THE DIAGONALIZABLE AND NILPOTENT PARTS OF A MATRIX

HERBERT A. MEDINA

1. Introduction

It is an easy consequence of the Jordan canonical form that a matrix $A \in M_{n\times n}(\mathbb{C})$ can be decomposed into a sum $A = D_A + N_A$ where D_A is a diagonalizable matrix, N_A a nilpotent matrix, and such that $D_A N_A = N_A D_A$. It is clear that both D_A and N_A also commute with A. This decomposition is often referred to as the *Jordan decomposition* and has found many applications throughout the years. For example, it is not hard to see that computing powers of A is much simpler if we know its Jordan decomposition. (The reader may refer to [1] for some of the recent applications of this decomposition.)

It is "often" the case that if a matrix B commutes with a given matrix A, then B is in fact a polynomial in A. This is indeed the case for the D_A and N_A described above. A proof of this fact can be found in [3, §6.8]. The proof presented therein, which perhaps can be best described as "semi constructive" as it invokes a fact about polynomials that depends on the Euclidean algorithm, relies on the decomposition of $\mathbb{C}^n$ into subspaces that are invariant under the operator induced by the matrix A. The recent work done on this problem is much more technical and has focused on finding fast algorithms to express D_A and N_A as polynomials in A (e.g., [1, 4]).

This article is devoted to an introduction and a new, computer-algebra-system motivated, very elementary solution (acessible to an undergraduate with an upper-division background in linear algebra) to the problem of expressing D_A and N_A as polynomials in A. Specifically, we 1.) introduce the diagonalizable + nilpotent decomposition; 2.) present a simple (perhaps the simplest so far as it relies almost exclusively on the Jordan form and matrix algebra), constructive proof of the fact that D_A and N_A can be expressed as polynomials in A; 3.) discuss some of the consequences our proof; and 4.) state a few questions that arise from our construction.

We begin by stating the theorem whose proof will be our focus for the majority of the article.

Theorem 1. *Let $A \in M_{n\times n}(\mathbb{C})$ and let $A = D_A + N_A$ be a decomposition of A where D_A is diagonalizable, N_A is nilpotent and $N_A D_A = D_A N_A$. Then there exists a polynomials $p(x), q(x) \in P(\mathbb{C})$ such that $p(A) = N_A$ and $q(A) = D_A$. Moreover, N_A and D_A are unique.*

2. Preliminaries and a Reduction

Let us begin with a discussion of the decomposition central to this article. The reader can refer to many linear algebra texts (e.g., [2, 3]) to see a development and proof of the fact that any matrix $A \in M_{n\times n}(\mathbb{C})$ can be written in the form $A =$

QJQ^{-1}, where $Q \in M_{n\times n}(\mathbb{C})$ is invertible and J is a matrix in Jordan canonical form. That is,

$$J = \begin{pmatrix} J^1_{\lambda_1} &&&&&&&& \\ & \ddots &&&&&&& \\ && J^{m_1}_{\lambda_1} &&&&&& \\ &&& J^1_{\lambda_2} &&&&& \\ &&&& \ddots &&&& \\ &&&&& J^{m_2}_{\lambda_2} &&& \\ &&&&&& \ddots && \\ &&&&&&& J^1_{\lambda_q} & \\ &&&&&&&& \ddots \\ &&&&&&&&& J^{m_q}_{\lambda_q} \end{pmatrix},$$

where each *Jordan block* $J^j_{\lambda_i}$ is either a 1×1 matrix containing the eigenvalue λ_i or a matrix of the form

$$J^j_{\lambda_i} = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_i & 1 & & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ & & & & \lambda_i & 1 \\ & & & & & \lambda_i \end{pmatrix}.$$

This decomposition is unique up to rearrangement of the blocks.

It is evident that any one of the blocks $J^j_{\lambda_i}$ can be written uniquely as the sum of a diagonal matrix $D^j_i = \begin{pmatrix} \lambda_i && \\ & \ddots & \\ && \lambda_i \end{pmatrix}$ and a nilpotent matrix $N^j_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ & \ddots & & \ddots & \vdots \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$.

The decomposition of J into such a sum is done in the same way block by block. The matrix with the N^j_i submatrices along its diagonal is nilpotent because one can multiply matrices of the form $M = \begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_k \end{pmatrix}$ where M_i are submatrices of smaller size than M by multiplying the corresponding submatrices. We leave it to the reader to check that the decomposition $J = D_J + N_J$ into the sum of a diagonal and nilpotent matrix is unique and that these two summands commute; we will call D_J the *diagonal part* and N_J the *nilpotent part* of J.

The first reduction that we make is evident: if we can find $p(x)$ as in the theorem, then we can take $q(x) = p(x) - x$. Thus we may focus on finding $p(x)$. (We note that this approach is different than the one taken in [3] where they focus on finding $q(x)$.)

If $A = QJQ^{-1}$, and $J = D_J + N_J$ as described above, then $A = QD_JQ^{-1} + QN_JQ^{-1}$ decomposes A into a sum of a diagonalizable matrix and a nilpotent matrix and these two summands commute. The proof of Theorem 1 relies on showing that there is a polynomial $p(x)$ such that $p(A) = QN_JQ^{-1}$; once we have done this, we will be able to show that the decomposition in the statement of theorem is unique. That is, that

$N_A = QN_JQ^{-1}$. This will complete the argument. The existence of such a polynomial is an easy consequence of a similar result for the nilpotent part of J. We state this reduction as a lemma whose proof is immediate.

Lemma 1. *Let* $A \in M_{n\times n}(\mathbb{C})$ *and let* J *be its Jordan canonical form such that* $A = QJQ^{-1}$. *Let* $J = D_J + N_J$ *be the decomposition of* J *into the sum of its diagonal and nilpotent parts. If* $p(x)$ *is a polynomial such that* $p(J) = N_J$, *then* $p(A) = QN_JQ^{-1}$.

The existence of this polynomial $p(x)$ for J is the key to our proof of Theorem 1.

3. An Example

The construction of a polynomial $p(x)$ such that $p(J) = N_J$ is not hard and is best understood via an example. In fact, after the example is well understood, the existence of $p(x)$ in general is apparent. Indeed, the construction was found by using *Mathematica* to compute examples; at first the computations were ad-hoc, but eventually a systematic method for computing $p(x)$ was found.

To illustrate the construction procedure, it is best to use a large matrix. Let

$$J = \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{pmatrix}.$$

The idea is first to produce a polynomial that will give the nilpotent part of J that is present in the Jordan blocks corresponding to λ_1. We then do the same for λ_2, λ_3.

We begin our construction of $p(x)$ by considering

$$p_{1,1}(x) = \frac{1}{(\lambda_1 - \lambda_2)^2(\lambda_1 - \lambda_3)}(x - \lambda_1)(x - \lambda_2)^2(x - \lambda_3).$$

When applied to J, this polynomial will get rid of the entries in the blocks corresponding to λ_2, λ_3. It gives us the 1's in the nilpotent part of J contained in the blocks corresponding to λ_1, but also gives us entries that we do not want. Indeed,

$$p_{1,1}(J) = \begin{pmatrix} 0 & 1 & d_1(1) & d_1(2) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & d_1(1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & d_1(1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (1)$$

where $d_1(1)=-\frac{3\lambda_1-\lambda_2-2\lambda_3}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)}$, $d_1(2)=\frac{3\lambda_1-2\lambda_2-\lambda_3}{(\lambda_1-\lambda_2)^2(\lambda_1-\lambda_3)}$.

We now would like a polynomial $p_{1,2}(x)$ such that $p_{1,1}(J)-d_1(1)p_{1,2}(J)$ will get rid of the unwanted $d_1(1)$'s. We take

$$p_{1,2}(x)=\frac{1}{(\lambda_1-\lambda_2)^2(\lambda_1-\lambda_3)}(x-\lambda_1)^2(x-\lambda_2)^2(x-\lambda_3).$$

It is not difficult to see that the extra $(J-\lambda_1 I)$ factor will push all the entries up one row:

$$p_{1,2}(J)=\begin{pmatrix}0&0&1&d_1(1)&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{pmatrix}.$$

Now, $p_{1,1}(J)-d_1(1)p_{1,2}(J)$ only has an unwanted $d_1(2)-(d_1(1))^2$ in its 1st row 4th column. To get rid of it we use

$$p_{1,3}(x)=\frac{1}{(\lambda_1-\lambda_2)^2(\lambda_1-\lambda_3)}(x-\lambda_1)^3(x-\lambda_2)^2(x-\lambda_3),$$

because

$$p_{1,3}(J)=\begin{pmatrix}0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{pmatrix}.$$

Therefore, if we let $p_1(x)=p_{1,1}(x)-d_1(1)p_{1,2}(x)-(d_1(2)-(d_1(1))^2)p_{1,3}(x)$, then

$$p_1(J)=\begin{pmatrix}0&1&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{pmatrix}.$$

We now repeat the process for the block(s) corresponding to λ_2. We note that only one step is needed for blocks of size 2. We get

$$p_2(x) = \frac{1}{(\lambda_2-\lambda_1)^4(\lambda_2-\lambda_3)}(x-\lambda_2)(x-\lambda_1)^4(x-\lambda_3).$$

There is no need to do anything for λ_3 because the $(x-\lambda_3)$ in $p_1(x)$ and $p_2(x)$ will zero out the diagonal entry for that block. (This is true in general for eigenvalues having only one-blocks.) For the sake of the general procedure, we set $p_3(x) = 0$.

The polynomial we want is

$$p(x) = p_1(x) + p_2(x) + p_3(x).$$

In this case, after some *Mathematica* simplification,

$$\begin{aligned} p(x) \;\; = \;\; & \frac{(x-\lambda_1)\,(x-\lambda_2)\,(x-\lambda_3)}{(\lambda_1-\lambda_2)^4}\Bigg(\frac{(x-\lambda_2)\,(\lambda_1-\lambda_2)^2}{\lambda_1-\lambda_3}+\frac{(x-\lambda_1)^3}{\lambda_2-\lambda_3}+ \\ & \frac{(x-\lambda_1)\,(x-\lambda_2)\,(\lambda_1-\lambda_2)\,(-3\,\lambda_1+\lambda_2+2\,\lambda_3)}{(\lambda_1-\lambda_3)^2}+ \\ & \frac{(x-\lambda_1)^2\,(x-\lambda_2)\,\left(6\,{\lambda_1}^2+{\lambda_2}^2+2\,\lambda_2\,\lambda_3+3\,{\lambda_3}^2-4\,\lambda_1\,(\lambda_2+2\,\lambda_3)\right)}{(\lambda_1-\lambda_3)^3}\Bigg). \end{aligned}$$

We note that the construction forces $p(x)$ to contain as a factor the product of the linear factors of the minimal polynomial of J.

4. Proof Theorem 1

We begin with a couple of lemmas from elementary linear algebra; the proof of the first is left to the reader.

Lemma 2. *Let* $N = \begin{pmatrix} 0 & 1 & 0 & \cdots & & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & \vdots \\ 0 & \cdots & & & 0 & 1 \\ 0 & \cdots & & & 0 & 0 \end{pmatrix}$ *be an* $n \times n$ *matrix.*

(i.) For $k \geq 1$, $(N^k)_{ij} = \begin{cases} 1 & j-i=k; \\ 0 & \text{otherwise.} \end{cases}$

(ii.) Let $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ *be any* $n \times n$ *matrix. Then*

$$N^k A = \begin{pmatrix} a_{(k+1)1} & \cdots & a_{(k+1)n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \\ 0 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

In a matrix $A = (a_{ij})$ we will call the set of elements $\{a_{ij} : j - i = k\}$ the k^{th} *diagonal of* A. The 0^{th} diagonal will be referred to as the *main diagonal.*

Lemma 3. *Let* $A = (a_{ij})$, $B = (b_{ij})$ *be* $n \times n$ *upper triangular matrices such that* $a_{ij} = a_{km}$ *and* $b_{ij} = b_{km}$ *whenever* $j - i = m - k$. *Then* $AB = (c_{ij})$ *also has the property that* $c_{ij} = c_{km}$ *whenever* $j - i = m - k$. *That is, if* A, B *are upper triangular and constant in each of their diagonals, then their product will be also. Moreover, if* $A' = (a'_{ij})$, $B' = (b'_{ij})$ *are upper triangular* $n' \times n'$ *such that the* k^{th} *diagonal of* $A = k^{\text{th}}$ *diagonal of* A' *and the* k^{th} *diagonal of* $B = k^{\text{th}}$ *diagonal of* B' *for all* k *that make sense, then the* k^{th} *diagonal of* $(AB) = k^{\text{th}}$ *diagonal of* $(A'B')$.

Proof. Let $j \geq i$ and $m \geq k$ with $j - i = m - k$.

$$c_{ij} = \sum_{r=i}^{j} a_{ir}\, b_{rj}.$$

By hypothesis, $a_{ir} = a_{k(r-i+k)}$ and $b_{rj} = b_{(j-r+m)m}$. Therefore,

$$c_{ij} = \sum_{r=i}^{j} a_{k(r-i+k)}\, b_{(r-j+m)m}.$$

Let $s = r - i + k$. When $r = i$, $s = k$; and when $r = j$, $s = j - i + k = m - k + k = m$. Also, $r - j + m = s + i - k - j + m = s - (j - i) + (m - k) = s$. Thus

$$c_{ij} = \sum_{s=k}^{m} a_{ks}\, b_{sm} = c_{km}. \tag{2}$$

Equation (2) shows that the elements in the product matrix only depend on the values in the diagonal entries and not on the size of the matrix. □

It is interesting (albeit not surprising) to note that the lemma is false without the upper triangular hypothesis.

We now turn to the general construction of $p(x)$.

The case when J is a diagonal matrix is immediate, because if we take $p(x)$ to be the minimal polynomial of J, then $p(J) = 0$, the nilpotent part of J.

Let $J \in M_{n\times n}(\mathbb{C})$ be in Jordan canonical form

$$J = \begin{pmatrix} J^1_{\lambda_1} & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & J^{m_1}_{\lambda_1} & & & & & & & \\ & & & J^1_{\lambda_2} & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & J^{m_2}_{\lambda_2} & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & J^1_{\lambda_q} & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & J^{m_q}_{\lambda_q} \end{pmatrix}, \tag{3}$$

where $\lambda_1, \ldots, \lambda_q$ are the distinct eigenvalues of J, $J^1_{\lambda_i}, \ldots, J^{m_i}_{\lambda_i}$ are the distinct Jordan blocks corresponding to λ_i, and where at least one of the blocks is of length 2. We assume that the blocks are ordered so that $J^1_{\lambda_i}$ is at least as large as any other Jordan block corresponding to λ_i; let s_i be its size. (The order of the blocks does not matter for the proof, but it helps to understand the construction.)

Define

$$p_{1,1}(x) = \frac{1}{(\lambda_1 - \lambda_2)^{s_2} \cdots (\lambda_1 - \lambda_q)^{s_q}} (x - \lambda_1)(x - \lambda_2)^{s_2} \cdots (x - \lambda_q)^{s_q}.$$

We first note that

$$p_{1,1}(J) = \begin{pmatrix} p_{11}(J^1_{\lambda_1}) & & & & & \\ & \ddots & & & & \\ & & p_{11}(J^{m_1}_{\lambda_1}) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

This is because the product in $p_{1,1}(J)$ can be computed by multiplying the submatrices of the size of the Jordan blocks along the diagonal of J in each of the factors of $p_{1,1}(J)$. For $i \geq 2$, the factor $(J - \lambda_i)^{s_i}$ will have zero submatrices in the positions corresponding to the Jordan blocks for λ_i. Therefore, $p_{1,1}(J)$ also will have zero submatrices in these positions.

To compute $p_{1,1}(J^i_{\lambda_1})$, we write

$$p_{1,1}(J^i_{\lambda_1}) = (J^i_{\lambda_1} - \lambda_1 I)\frac{(J^i_{\lambda_1} - \lambda_2 I)^{s_2}}{(\lambda_1 - \lambda_2)^{s_2}} \cdots \frac{(J^i_{\lambda_1} - \lambda_q I)^{s_q}}{(\lambda_1 - \lambda_q)^{s_q}}.$$

We will need to show that each of the $\frac{(J^i_{\lambda_1} - \lambda_l I)^{s_l}}{(\lambda_1 - \lambda_l)^{s_l}}$ factors has constant diagonals. The next lemma shows this and more as it gives a formula for the entries in the diagonals.

Lemma 4. *For each i, k,*

$$\frac{(J^i_{\lambda_1} - \lambda_k I)^{s_k}}{(\lambda_1 - \lambda_k)^{s_k}} = \begin{pmatrix} 1 & b_{s_k}(1) & b_{s_k}(2) & \cdots & \cdots & b_{s_k}(m-2) & b_{s_k}(m-1) \\ 0 & 1 & b_{s_k}(1) & \cdots & \cdots & \cdots & b_{s_k}(m-2) \\ & & \ddots & & \ddots & & \vdots \\ & & & & 1 & b_{s_k}(1) & b_{s_k}(2) \\ & & & & & 1 & b_{s_k}(1) \\ & & & & & & 1 \end{pmatrix}, \tag{4}$$

where $b_{s_k}(j) = \begin{cases} \frac{\binom{s_k}{j}}{(\lambda_1 - \lambda_k)^j} & j \leq s_k; \\ 0 & j > s_k. \end{cases}$

Proof. We argue by induction on s_k. The base case is clear. Assume the result for $s_k = l$.

$$\frac{(J^i_{\lambda_1} - \lambda_k I)^{l+1}}{(\lambda_1 - \lambda_k)^{l+1}} = \begin{pmatrix} 1 & \frac{1}{\lambda_1 - \lambda_k} & 0 \cdots & \cdots & 0 \\ 0 & 1 & \frac{1}{\lambda_1 - \lambda_k} & 0 \cdots & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & \frac{1}{\lambda_1 - \lambda_k} \\ & & & & 1 \end{pmatrix} \times$$

$$\begin{pmatrix} 1 & b_l(1) & b_l(2) & \cdots & \cdots & b_l(m-2) & b_l(m-1)\\ 0 & 1 & b_l(1) & \cdots & \cdots & \cdots & b_l(m-2)\\ & \ddots & \ddots & \ddots & & & \vdots \\ & & & & 1 & b_l(1) & b_l(2)\\ & & & & & 1 & b_l(1) \\ & & & & & & 1\end{pmatrix}.$$

The main diagonal of the resulting product is clear. Note that for $j \geq i$, the i, j component of matrix in (4), is $b_{s_k}(j-i)$. Therefore, we have to show that the i, j component of the matrix resulting from the product above is $b_{l+1}(j-i)$. The i, j component in the product is $1 \cdot b_l(j-i) + \frac{1}{(\lambda_1-\lambda_k)} \cdot b_l(j-i-1)$. This expression is computed in three cases:

$$\begin{cases} \frac{\binom{l}{j-i}}{(\lambda_1-\lambda_k)^{j-i}} + \frac{1}{(\lambda_1-\lambda_k)}\frac{\binom{l}{j-i-1}}{(\lambda_1-\lambda_k)^{j-i-1}} = \frac{\binom{l+1}{j-i}}{(\lambda_1-\lambda_k)^{j-i}} = b_{l+1}(j-i) & j-i \leq l;\\ 0 + \frac{1}{(\lambda_1-\lambda_k)}\frac{\binom{l}{j-i-1}}{(\lambda_1-\lambda_k)^{j-i-1}} = \frac{\binom{l+1}{j-i}}{(\lambda_1-\lambda_k)^{j-i}} = b_{l+1}(j-i) & j-i = l+1;\\ 0 = b_{l+1}(j-i) & j-i \geq l+2.\end{cases}$$

□

Lemma 4 not only shows that the diagonals in each of the factors of $p_{1,1}(J^i_{\lambda_1})$ are constant, but that the k^{th} diagonal in the corresponding factor of $p_{1,1}(J^i_{\lambda_1})$ and $p_{1,1}(J^{i'}_{\lambda_1})$ have the same constant entry. (The reader may want to refer to equation (1) for an example of this observation.)

Using Lemmas 2 – 4, we conclude that

$$p_{1,1}(J^i_{\lambda_1}) = \begin{pmatrix} 0 & 1 & d_1(1) & d_1(2) & \cdots & \cdots & d_1(m-2) & d_1(m-1)\\ 0 & 0 & 1 & d_1(1) & \cdots & \cdots & \cdots & d_1(m-2)\\ & & \ddots & \ddots & \ddots & & & \vdots\\ & & & & 0 & 1 & d_1(1) & d_1(2)\\ & & & & & 0 & 1 & d_1(1)\\ & & & & & & 0 & 1\\ & & & & & & \cdots & 0\end{pmatrix}, \tag{5}$$

where d_1 is a function independent of i. That is, the k^{th} diagonal in $p_{1,1}(J^i_{\lambda_1}) = p_{1,1}(J^{i'}_{\lambda_1})$ for any two $i, i' \in \{1, \ldots, m_1\}$.

We now define

$$p_{1,2}(x) = \frac{1}{(\lambda_1-\lambda_2)^{s_2}\cdots(\lambda_1-\lambda_q)^{s_q}}\,(x-\lambda_1)^2(x-\lambda_2)^{s_2}\cdots(x-\lambda_q)^{s_q}.$$

Using Lemma 2, we see that

$$p_{1,2}(J^i_{\lambda_1}) = \begin{pmatrix} 0 & 0 & 1 & d_1(1) & \cdots & d_1(m-3) & d_1(m-2) \\ & 0 & 0 & 1 & d_1(1) & \cdots & d_1(m-3) \\ & & & \ddots & \ddots & \ddots & \vdots \\ & & & & 0 & 1 & d_1(1) \\ & & & & & 0 & 1 \\ & & & & & \cdots & 0 \\ & & & & & \cdots & 0 \end{pmatrix}.$$

In general, we define

$$p_{1,k}(x) = \frac{1}{(\lambda_1 - \lambda_2)^{s_k} \cdots (\lambda_1 - \lambda_q)^{s_q}} (x - \lambda_1)^k (x - \lambda_2)^{s_2} \cdots (x - \lambda_q)^{s_q}.$$

Lemma 2 shows that $p_{1,k}(J^i_{\lambda_1})$ will have 1's in its k^{th} diagonal, $d_1(1)$ in its $(k+1)^{\text{st}}$, $d_1(2)$ in its $(k+2)^{\text{nd}}$, etc. We define

$$p_1(x) = p_{1,1}(x) - \alpha_1(2)p_{1,2}(x) - \alpha_1(3)p_{1,3}(x) - \cdots - \alpha_1(s_1 - 1)p_{1,(s_1-1)}(x), \tag{6}$$

where α_1 is a function defined recursively by

$$\begin{array}{rcl} \alpha_1(2) & = & d_1(1), \\ \alpha_1(3) & = & -\alpha_1(2)d_1(1) + d_1(2), \\ & \vdots & \\ \alpha_1(n) & = & -\alpha_1(n-1)d_1(1) - \alpha_1(n-2)d_1(2) - \cdots - \alpha_1(2)d(n-2) + d_1(n). \end{array} \tag{7}$$

This polynomial has been constructed so that $p_1(J^i_{\lambda_1})$ will be the nilpotent part $N^i_{\lambda_1}$ of $J^i_{\lambda_1}$ for each i. Hence,

$$p_1(J) = \begin{pmatrix} N^1_{\lambda_1} & & & & & \\ & \ddots & & & & \\ & & N^{m_1}_{\lambda_1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

In a similar fashion, we construct polynomials $p_k(x)$ for each λ_k having a block of at least size 2. If λ_i has only one-blocks, then we let $p_i(x) = 0$. Finally, we let

$$p(x) = \sum_{k=1}^{q} p_k(x).$$

The construction guarantees that $p(J) = N_J$, the nilpotent part of J.

The hard part is over. We have shown that if J is the Jordan canonical form of A such that $A = QJQ^{-1}$, and $J = D_J + N_J$ where D_J is diagonal and N_J nilpotent, then there exists a polynomials $p(x), q(x)$ such that $p(A) = QN_JQ^{-1}$ and $q(A) = QD_JQ^{-1}$. To complete the proof of the theorem, we only need to show that the decomposition in the hypothesis is unique. That is, that any D_A and N_A are in fact QD_JQ^{-1} and QN_JQ^{-1} respectively. (We note that decompositions $A = D_A + N_A$ where the nilpotent and diagonalizable matrices are not required to commute are not unique.)

Lemma 5. *Let* $A \in M_{n\times n}(\mathbb{C})$ *and let* J *be its Jordan canonical form such that* $A = QJQ^{-1}$*. Let* $J = D_J + N_J$ *be* J*'s decomposition into the sum of a diagonal matrix* D_J *and a nilpotent matrix* N_J*. Suppose that* $A = D_A + N_A$ *is a decomposition of* A *where* D_A *is diagonalizable,* N_A *is nilpotent, and* $D_A N_A = N_A D_A$*. Then* $D_A = QD_JQ^{-1}$ *and* $N_A = QN_JQ^{-1}$*.*

Proof. We can write $A = QD_JQ^{-1} + QN_JQ^{-1}$. The construction shows that there exists polynomials $p(x), q(x)$ such that $p(A) = QN_JQ^{-1}$ and $q(A) = QD_JQ^{-1}$. Since D_A commutes with A, it commutes with QD_JQ^{-1}; and since N_A commutes with A, it commutes with QN_JQ^{-1}. In particular, we can find an invertible matrix P that simultaneously diagonalizes D_A and QD_JQ^{-1}; that is, PD_AP^{-1} and $P(QD_JQ^{-1})P^{-1}$ are diagonal [3, §6.5]. Therefore, $PD_AP^{-1} + PN_AP^{-1} = P(QD_JQ^{-1})P^{-1} + P(QN_JQ^{-1})P^{-1}$. Because, QN_JQ^{-1} and N_A commute, $P(QN_JQ^{-1})P^{-1} - PN_AP^{-1}$ is nilpotent. Hence, $PD_AP^{-1} - P(QD_JQ^{-1})P^{-1}$ is both diagonal and nilpotent; therefore it must be zero. Therefore, $D_A = (QD_JQ^{-1})$ and $N_A = (QN_JQ^{-1})$. ☐

5. Other Observations

Theorem 1 allows us to make the following definition.

Definition 1. Let $A \in M_{n\times n}(\mathbb{C})$ and write $A = D_A + N_A$ where D_A is diagonalizable, N_A is nilpotent, and $D_AN_A = N_AD_A$. We call D_A the *diagonalizable part* and N_A the *nilpotent part* of A.

Theorem 1 gives an immediate (albeit slight) improvement of the well known linear algebra result used in the proof of Lemma 5.

Corollary 1. *Let* $A, B \in M_{n\times n}(\mathbb{C})$*. If* $AB = BA$*, then the diagonalizable parts of* A *and* B *are simultaneously diagonalizable.*

Proof. The diagonalizable parts of A and B are polynomials in A and B respectively. Thus, they commute and are simultaneously diagonalizable. ☐

The converse of the corollary is false; one can find a counter example in $M_{2\times 2}(\mathbb{C})$.

It is not clear that the construction presented herein produces the "simplest" polynomial with the desired property. The next theorem proves that it does.

Theorem 2. *Suppose that* $A \in M_{n\times n}(\mathbb{C})$ *is not a diagonalizable matrix. Then the polynomial* $p(x)$ *constructed herein is the polynomial of minimal degree such that* $p(A) = N_A$*, the nilpotent part of* A*. (If* A *is diagonalizable, then its Jordan canonical form is diagonal, and* $p(x)$ *is the minimal polynomial of* A*. Thus it is unique up to a constant multiple.)*

Proof. We assume that $A \in M_{n\times n}(\mathbb{C})$ has Jordan canonical form as in (3); its Jordan canonical form has at least one block of size ≥ 2. Suppose $p(x)$ is the polynomial given by our construction and that $r(x)$ is another polynomial such that $r(A) = p(A) = N_A$, the nilpotent part of A. Since $r(A) - p(A) = 0$, $r(x) - p(x)$ is a multiple of the minimal polynomial of J. The minimal polynomial has degree $s_1 + s_2 + \cdots + s_q$ [3, §7.1]. Our construction gives a polynomial of degree at most $s_1 + s_2 + \cdots + s_q - 1$. Therefore, $\text{degree}(r(x)) > \text{degree}(p(x))$. ☐

The construction shows that the polynomial $p(x)$ always will have $\prod_{i=1}^{q}(x-\lambda_i)$ as a factor. We can say a bit more. We state the next result without our fairly easy proof.

Corollary 2. *Let $A \in M_{n\times n}(\mathbb{C})$ and let $p(x)$ be the polynomial of minimal degree such that $p(A) = N_A$, the nilpotent part of A. Then*

$$p(x) = \Big(\prod_{i=1}^{q}(x-\lambda_i)\Big)s(x),$$

where $\lambda_1, \ldots, \lambda_q$ are the distinct eigenvalues of A and such that $s(x)$ is relatively prime to $m(x)$, the minimal polynomial of A.

6. Possibilities for Further Work

The construction presented herein does not provide a closed form formula (in terms of the eigenvalues and size of the blocks) for the polynomial $p(x)$. Does one exist in general? We have been able to find such a formula for certain very special cases (e.g., for matrices with only Jordan one- or two-blocks). It would be interesting to find a closed form formula for $p(x)$ for other "uniform looking" J's. The problem reduces to finding the $s(x)$ in Corollary 2.

An easier problem is to find the recurrence (7) in terms of the eigenvalues of the matrix and the size of its Jordan blocks. We have been able to find it for matrices with two eigenvalues: λ_1 with blocks of size 1–3 and λ_2 with arbitrarily large blocks.

It would be interesting to analyze the "expense" of the algorithm for computing $p(x)$ presented herein. Theorem 2 shows that our $p(x)$ is the polynomial that we are after. Is there a faster way to get to it than via our construction? (We note that we are asking the question in the case where we know the Jordan form of A.)

Although all the results presented herein are stated for matrices in $A \in M_{n\times n}(\mathbb{C})$, the arguments are true for matrices in $M_{n\times n}(\mathbb{F})$, where $\mathbb{F}$ is a field over which the characteristic polynomial of A splits (e.g., algebraically closed). Is the construction or formula for $p(x)$ easier over other fields? Suppose that $A \in M_{n\times n}(\mathbb{R})$ but that at least one of its eigenvalues is in $\mathbb{C}$. When (if ever) is the polynomial $p(x)$ in $P(\mathbb{R})$?

As mentioned in the Introduction, the semi-constructive proof of Theorem 1 given in [3] relies on the Euclidean algorithm applied to certain polynomials. How is our construction related to (perhaps equivalent to) the Euclidean algorithm for these polynomials?

References

[1] N. Burgoyne & R. Cushman, "The Decomposition of a Linear Mapping", *Linear Algebra and its Applications*, **8** (1974), 515–519.

[2] S. H. Friedberg, A. J. Insel & L. E. Spence, *Linear Algebra*, 3rd ed., Prentice-Hall, New Jersey, 1997.

[3] K. Hoffman & R. Kunze, *Linear Algebra*, 2nd ed., Prentice-Hall, New Jersey, 1971.

[4] D. Schmidt, "Construction of the Jordan decomposition by means of Newton's method", *Linear Algebra and its Applications*, **314** (2000), 75–89.

Department of Mathematics, Loyola Marymount University, Los Angeles, CA 90045

E-mail address: `hmedina@lmu.edu`