THE CAYLEY-HAMILTON AND JORDAN NORMAL FORM THEOREMS

GABRIEL DAY

Abstract. We present three proofs for the Cayley-Hamilton Theorem. The final proof is a corollary of the Jordan Normal Form Theorem, which will also be proved here.

Contents

1. Introduction

The Cayley-Hamilton Theorem states that any square matrix satisfies its own characteristic polynomial. The Jordan Normal Form Theorem provides a very simple form to which every square matrix is similar, a consequential result to which the Cayley-Hamilton Theorem is a corollary.

In order to maintain the focus of the paper on the Cayley-Hamilton Theorem (often abbreviated C-H here), we anticipate that the reader has prior knowledge of basic linear algebra in order to omit a number of tedious definitions and proofs. Specifically, some understanding of linear combinations, linear independence and dependence, vector spaces, bases for vector spaces, direct sums, eigenvalues/vectors, and the determinate operation, as well as the Fundamental Theorem of Algebra, is presumed.

Throughout the paper, we always presume that the vector spaces with which we are dealing are of finite dimension. We also assume that our field of scalars is the complex numbers.

One object of particular importance to this paper, and much of linear algebra as a whole, is that of an operator. An **operator** is a linear transformation from a vector space to itself. Operators can be represented by square matrices, and this paper will often refer to an operator and its matrix interchangeably.

Each of the three following sections contains one proof of C-H, employing different methods. In Section 2, the key result is the representation of every square

Date: September 8, 2017.

matrix by an upper triangular matrix. While this lemma is also used in Section 3, the proof presented there relies on analysis, namely, the density of diagonalizable matrices among all matrices. The third proof follows from the Jordan Normal Form Theorem.

2. Proof of the Cayley-Hamilton Theorem Using Generalized Eigenvectors

Our first proof of the Cayley-Hamilton Theorem, originally found in Axler's *Linear Algebra Done Right*, is founded on an extension of the basic concepts of eigenvalues and eigenvectors. The definition of the characteristic polynomial here uses these "generalized eigenvectors," which we will define below. Crucial to this proof is that every matrix may be represented as upper triangular with respect to some basis.

Several lemmas are required in order to reach the proof. First, we must guarantee that every operator has at least one eigenvalue.

Lemma 2.1. *Every nonzero operator on a complex vector space has an eigenvalue.*

Proof. Suppose that V is a complex vector space with $\dim V = n > 0$. Let T be an operator on V. Choose $v \neq 0 \in V$. We know that the list of vectors

$$(v, Tv, T^2v, \cdots, T^nv)$$

must be linearly dependent, as it contains $n+1$ vectors, while the dimension of V is n. Thus there exist scalars $a_0, a_1, \cdots, a_n \in \mathbb{C}$, not all of which equal zero, such that

$$0 = a_0v + a_1Tv + a_2T^2v + \cdots + a_nT^nv. \tag{2.2}$$

The Fundamental Theorem of Algebra ensures that there exists a largest nonzero integer m such that the coefficients $a_0, \cdots a_m$ are also the coefficients for a polynomial with m roots, $\lambda_1, \cdots, \lambda_m \in \mathbb{C}$. That is, for all $x \in \mathbb{C}$,

$$a_0 + a_1x + a_2x^2 + \cdots, +a_mx^m = c(x-\lambda_1)\cdots(x-\lambda_m), \tag{2.3}$$

where $c \in \mathbb{C}$ is a nonzero constant.

Using the result of (2.2) and the right hand side of (2.3), we see that

$$0 = c(T-\lambda_1)\cdots(T-\lambda_m)v.$$

Thus, for at least one j, $T - \lambda_j I$ is not injective. So T has an eigenvalue.

□

Using this result, we prove that every operator is upper triangular with respect to some basis, a lemma critical to this and later proofs of C-H. First, however, it is useful to note the equivalence of the following statements:

Remark 2.4. *For an operator T on a vector space V with basis $(v_1, \cdots, v_n)$, the matrix of T with respect to $(v_1, \cdots, v_n)$ is upper triangular if, and only if*

$$Tv_k \in span(v_1, \cdots, v_k)$$

for each $k = 1, \cdots, n$.

This equivalence follows from the definitions of upper triangular matrix and the matrix for an operator with respect to a basis.

Now we prove the crucial lemma for this and the following proof of C-H.

Lemma 2.5. *Suppose V is a complex vector space and T is an operator on V. Then T is represented by an upper triangular matrix under some suitable basis.*

Proof. We prove by induction on $\dim V$. The result clearly holds for $\dim V = 1$. Suppose that it holds for all vector spaces with dimension less than that of V, which is greater than 1. Let λ be the eigenvalue for T whose existence is guaranteed by the previous lemma, and define

$$U = \mathrm{Im}(T - I\lambda).$$

The kernel of $T - \lambda I$ at least contains the eigenvector corresponding to λ, so $\dim U < \dim V$ follows from the Rank-Nullity Theorem. To show the T-invariance of U, simply take $u \in U$.

$$Tu = (T - \lambda I)u + \lambda u,$$

and since each term of this decomposition is in U, $Tu \in U$.

We conclude from these last statements that $T|_U$ is an operator on U. Thus by the inductive hypothesis, there is some basis $(u_1, \cdots, u_m)$ of U with respect to which the matrix of $T|_U$ is upper triangular. We see from (2.4) that this implies that for all k,

$$Tu_k = T|_U u_k \in \mathrm{span}(u_1, \cdots, u_k).$$

Extend $(u_1, \cdots, u_m)$ to a basis for V, $(u_1, \cdots, u_m, v_1, \cdots, v_n)$. For each j,

$$Tv_j = (T - \lambda I)v_j + \lambda v_j. \tag{2.6}$$

By the definition of U, $(T - \lambda I)v_j \in U = \mathrm{span}(u_1, \cdots, u_m)$. Thus the linear combination in (2.6) gives that

$$Tv_j \in \mathrm{span}(u_1, \cdots, u_m, v_j) \subset \mathrm{span}(u_1, \cdots, u_m, v_1, \cdots, v_j).$$

Employing (2.4) again, we conclude that the matrix of T with respect to the constructed basis is upper triangular. □

Next, we introduce the concepts of generalized eigenvectors, multiplicities of eigenvalues, and the characteristic polynomial.

Definition 2.7. A vector $v \in V$ is a **generalized eigenvector** of the operator T to eigenvalue λ if

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

So the set of generalized eigenvectors to an eigenvalue λ is the union of $\ker(T - \lambda I)^j$ for all values of j. Clearly, if $v \in \ker(T - \lambda I)^j$, then $v \in \ker(T - \lambda I)^{j+1}$, and the same is true for all higher powers. After the exponent $\dim V$ is reached, the kernels of all greater exponents are equal; otherwise the Rank-Nullity theorem would be violated. Thus, we may restate the set of generalized eigenvectors as equal to $\ker(T - \lambda I)^{\dim V}$.

Definition 2.8. The **eigensubspace** of eigenvalue λ is $\ker(T - \lambda I)^{\dim V}$.

Definition 2.9. The **multiplicity** of an eigenvalue λ is the dimension of the eigensubspace for λ, that is, $\dim \ker(T - \lambda I)^{\dim V}$.

A rather arduous but uncomplicated induction proof gives the following lemma:

Lemma 2.10. *An eigenvalue λ appears precisely its multiplicity number of times on the diagonal of an upper triangular matrix for T.*

As a result of this conclusion, if $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of T, listed with multiplicities, then an upper triangular matrix for T is written in the following way:

$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \tag{2.11}$$

Since each eigenvalue appears its multiplicity number of times in the diagonal, and the diagonal has length $\dim V$, if $U_1, \cdots, U_m$are the eigensubspaces of T, then

$$\dim V = \dim U_1 + \cdots + \dim U_m. \tag{2.12}$$

We will give here a brief proof of an important structural theorem, which, though not used in this first proof of C-H, intimately relates to the concepts just stated, and will be employed later on in the paper to prove the Jordan Normal Form Theorem.

Theorem 2.13. *Suppose that T is an operator over a vector space V, with distinct eigenvalues $\lambda_1, \cdots, \lambda_m$ and corresponding eigensubspaces $U_1, \cdots, U_m$. Then V is equal to the direct sum of these subspaces:*

$$V = U_1 \oplus \cdots \oplus U_m.$$

Proof. Because $U_1 + \cdots + U_m$ is a subset of of V, and (2.12) holds, this sum and V must be equal. So

$$V = U_1 + \cdots + U_m.$$

Along with (2.12), this is sufficient to conclude that the sum is direct. □

Definition 2.14. The polynomial $(x - \lambda_1)^{d_1} \cdots (x - \lambda_m)^{d_m}$ is the **characteristic polynomial** for T, where $\lambda_1, \cdots, \lambda_m$ are the eigenvalues of T and $d_1, \cdots, d_m$ denote their respective multiplicities. We denote the characteristic polynomial of T by f_T.

With that, we have arrived at the first proof of the Cayley-Hamilton Theorem.

Theorem 2.15. *Suppose that T is an operator on a complex vector space V. Then*

$$f_T(T) = 0.$$

Proof. Suppose that $(v_1, \cdots, v_n)$ is a basis for V with respect to which T has the form given in (2.11). To show that the matrix $f_T(T) = 0$, we need to show that $f_T(T)v_j = 0$ for all values of j. To do this, it suffices to show that

$$(T - \lambda_1 I) \cdots (T - \lambda_j I)v_j = 0, \tag{2.16}$$

as the polynomial by which v_j is multiplied is a factor of f_T.
We prove this by induction. The case of $j = 1$ is given by $Tv_1 = \lambda_1 v_1$, the definition of an eigenvector. Suppose that for $1 < j \leq n$, (2.16) holds.
The form in (2.11) gives that the jth column of $T - \lambda_j I$ will have zeros in the jth entry and all below, so $(T - \lambda_j I)v_j$ is a linear combination of $v_1, \cdots, v_{j-1}$.
By the induction hypothesis, applying $(T - \lambda_1 I) \cdots (T - \lambda_{j-1} I)$ to $(T - \lambda_j I)v_j$ gives zero. Therefore (2.16) is satisfied, and the proof is complete. □

3. Proof of the Cayley-Hamilton Theorem Using Density of Diagonalizable Matrices

Another definition of the characteristic polynomial for a matrix, which leads to a simple proof of C-H, makes use of the determinant. We presume the reader is aware of the definition of determinant and its basic properties.

The **characteristic polynomial** f_T of an operator T on a vector space V is

$$f_T(t) = \det(T - tI).$$

With this definition, C-H is stated precisely as in (2.15). This proof requires several steps. First, we prove C-H for operators with diagonal matrices, then diagonalizable matrices, and finally for all operators.

A matrix is **diagonal** if its entries are all zero, except possibly those on the diagonal. One can easily see that if $A = \operatorname{diag}(a_1, \cdots, a_n)$ and $B = \operatorname{diag}(b_1, \cdots, b_n)$ are two diagonal matrices,

$$A + B = \operatorname{diag}(a_1 + b_1, \cdots, a_n + b_n).$$

Likewise,

$$AB = \operatorname{diag}(a_1 b_1, \cdots, a_n b_n).$$

So if a polynomial f is applied to A, then

$$f(A) = \operatorname{diag}(f(a_1), \cdots, f(a_n)).$$

The determinate of a diagonal matrix is simply the product of the diagonal entries, so clearly diagonal matrices satisfy C-H.

Definition 3.1. Two matrices A and B are said to be **similar**, denoted $A \sim B$, if there exists some invertible matrix S such that

$$A = S^{-1}BS.$$

Stating that two matrices are "similar" is synonymous with saying that they are the same matrix with respect to different bases; the invertible matrix represents a change of basis. So (2.5) can be rephrased "every matrix over a complex vector space is similar to an upper triangular matrix."

Lemma 3.2. *If two matrices A and B are similar, then their characteristic polynomials are equal.*

Proof. Let f_A and f_B denote the characteristic polynomials of A and B, respectively. Since $A \sim B$, there exists an invertible matrix S such that $A = S^{-1}BS$. The following lines proceed from the definitions of characteristic polynomial and similarity, as well as some basic determinate properties.

$$f_A = \det(tI - S^{-1}BS) = \det(tS^{-1}S - S^{-1}BS) = \det(S^{-1}(tI - B)S) =$$

$$\det(S^{-1})\det(tI - B)\det(S) = \det(S^{-1}S)\det(tI - B) = \det(tI - B) = f_B.$$

□

Definition 3.3. A matrix is **diagonalizable** if it is similar to a diagonal matrix.

One more brief lemma is required to present the proof of C-H for diagonalizable matrices.

Lemma 3.4. *If g is a polynomial and A and B are similar matrices, then*

$$g(A) \sim g(B).$$

Proof. First, consider the case where $g(t) = t^k$.

$$g(A) = A^k = (S^{-1}BC)^k = (S^{-1}BS)\cdots(S^{-1}BS) = S^{-1}B^kS = S^{-1}g(B)S,$$

where the second to last equality comes from the cancellation of the inner pairs of S^{-1} and S to the identity. Thus $g(A) \sim g(B)$.

For an arbitrary polynomial g, a similar process to the above can be applied to each term. Clearly, coefficients will be unaffected, and we can factor out S^{-1} and S from each term to reach the desired conclusion. □

Theorem 3.5. *If A is diagonalizable, then $f_A(A) = 0$.*

Proof. $A \sim D$, where D is a diagonal matrix. As we proved above, $f_D(D) = 0$. From (3.2), we have that $f_A(t) = f_D(t)$. So $f_A(D) = 0$. We know from (3.4) that $f_A(A) \sim f_A(D)$. Thus

$$f_A(A) = S^{-1}f_A(D)S = S^{-1}0S = 0.$$

□

The next two results will establish a connection between eigenvalues and diagonalizability, a necessary ingredient for the density argument.

Theorem 3.6. *Eigenvectors to distinct eigenvalues are linearly independent.*

Proof. We induct on the number of eigenvectors. If there is only one eigenvector v_1, then $\lambda_1 \cdot v_1 = 0$ only if $\lambda_1 = 0$, as $v_1 \neq 0$.
Suppose that k eigenvectors to distinct eigenvalues are linearly independent.
Now suppose we have $k+1$ eigenvectors to distinct eigenvalues. Suppose

$$a_1v_1 + \cdots + a_kv_k + a_{k+1}v_{k+1} = 0, \tag{3.7}$$

where $a_1, \cdots, a_{k+1} \in \mathbb{R}$
Multiplying by λ_{k+1}, the eigenvalue for v_{k+1}, we get

$$\lambda_{k+1}a_1v_1 + \cdots + \lambda_{k+1}a_kv_k + \lambda_{k+1}a_{k+1}v_{k+1} = 0. \tag{3.8}$$

Multiplying (3.7) by A, we get

$$A(a_1v_1+\cdots+a_kv_k+a_{k+1}v_{k+1}) = \lambda_1a_1v_1+\cdots+\lambda_ka_kv_k+\lambda_{k+1}a_{k+1}v_{k+1} = 0, \tag{3.9}$$

the last equality following from the definition of eigenvectors and eigenvalues. Subtracting (3.8) from (3.9), we get the following equation, with the last terms canceling:
$(\lambda_1 - \lambda_{k+1})a_1v_1 + \cdots + (\lambda_k - \lambda_{k+1})a_kv_k = 0.$

Since all eigenvalues are distinct, for all i, $\lambda_i - \lambda_{k+1} \neq 0$. So each coefficient a_i could be multiplied by $\frac{1}{\lambda_i - \lambda_{k+1}}$ to get $a_1v_1+\cdots+a_kv_k = 0$, a linear combination of k

eigenvectors to distinct eigenvalues. Because these vectors are linearly independent by the inductive hypothesis, $a_i = 0$ for all i.

Thus, returning to (3.7), we find that $a_{k+1}v_{k+1} = 0$, and thus $a_{k+1} = 0$. Thus, the eigenvectors are linearly independent.

□

Lemma 3.10. *If a $k \times k$ matrix A has k linearly independent eigenvectors, then it is diagonalizable.*

Proof. Suppose A has k linearly independent eigenvectors $x_1, \cdots, x_k$. Let $S = [x_1, \cdots, x_k]$. By the definition of eigenvector, $AS = SD$, where D is defined as above. The columns of S are linearly independent, so it is invertable.
So $SDS^{-1} = ASS^{-1} = A$. Thus A is diagonalizable.

□

One more lemma is needed to prove the general case of C-H, employing a bit of analysis.

Lemma 3.11. *Diagonalizable matrices are dense in all complex, square matrices.*

Proof. Consider a non-diagonalizable matrix A. By (2.5), there exists upper-triangular matrix $B = \begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$ such that $A \sim B$. Take some $\epsilon > 0$. Consider the matrix B_ϵ produced by making slight variations to the diagonal entries of B by factors of ϵ, so that all of the diagonal entries of B_ϵ are mutually distinct. For example, if B is a 2×2 matrix with diagonal values of 1, then B_ϵ could have diagonal entries 1 and $1 - \frac{\epsilon}{2}$. Because B_ϵ has n distinct eigenvalues equal to the unique diagonal entries (we know these are eigenvalues from (2.11)), B_ϵ has n linearly independent eigenvectors and is thus diagonalizable. So as $\epsilon \to 0$, the B_ϵ go to B, and $B \sim A$. Thus diagonalizable matrices are dense among all complex square matrices. □

Finally, we can prove C-H using density of diagonalizable matrices:

Theorem 3.12. *For all matrices A, $f_A(A) = 0$.*

Proof. Consider a non-diagonalizable matrix A as above. For each ϵ,

$$f_{B_\epsilon}(B_\epsilon) = 0.$$

Since $B_\epsilon \to B$, $f_{B_\epsilon} \to f_B = f_A$, and taking the determinate is continuous (changing the values on the diagonal is simply changing the coefficients of the polynomial),

$$f_A(A) = \lim_{\epsilon \to 0} f_{B_\epsilon}(B_\epsilon) = 0.$$

□

4. The Jordan Normal Form Theorem

The Jordan Normal Form Theorem (JNFT) (also known as the Jordan Canonical Form) provides an answer to one of the essential questions of linear algebra: what is the simplest form in which we can write an arbitrary matrix? More precisely, with respect to some basis, what form for a matrix contains the most zero entries?

This important result is probably more consequential than C-H, and once we have it in hand a third and final proof for C-H is simple.

The proof of JNFT presented here, originally found in Axler, makes extensive use of nilpotent operators.

Definition 4.1. An operator (or matrix) A is **nilpotent** if there exists a positive integer k such that $A^k = 0$.

This proof of JNFT relies primarily on a lemma which guarantees that a basis for the vector space can be found using a nilpotent operator.

For a vector v and nilpotent operator N, let $m(v)$ denote the maximum nonnegative integer such that $N^{m(v)}v \neq 0$.

Lemma 4.2. *If an operator N over a complex vector space V is nilpotent, then there exist vectors $v_1, \cdots, v_k \in V$ such that*

(a) the list

$$(v_1, Nv_1, \cdots, N^{m(v_1)}v_1, \cdots, v_k, Nv_k, \cdots, N^{m(v_k)}v_k) \tag{4.3}$$

is a basis for V, and

(b) $(N^{m(v_1)}v_1, \cdots, N^{m(v_k)}v_k)$ is a basis for $\ker(N)$.

Proof. We prove by induction on $\dim V$. The $\dim V = 1$ case is trivial, as N must equal [0]. So we assume that the lemma holds for vector spaces of dimension less than that of V. Since N is nilpotent it is not injective, hence $\dim \mathrm{Im}N < \dim V$. Applying our inductive hypothesis (where we replace V with $\mathrm{Im}N$ and N with $N|_{\mathrm{Im}N}$), there is a basis $u_1, \cdots u_j \in \mathrm{Im}N$ such that

(i) $(u_1, Nu_1, \cdots, N^{m(u_1)}u_1, \cdots, u_j, Nu_j, \cdots, N^{m(u_j)}u_j)$ is a basis for $\mathrm{Im}N$, and

(ii) $(N^{m(u_1)}u_1, \cdots, N^{m(u_j)}u_j)$ is a basis for $\mathrm{Im}N \cap \ker N$.

Define $v_1, \cdots, v_j$ by $Nv_r = u_r$, for all r. Note that $m(v_r) = m(u_r) + 1$.

There exists a subspace W such that

$$\ker N = (\ker N \cap \mathrm{Im}N) \oplus W. \tag{4.4}$$

We may choose a basis $v_{j+1}, \cdots, v_k$ for W. Each of these is in $\ker N$ and thus $0 = m(v_{j+1}) = \cdots = m(v_m)$.

The list $v_1, \cdots v_k$ is our suspect list of vectors satisfying (a) and (b). Showing that the list of vectors in (4.3) is linearly independent involves applying N to a linear combination of all of these vectors, along with careful use of the $m(v_r)$ values.

We can clearly see from the bases we have defined that

$$\dim \ker N = k.$$

(i) gives that

$$\dim \mathrm{Im}N = \sum_{r=1}^{j}(m(u_r) + 1) = \sum_{r=1}^{j} m(v_r).$$

Combining these last two equations, we see that the list of vectors in (a) has length

$$\sum_{r=1}^{k}(m(v_r)+1) = k + \sum_{r=1}^{j} m(v_r)$$
$$= \dim\ker N + \dim \mathrm{Im} N = \dim V,$$

where the last equality holds by the Rank-Nullity Theorem. Since this list of vectors is linearly independent, the proof of (a) is complete.

Note that

$$(N^{m(v_1)}v_1, \cdots, N^{m(v_k)}v_k) = (N^{m(u_1)}u_1, \cdots, N^{m(u_j)}u_j, v_{j+1}, \cdots, v_k),$$

a list which (4.4) and (ii) show to be a basis for $\ker N$.

□

We can now move on to the statement and proof of the JNFT. First,

Definition 4.5. A **Jordan Basis** for an operator T on V is a basis for V with respect to which T is represented by

$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_i, known as a **Jordan block**, corresponds to an eigenvalue λ_i and has the form

$$\begin{bmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}.$$

That is, each Jordan block contains one copy of a particular eigenvalue in each diagonal entry, and the line above the diagonal contains all ones. Elsewhere, the block is zero.

Theorem 4.6. *If T is an operator on a complex vector space V, then there is a basis for V which is a Jordan basis for T.*

Proof. First, we prove the theorem for nilpotent operators. Consider a nilpotent operator N on V, and the basis whose existence is guaranteed by (4.2). For each j, when N is applied to $N^{m(v_j)}v_j$ the result is zero, and when N is applied to any other element of $(N^{m(v_j)}v_j, \cdots, Nv_j, v_j)$, the result is the previous element. If we consider the block of the matrix for N given by this list, applying N gives

$$\begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}. \tag{4.7}$$

Thus we can see that if we reverse the order of the list of vectors in (4.3), N is in Jordan Normal Form. So JNFT holds for nilpotent operators.

Now consider an operator T with distinct eigenvalues $\lambda_1, \cdots, \lambda_m$ and corresponding eigensubspaces $U_1, \cdots U_m$. From (2.13),

$$V = U_1 \oplus \cdots \oplus U_m.$$

Each $(T - \lambda_j I)|_{U_j}$ is nilpotent, a fact which arises from the definition of U_j: since U_j is the null space of a power of $T - \lambda_j I$, there is some power such that if $(T - \lambda_j I)|_{U_j}$ is raised to it, multiplying this matrix by any vector in U_j yields zero. So this restriction is nilpotent. Thus there exists a basis for U_j which is a Jordan basis for $(T - \lambda_j I)|_{U_j}$. Putting these together into a single basis, we get a Jordan basis for T.

□

As previously stated, C-H follows almost directly from JNFT. So here is one last proof of C-H:

Corollary 4.8. *If T is an operator on a vector space V, then*

$$f_T(T) = 0.$$

Proof. By (3.2) and (3.4), we may assume T is in Jordan Normal Form. We use Axler's definition of the characteristic polynomial, (2.14). Each binomial multiplied to get $f_T(T)$ is of the form $T - \lambda_j I$, so the Jordan block for λ_j has the form (4.7). Taking powers of this Jordan block, we may observe that the diagonal line of ones recedes and disappears. Thus this Jordan block goes to zero when taken to the d_j power, the multiplicity of λ_j. Since the power to which $T - \lambda_j$ is raised is d_j, this Jordan block is annihilated by the characteristic polynomial. Since this holds for each Jordan block, the entire matrix is annihilated.

□

The power and usefulness of the Cayley-Hamilton Theorem arises from the fact that it holds not only over the complex numbers, as we have proved, but over all fields. We hope that through the proofs presented here, readers have a better understanding of the interactions between eigenvalues, the determinant, and matrix form, and of linear algebra as a whole.

Acknowledgments. Many thanks to my mentor, Jingren Chi, for his suggestions for the topic of my paper and reading material, and helpfulness in editing and improving my work. I would also like to thank Dr. Edwin Ihrig for discussing the paper with me and offering his suggestions. I owe much of my understanding of linear algebra to Professor Babai and his apprentice program this summer, in which he provided the second proof presented in this paper. Finally, thank you to Professor Peter May for organizing this excellent REU.

References

[1] Axler, S, *Linear Algebra Done Right, 2nd ed.* Springer, New York, 1997, 81, 83-85, 164, 167-168, 171-175, 183-187pp.

[2] Babai, L., Tulsiani, M., 2013: Apprentice Program - Summer 2013 REU, Class 7/25. Accessed 27 July 2017. Available online at `http://ttic.uchicago.edu/~madhurt/courses/reu2013/class725.pdf`.

[3] Wikipedia, the Free Encyclopedia 2017: Jordan normal form. Accessed 3 August 2017. Available online at `https://en.wikipedia.org/wiki/Jordan_normal_form#cite_note-1`.