariant subspaces we must have $p_T = \prod_{i=1}^k p_T|_{V_i}$. On the other hand since $m_T|_{V_i} = (x - \lambda_i)^{r_i}$ and $p_T|_{V_i}$ have the same roots then $p_T|_{V_i} = (x - \lambda_i)^{d_i}$ where $d_i = \dim(V_i)$ and $d_i$ is the same as in the factorization of $p_T$. Also, because $T|_{V_i} - \lambda_i I$ is nilpotent of order $r_i$, by Lemma 3.3 we have $r_i \leq d_i$.

Next we show that $V_i = \ker((T - \lambda_i I)^{d_i})$. Since we already know that $V_i = \ker((T - \lambda_i I)^{r_i})$ and $r_i \leq d_i$ the containment $V_i \subseteq \ker((T - \lambda_i I)^{d_i})$ is immediate. On the other hand if $x = \sum j x_j$, where $x_j \in V_j$, is such that $x \in \ker((T - \lambda_i I)^{d_i})$, then we know from invariance and independence that individually $x_j \in \ker((T - \lambda_i I)^{d_i})$. For $j \neq i$ we can write $T|_{V_j} - \lambda_i I = M + N$ where $\lambda = \lambda_j - \lambda_i \neq 0$ and $N = T|_{V_j} - \lambda_i I$ is nilpotent. Thus by Lemma 3.3 $T|_{V_j} - \lambda_i I$ is invertible when $i \neq j$, so $\ker(T|_{V_i} - \lambda_i I)^{d_i} = \{0\}$. Thus $x_j = 0$ for all $j \neq i$ which shows $x \in V_i$.

Finally, to get the $T = S + N$ decomposition we just need to construct $S$ and $N$ in terms of blocks for each $V_i$. Let $S|_{V_i} = \lambda_i I$, and let $N|_{V_i} = T|_{V_j} - \lambda_i I$. Then it is clear that $S$ is semisimple and $N$ is nilpotent with $[S,N]$. Now let $T = S' + N'$ be any other such decomposition. By the condition $[S',T] = 0$ we see that $S',E_i = 0$ for each of the projections constructed in Theorem 2.2, so $S'(V_i) \subseteq V_i$ for each $i$. Therefore $V = \bigoplus_{i=1}^k V_i$ is also an invariant direct sum decomposition for $S'$, and hence for $N'$ as well. Now if $S'x = \lambda x$, where $x = \sum x_i$ and $x_i \in V_i$ we must have $Sx_i = \lambda x_i$ by independence. Thus, if $B = \{e_1, \ldots, e_n\}$ is a basis for $V$ consisting of eigenvectors of $S'$, then $S_i = \{E_i e_1, \ldots, E_i e_n\}$ must be a collection of eigenvectors of $S'|_{V_i}$ which spans $V_i$. In other words $S'|_{V_i}$ is also diagonalizable for each $i$. Now let $x \in V_i$ be any nonzero vector in $S'|_{V_i}$, and let $k_x \geq 1$ be the smallest power such that $(N')^{k_x}x = 0$. Then if we set $y = (N')^{k_x - 1}x$ have $\tilde{y} = y \in V_i$ and $Ty = (S' + N')y = \lambda y$. Thus, we have shown that any eigenvalue of $S'|_{V_i}$ must also be an eigenvalue of $T|_{V_i}$, so $S'|_{V_i} = \lambda I_i$. Now $N'|_{V_i} = T|_{V_j} - \lambda_i I$ and we are done.

We’ll end this section we a few additional definitions regarding the subspaces $V_i$ constructed in the previous theorem.

Definition 3.5. Let $T \in \mathcal{L}(V)$ and assume that its characteristic polynomial factors as $p_T(x) = (x - \lambda)^{d_\lambda} q(x)$ where $(x - \lambda), q(x)$ are relatively prime. Then we call $V_\lambda = \ker(T - \lambda I)^{d_\lambda}$ the “generalized eigenspace associated to $\lambda$”. Note that $d_\lambda = \dim(V_\lambda)$. We call $d_\lambda$ the “algebraic multiplicity of eigenvalue $\lambda$”, and we note $e_\lambda = \dim(\ker(T - \lambda I))$ the “geometric multiplicity of eigenvalue $\lambda$”.

Note that we always have $1 \leq e_\lambda \leq d_\lambda$, and that $e_\lambda$ is exactly the number of linearly independent eigenvectors of $T$ in $V_\lambda$ (which all must have eigenvalue $\lambda$). In particular $T$ is diagonalizable iff $p_T(x)$ factors and $e_\lambda = d_\lambda$ for all eigenvalues $\lambda$. Finally, note it is possible for $e_\lambda$ to be much smaller than $d_\lambda$. For example if $r_\lambda = d_\lambda$, where $r_\lambda$ is the power of $(x - \lambda)$ in the minimal polynomial $m_T$, then we must have $e_\lambda = 1$.

4. The Structure of Nilpotent Operators

In light of Theorem 3.4 we see that to get the simplest form for a general $T \in \mathcal{L}(V)$ is suffices to study a general nilpotent operator $N \in \mathcal{L}(V)$. It turns out that the key notion in this regard is the following:

Definition 4.1. Let $T \in \mathcal{L}(V)$ be any linear operator, and let $x \in V$ be a fixed vector. Then we define:

$$Z(x,T) = \text{span}\{x,Tx,T^2x,\ldots,T^{\dim(V) - 1}x\}.$$
We call $Z(x, T)$ the “$T$-cyclic subspace of $V$ generated by $x$”. We call $\dim(Z(x, T)) = O(x, T)$ the “$T$-order of $x$”.

Note that by the properties of the characteristic polynomial we have $T^{\dim(V) + j}x \in Z(x, T)$ for all $j \geq 0$. Our main theorem now is the following:

**Theorem 4.2.** Let $N \in L(V)$ be a nilpotent operator of order $k_0$. Then there exists a collection of vectors $x_1, \ldots, x_k \in V$ with $O(x_1, N) \subseteq O(x_2, N) \subseteq \ldots \subseteq O(x_k, N) = k_0$ and $V = \bigoplus_{i=1}^k Z(x_i, N)$. This decomposition is unique in the sense that if $V = \bigoplus_{i=1}^{k'} Z(y_i, N)$ is any other such decomposition we must have $k = k'$ and $O(x_i, N) = O(y_i, N)$ for all $i = 1, \ldots, k$.

In particular there is a basis $B$ consisting of vectors of the form $N^i x_j$ such that:

\[
[N]_B = \begin{bmatrix}
N_{\ell_1} & N_{\ell_2} & \cdots & N_{\ell_k}
\end{bmatrix},
\]

where each $\ell_i = O(x_i, N)$, and $N_{\ell} \in M(\ell \times \ell)$ is the “elementary nilpotent matrix of order $\ell$”:

\[
N_{\ell} = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & \cdots & 0 & 1 \\
0 & \cdots & \cdots & \cdots & 0
\end{bmatrix}.
\]

Any such form of $N$ is unique with the proviso that $\ell_1 \leq \ell_2 \leq \ldots \leq \ell_k$.

**Proof.** The proof of the cyclic decomposition $V = \bigoplus_{i=1}^k Z(x_i, N)$ we give here is the same as in Appendix III of the text. The proof is by induction on the dimension of $V$. If $\dim(V) = 1$ then the only nilpotent operator is $N = 0$ and so $V = Z(x, N)$ for any $x \neq \vec{0}$. Now let $N, V$ be arbitrary, and set $W = N(V) \subseteq V$. Since $N$ is nilpotent we must have $\ker(N) \neq \{\vec{0}\}$ so $W \neq \vec{0}$. Now $N(W) \subseteq W$, and $\dim(W) < \dim(V)$ so by induction we can decompose $W = \bigoplus_{i=1}^k Z(y_i, N|_W)$ for some $y_i \in W$. Choose any $x_i \in V$ with $y_i = N x_i$.

First suppose that there exists $p_i \in \mathbb{F}[x]$ with $\sum_i p_i(N)x_i = \vec{0}$ and each $\deg(p_i) < O(x_i, N) = O(y_i, N) + 1$. Applying $N$ to this and using the independence of $Z(y_i, N|_W)$ we see that for each $i$ either $p_i = 0$ or $p_i \neq 0$ (as a polynomial) and $p_i(N)y_i = \vec{0}$. In other words $p_i = 0$ or else $p_i(x) = a_i x^{O(y_i, N)}$. But this also implies that $\sum_{a_i \not= 0} a_i N^{O(y_i, N) - 1} y_i = \vec{0}$, so $a_i = 0$ for each term in that sum. This shows that the $Z(x_i, N)$ themselves are independent.

It remains to show $V = \bigoplus_{i=1}^k Z(x_i, N) \oplus L$ where $L = \bigoplus_{j=k+1}^l Z(x_j, N)$ for some other collection of vectors $x_j$. Since $N(\bigoplus_{i=1}^k Z(x_i, N)) = N(V)$, we must have $V = \operatorname{span}(\bigoplus_{i=1}^k Z(x_i, N), \ker(N))$. Let $\ker(N) = L \oplus L'$ where $L' = \ker(N) \cap \bigoplus_{i=1}^k Z(x_i, N)$ and $L$ is any other complimentary subspace. Then $N(L) = \{\vec{0}\} \subseteq L$, so $V = \bigoplus_{i=1}^k Z(x_i, N) \oplus L$ is an invariant direct sum decomposition. If $\{x_{k+1}, \ldots, x_l\}$ is any basis for $L$ we have $L = \bigoplus_{j=k+1}^l \operatorname{span}(x_j)$, and automatically $\operatorname{span}(x_j) = Z(x_j, N)$, so we are done.

Notice that this inductive procedure also establishes uniqueness in terms of the orders $O(x_i, N)$ because if the orders of $O(y_i, N|_W)$ are uniquely determined for $i = 1 \ldots k$, then so are the orders $O(x_i, N) = O(y_i, N|_W) + 1$. And the number of remaining vectors, which all have order 1, is also determined by the number $\dim(V) - \sum_{i=1}^k O(x_i, N)$.

Finally, to see the canonical form (1) is valid suppose that $V = \operatorname{span}(x, Nx, \ldots, N^{\ell-1}x)$ where $\ell = \dim(V) = O(x, N)$. Then in the basis $B = \{e_1 = N^{\ell-1}x, e_2 = N^{\ell-2}x, \ldots, e_\ell = x\}$ we have $Ne_1 = 0$ and $Ne_i = e_{i-1}$ for all $i = 2, \ldots, \ell$. Thus $[N]_B = N_{\ell}$ where $N_{\ell}$ is defined by (2) above. \qed
Remark 4.3. Note that if \( V = \text{span}(x, Nx, \ldots, N^{\ell - 1}x) \) where \( \ell = \dim(V) = O(x, N) \), and instead we use the basis \( \mathcal{B} = \{e_1 = x, e_2 = Nx, \ldots, e_\ell = N^{\ell - 1}x\} \) we will get:

\[
[N]_\mathcal{B} = N^\dagger_\ell = \begin{bmatrix}
0 & 1 & 0 & \cdots & \cdots & 1 \\
1 & 0 & & & & \\
& & \ddots & & & \\
& & & \ddots & & \\
& & & & 0 & 1
\end{bmatrix}.
\]

This is the basic form for nilpotent matrices used in the text. I have chosen to employ the upper triangular form here which is a little more common by modern standards.

5. The Real and Complex Jordan Forms

Combining what we have done so far yields the so called Jordan canonical form of a matrix:

**Theorem 5.1** (Jordan form). Let \( V \) be a vector space over \( \mathbb{F} \) and let \( T \in \mathcal{L}(V) \) be such that its characteristic polynomial \( p_T(x) \) factors over \( \mathbb{F} \) as \( p_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{d_i} \). In addition suppose the minimal polynomial of \( T \) factors as \( m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{r_i} \). Then there exists a basis \( \mathcal{B} \) for \( V \) such that:

\[
[T]_\mathcal{B} = \begin{bmatrix}
J_{\ell_{11}}(\lambda_1) & J_{\ell_{12}}(\lambda_1) & \cdots & & \\
& J_{\ell_{11}}(\lambda_1) & \cdots & & \\
& & \ddots & \ddots & \\
& & \cdots & J_{\ell_{11}}(\lambda_1) & \\
& & & & J_{\ell_{11}}(\lambda_k)
\end{bmatrix}
\]

where \( \ell_{ij} \leq \ell_{ij+1} \) for each \( ij \) and \( \ell_{in_i} = r_i \), and \( \sum_{j=1}^{n_i} \ell_{ij} = d_i \), and each elementary block \( J_{\ell_{ij}}(\lambda_i) \) is an \( \ell_{ij} \times \ell_{ij} \) “Jordan matrix” of the form:

\[
J_{\ell}(\lambda) = \lambda I + N_{\ell} = \begin{bmatrix}
\lambda & 1 & \cdots & 0 \\
& \lambda & \ddots & \vdots \\
& & \ddots & 1 \\
& & & \lambda
\end{bmatrix}.
\]

**Proof.** This follows more or less directly by combining Theorems 3.4 and 4.2. First note that \( T \) takes a block form along the primary direct sum decomposition \( V = \oplus_{i=1}^{k} V_i \), and on each \( V_i \) we have \( T|_{V_i} = \lambda_i I + N_i \) for some nilpotent operator \( N_i \in \mathcal{L}(V_i) \) or order \( r_i \). By applying Theorem 4.2 to \( N_i \) we are done. \( \square \)

The previous Theorem is not quite satisfactory when we want to compute solutions to \( \dot{x} = Tx \) on a real vectors space \( V \) when the characteristic polynomial \( T \in \mathcal{L}(V) \) does not factor over \( \mathbb{R} \). However, the only obstruction here is complex conjugate pairs of roots in the factorization of \( p_T(x) \). Using the material from the previous notes on complexifications and the previous Theorem we have:

**Theorem 5.2** (Real Jordan form). Let \( V \) be a real vector space and let \( T \in \mathcal{L}(V) \) be such that its characteristic polynomial \( p_T(x) \) factors over \( \mathbb{C} \) as \( p_T(x) = \prod_{i=1}^{k_r} (x - \lambda_i)^{d_i} \prod_{j=1}^{k_c} (x - \mu_j)^{e_j} \), where \( \lambda_i \in \mathbb{R} \) and \( \mu_j = a_j + \sqrt{-1}b_j \) and \( b_j \neq 0 \). In addition suppose the minimal polynomial of \( T \) factors as \( m_T(x) = \prod_{i=1}^{k_r} (x - \lambda_i)^{r_i} \prod_{j=1}^{k_c} (x - \mu_j)^{s_j} \), where \( \lambda_i \in \mathbb{R} \).
\[ \prod_{i=1}^{k_i} (x - \lambda_i)^{r_i} \prod_{j=1}^{k_j} (x - \mu_j)^{s_j} (x - \overline{\mu}_j)^{s_j}. \] Then there exists a (real) basis \( \mathcal{B} \) for \( V \) such that:

\[
[T]_{\mathcal{B}} = \begin{bmatrix}
J_{\ell_{i,1}}(\lambda_1) & & \\
& \ddots & \\
& & J_{\ell_{k,r}}(\lambda_{k,r}) & \\
& & & K_{s_{1,1}}(\mu_1) & \\
& & & & \ddots & \\
& & & & & K_{s_{k,m_k}}(\mu_{k,c})
\end{bmatrix}
\]

where \( \ell_{ij} \leq \ell_{i,j+1} \) for each \( ij \) and \( \ell_{im} = r_i \), and \( \sum_{j=1}^{n_i} \ell_{ij} = d_i \), and each real elementary Jordan block \( J_i(\lambda) \) is the same as before. Furthermore \( s_{ij} \leq s_{i,j+1} \) for each \( ij \) and \( s_{im} = t_i \) while \( \sum_{j=1}^{m_i} s_{ij} = e_i \) and each matrix \( K_{s_{ij}}(\mu_i) \), where \( \mu_i = a_i + \sqrt{-1}b_i \), is a \( 2s_{ij} \times 2s_{ij} \) block matrix of the form:

\[
K_s(\mu) = \begin{bmatrix}
a & -b & I & & \\
& a & -b & \ddots & \\
& & a & -b & \\
& & & \ddots & I & \\
& & & & a & -b & \\
& & & & & b & a
\end{bmatrix}.
\]

**Proof.** Using Theorem 3.4 we can reduce to the invariant factor associates to a single pair of conjugate complex eigenvalues. In other words we may assume \( p_\mu(x) = (x - \mu)^r(x - \overline{\mu})^r \) and \( m_\mu(x) = (x - \mu)^t(x - \overline{\mu})^t \) where \( \mu = a + ib \) with \( b \neq 0 \). Then using the complex Jordan normal form we know that \( V^C = V_\mu \oplus V_\overline{\mu} \) where each factor is the invariant subspace associated to eigenvalues \( \mu \) and \( \overline{\mu} \) respectively.

Let \( \sigma \) be the complex conjugation on \( V^C \). The key observation now is that \( \sigma(V_\mu) = V_\overline{\mu} \). To see this recall that \( V_\mu = \ker(T^C - \mu I)^r \), and an easy calculation shows that \( \sigma(\ker(T^C - \mu I)^r) = \ker(T^C - \overline{\mu} I)^r \) thanks to \( \sigma T^C = T^C \sigma \). Because of this we further have \( T^C = S + N \) where \( S = \mu E_\mu + \overline{\mu} \sigma E_\mu \sigma \) where \( \sigma E_\mu \sigma = E_{\overline{\mu}} \) and the \( E \) are projections onto the generalized eigenspaces. This implies that both \( \sigma S = S \sigma \) and \( \sigma N = N \sigma \), so both \( S \) and \( N \) in the semisimple/nilpotent decomposition of \( T \) are real operators.

Now suppose \( V_\mu = \bigoplus_{j=1}^{k} Z(z_j, N) \). Then using the mirror symmetry \( \sigma(V_\mu) = V_{\overline{\mu}} \) and the fact \( N \) is real we must have \( V_{\overline{\mu}} = \bigoplus_{j=1}^{k} Z(\sigma(z_j), N) \). In particular we can pair cyclic spaces to get the real decomposition:

\[ V = \bigoplus_{j=1}^{k} \left( Z(z_j, N) \oplus Z(\sigma(z_j), N) \right)^R. \]

For each fixed value of \( j \) let \( s_j = O(z_j, N) \) denote the (complex) dimension of the cyclic factor \( Z(z_j, N) \), and choose the real basis of \( (Z(z_j, N) \oplus Z(\sigma(z_j), N))^R \) to be \( \mathcal{B}_j = \{ y_{1,j}, x_1, \ldots, y_{s_j}, x_{s_j} \} \) where \( x_1 + iy_1 = N^{s_j-1}z_j \). Then one can readily check that in this basis the matrix of \( T \) restricted to the factor \( (Z(z_j, N) \oplus Z(\sigma(z_j), N))^R \) is exactly \( K_{s_j}(\mu) \) from line (3). \( \Box \)

**E-mail address**: jsterben@math.ucsd.edu

---

**Department of Mathematics, University of California, San Diego (UCSD) 9500 Gilman Drive, Dept 0112 La Jolla, CA 92037-0112 USA**