ISSN: 2278-0181
ICADEMS - 2017 Conference Proceedings

Different Approaches to Prove Cayley-Hamilton Theorem

Dr. Paramjeet [1] , J K Narwal [2]
Assistant Professor,
Ganga Institute of Technology and Management,
Kablana, Jhajjar[1]

Abstract: **This paper covers different approach to prove the Cayley – Hamilton Theorem using different derivation of the determinant of a matrix in conjuction *i_{th}* elementary graph theory. In this paper various alternative proofs have been discussed via Schur's Triangularization, a variation of topological proofs, combinatorial proof. The Topological proof uses continuity properties and matrix norms.**

Keywords: Matrices, Partial Permutation, Characteristics Polynomial, Topology, Eigen values.

I. INTRODUCTION

Defintion 1.1. If A is an n×n matrix ,then the characteristic polynomial of A is defined to be $P_A(x)$= det(xI-A). This is a polynomial in x of degree n with leading term x^n . the constant term c_0 of a polynomial q(x) is interpreted as c_0 I in q(A).

Theorem 1.2 (Cayley – Hamilton Theorem). If A is an n×n matrix ,then $p_A(A)=0$, the zero matrix.

Theorem 1.3 If q≠0 is a quaternion of the form q-= a+bi+cj+dk with a,b,c,d, being real , then $q^2 - 2aq + (a^2+b^2+c^2+d^2)=0$

$$q^{-1} = \frac{\bar{q}}{|q|} = \frac{a-bi-cj-dk}{a^2+b^2+c^2+d^2} = \frac{2a}{a^2+b^2+c^2+d^2} - \frac{a+bi+cj+dk}{a^2+b^2+c^2+d^2}$$

$$= \frac{1}{a^2+b^2+c^2+d^2}(2a-q) \Rightarrow a^2+b^2+c^2+d^2 = 2aq - q2$$

$$\Rightarrow q2 - 2aq + (a^2+b^2+c^2+d^2) = 0$$

If one represents a quaternion q= a+bi+cj+dk as a matrix,

$$A = \begin{bmatrix} a+bi & c+di \\ -c+di & a-bi \end{bmatrix},$$

$P_A(A) = A^2 - 2aA + (a^2+b^2+c^2+d^2)I = 0,$ and the polynomial given in Theorem 1.3 is characteristic polynomial of A

II. GENERALIZATION OF CAYLEY HAMILTON THEOREM

Theorem 2.1 (Cayley-Hamilton Theorem). For any n × n Matrix A, $P_A(A)=0$.

Proof. Let D(x) be the matrix with polynomial entries *D(x)= adj(xI_n-A)*, So *D(x)(xI-A)= det(xI_n-A)I_n*. Since each entry in D(x) is the determinant of an *(n-1) X((n-1)* submatrix of *(xI_n-A)*, each entry of *D(x)* is a polynomial of degree less than or equal to n-1. It folowws that there exist matrices D_0, D_1,………D_{n-1} with entries from C such that $D(x) = D_{n-1} x^{n-1} + \dots + D_1 x + D_o$. Then the matrix equation follows

$$\det(xI_n-A)\, I_n = (x I_n-A)\, adj(xI_n-A) = (xI_n-A)D(x)$$

Substituting $p_A(x)= \det(xI_n-A)$, (and using the fact that scalars commute with matrix)

$$X^nI_n + b_{n-1}x^{n-1}I_n + \dots\dots + b_1xI_n + b_0I_n$$

$$= p_A(x)\, I_n = \det(xI_n-A)I_n = (xI_n-A)adj(xI_n-A)$$

$$= (xI_n-A)(x^{n-1}D_{n-1} + \dots + xD_1 + D_o)$$

$$= x^{nD}{}_{n-1} - x^{n-1}AD_{n-1} + x^{n-1}D_{n-2} - x^{n-2}AD_{n-2} + \dots + xD_0 - AD_o$$

$$= x^{nD}{}_{n-1} + x^{n-1}(-AD_{n-1} + D_{n-2}) + \dots + (-AD_1 + D_o) - AD_o$$

Since two polynomials are equal if and only if their coefficients are equal , the coefficient matrices are equal ; that is , $I_n=D_{n-1}$, $b_{n-1}I_n= (-AD_{n-1}+D_{n-2}),\dots,b_1I_n=-(AD_1+D_0)$, and $b_0I_n=-AD_0$. This means that A may be substituted for the variable x in the equation (2.1) to conclude

$$P_A(A) = A^n + b_{n-1} A^{n-1} + \dots + b_1A + b_0I_n$$

$$= A^nD_{n-1} + A^{n-1}(-AD_{n-1} + D_{n-2}) + \dots + A(-AD_1 + D_0) - AD_0$$

$$= A^nD_{n-1} - A^nD_{n-1} + A^{n-1}D_{n-2} - A^{n-1}D_{n-2} + \dots\dots AD_0 - AD_0 = 0$$

This proves the theorem

III. PROOF THROUGH SCHUR'S TRIANGULARIZATION

This proof synthesizes work of Issai Schur, According to this, If $S_1S_2,\dots,S_n$ are n × n upper –triangular matrices such that *(i,i)* element of S_i is zero for all I, then $S_1S_2\dots\dots S_n=0$

*Proof. (*Induction of n)

For n=1 , there is nothing to be proved since $S_1=0$

For n=2 , $S_1=\begin{bmatrix}0 & x\\0 & y\end{bmatrix}$, $S_2=\begin{bmatrix}u & v\\0 & 0\end{bmatrix}$, where x,y,u,and v are scalars . It is clear that $S_1S_2=0$

Assume now that the theorem is true for some integer m, then $S1S2_{\ldots\ldots\ldots}S_mS_{m+1}= T\, S_{m+1},$

Where, $S_1=\begin{bmatrix}T_i & a_i\\ \bar{0} & x_1\end{bmatrix}$, ………., $S_m=\begin{bmatrix}T_m & a_m\\ \bar{0} & x_m\end{bmatrix}$

And $T_1,\ldots\ldots\ldots T_m \in M_m$ are upper triangular matrices such that the (i,i) element of T_i is zero.

Then,

$T=\ S_1S_2\ldots\ldots. S_m= \begin{bmatrix}T_1T_{2\ldots\ldots\ldots}T_M & u_m\\ \bar{0} & x\end{bmatrix}=\begin{bmatrix}0_m & u_m\\ \bar{0} & x\end{bmatrix}$ and $S_{m+1}=\begin{bmatrix}A_m & t_m\\ \bar{0} & 0\end{bmatrix}$

Where $0_m\in M_m$, $A_m\in M_m$ is upper triangular , u_m and t_m are vector columns of order m X 1, $\bar{0}$ is zero row of order 1 X m , and x is a scalar. The Proof is complete.

Main Theorem.(Cayley- Hamilton Theorem).

Let $p_A(t)$ be the characteristic polynomial of $A\in M_m$. Then $P_A(A)=0$

Proof. Since $p_A(t)$ is of degree n with leading coefficient 1 and the roots of $p_A(t)$ are precisely the eigen values λ_1…………., λ_n of A, counting multiplicities , factor $p_A(t)$ as $P_A(t) =(t- \lambda_1)(t- \lambda_2)\ldots\ldots\ldots(t- \lambda_m)$

Using Schur's Theorem , write A as $A= UTU^*$

Where, T is upper triangular with λ_i in the ith diagonal position , i =1,……n. The theorem follows.

$P_A(A)$

$= P_A(UTU^*)=(UTU^*- \lambda_1I)(UTU^*- \lambda_2I)\ldots\ldots...(UTU^*- \lambda_nI)$

$=[U(T- \lambda_1I)U^*][U(T- \lambda_2I)U^*]\ldots\ldots\ldots\ldots [U(T- \lambda_nI)U^*]$

$=U[(T- \lambda_1I)(T- \lambda_2I)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (T- \lambda_nI)]U^*=0$

The last equality follows from theorem 2.1.

IV. TOPOLOGICAL PROOF

This is most concise alternate proof to the Frobenius's proof.

Theorem 3.1. The set $D_n= \{A\epsilon M_n | A \text{ is diagonalizable}\}$ of diagonalizable matrices dense in M_n

Proof. Fix $\varepsilon > 0$. It is sufficient to show that , given any matrix $A\epsilon M_n$, there exists diagonalizable matrix B such that $|||A-B|||_2<\varepsilon$

Where $|||A-B|||_2$ is the Frobenius norm given by

$$|||A-B|||_2= \left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

Given $A\epsilon M_n$,let $A=UTU^*$ where U is unitary and T is upper –Triangular , possible Schur's Triangularization theorem. Define $B=A+ UCU^*$, where

$$C_{fg}=\begin{Bmatrix}0 & if\ f\neq g\\ \frac{\varepsilon\delta}{2f\sqrt{n}} & if\ f=g\end{Bmatrix}$$

Choose δ>o , so that B will have distinct eigen values, hence B will be diagonalizable .Let $\sigma(A)=\{\lambda_f, f=1,2,\ldots\ldots\ldots n\}$ and $\lambda_{i+1}\geq\lambda_f$. Define δ as

$$\delta < \begin{matrix}min\\ \lambda_f\neq\lambda_g\end{matrix}\left\{\frac{(\lambda_g-\lambda_f)2\sqrt{n}}{\varepsilon}\left(\frac{1}{f}-\frac{1}{g}\right)\right\}\ and\ \delta<1$$

If $\lambda_h=\lambda_i$

$$\lambda_h+\frac{\varepsilon\delta}{2h\sqrt{n}}=\lambda_i+\frac{\varepsilon\delta}{2h\sqrt{n}}\neq\lambda_i+\frac{\varepsilon\delta}{2i\sqrt{n}}$$

If $\lambda_h\neq\lambda_i$

$$\frac{(\lambda_h-\lambda_i)2\sqrt{n}}{\varepsilon\left(\frac{1}{h}-\frac{1}{i}\right)}\geq\begin{matrix}min\\ \lambda_g\geq\lambda_f\end{matrix}\left\{\frac{(\lambda_g-\lambda_f)2\sqrt{n}}{\varepsilon}\left(\frac{1}{f}-\frac{1}{g}\right)\right\}>\delta$$

$$\Rightarrow(\lambda_i-\lambda_h)>\frac{\varepsilon\delta}{2i\sqrt{n}}\left(\frac{1}{h}-\frac{1}{i}\right)$$

$$\Rightarrow\lambda_i+\frac{\varepsilon\delta}{2i\sqrt{n}}>\lambda_h+\frac{\varepsilon\delta}{2h\sqrt{n}}$$

The Diagonal entries of T+C are distinct from the choice of , on B has distinct eigen values and is diagonalizable. Then

$|||A-B|||_2=|||A-(UCU^*|||_2$

$=|||C|||_{2,}$ because $|||A-B|||_2$ is unitarily invariant

$=\sqrt{\sum_{f=1}^{n}\left(\frac{\varepsilon\delta}{2f\sqrt{n}}\right)^2}$ because C is Diagonal $\leq \sqrt{\sum_{f=1}^{n}\left(\frac{\varepsilon\delta}{2}\right)^2}$

$=\frac{\varepsilon\delta}{2}<\varepsilon$ since every $\delta<1$

Example 3.1 If A is diagonalizable , then $p_A(A)= 0$

Proof. $A=PDP^{-1}$, where P is invertible.

$$D=\begin{bmatrix}\lambda_1 & & o\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix}$$

And λi *are the eigen value of* A. *Then,*

$P_A(A) = (A- \lambda_1 I)(A- \lambda_2 I)\ldots\ldots\ldots.(A- \lambda_n I)$

$= P(D- \lambda_1 I)(D- \lambda_2 I)\ldots\ldots\ldots.(D- \lambda_n I)P^{-1}$

$= P.0.P^{-1} = 0$

Main Theorem.(Cayley Hamilton).*If* $P_A(t)$ *is the characteristics polynomial of A then* $P_A(A) = 0$

Proof. Let P_n be the space of polynomials of degree n or less, with the Zariski topology. From Example 3.1 , the Cayley-Hamilton theorem is proved for all diagonalizable matrices. The mapping Ω: M_n X $P_N \rightarrow M_n$ given by $\Omega(A,f(x))=f(A)$ is continuous , and the mapping $\phi: M_n \rightarrow M_n$ X P_N given by $\phi(A) =(A, pA(x))$ is continuous . Hence the composition

$$\Omega o\, \phi: M_n \rightarrow M_n,$$

Which is given by $\Omega o\, \phi(A)= p_A(A)$ is continuous. This mapping is identically zero on a dense subject of M_n, so by continuity vanishes everywhere.

V. COMBINATORIAL PROOF OF CAYLEY-HAMILTON THEOREM

3.1.1. Partial permutation σ

A partial permutation of {1,…,n}is a bijection σ of a subset of {1,…,n} onto itself. The domain of σ is denoted by dom σ.The cardinality of dom σ is called the degree of σ and is denoted by$|\sigma|$.

A complete permutation whose domain is {1,….,n}. If σ is a partial permutaion of {1,...,n} , then the completion of σ , denoted $\hat{\sigma}$, is the complete permutation of {1,...,n} defined by

*Defintion 3.1.***1**. $\hat{\sigma}(i) = \begin{Bmatrix} \sigma(i) & if\ i \in dom\ \sigma \\ i & if\ i \in \{1,\ldots,n\} \backslash dom\sigma \end{Bmatrix}$

Definition 3.1.2. The signature of a complete permutation $\hat{\sigma}$ *denoted* $sgn(\hat{\sigma})$*, is +1if the total number of inversion in* $\hat{\sigma}$ *is even and -1 if that number is odd.*

Definition 3.1.3. The signature of a complete permutation σ *, denoted* $sgn(\sigma)$*, is defined by* $sgn(\sigma)=(-1)^{|\sigma|}sgn(\hat{\sigma})$.

The characteristics polynomial , $p_A(x)$, of a matrix is the sum o certain products of elements of that matrix and poers of x . It is shonbelow that the pairs of indices (i,j) appearing in one of these products an be described using partial permutations. There is a relation between the elements of a given product. Namely , their subscripts are ordered pairs (i, σ(i)) where σ is a partial permutation of {1,…n} and $if\ i \in dom\,\sigma$. For example , if $A \in M_3$, the terms of $p_A(x)$ are:

$\lvert\sigma\rvert = 0$	$\lvert\sigma\rvert = 1$	$\lvert\sigma\rvert = 2$	$\lvert\sigma\rvert = 3$
x^3	$-a_{11}x^2$ $-a_{22}x^2$ $-a_{33}x^2$	$a_{11}a_{22}\,x$ $a_{11}a_{33}\,x$ $a_{22}a_{33}\,x$ $-a_{12}a_{21}\,x$ $-a_{13}a_{31}\,x$ $-a_{23}a_{32}\,x$	$-a_{11}a_{22}\,a_{33}$ $a_{11}a_{23}\,a_{32}$ $a_{22}a_{13}\,a_{31}$ $a_{33}a_{12}\,a_{21}$ $-a_{12}a_{23}\,a_{31}$ $-a_{13}a_{32}\,a_{21}$

Table 1.1: Terms of $p_A(x)$ of M_3

Note that here the coeffiecient of x is the sum of signed terms whose indices come from the 6 partial permutations of {1,2,3} of order 2:

$a_{11}a_{22}$ corresponds to σ_1=(1)(2),

$a_{12}a_{21}$ corresponds to σ_4=(1,2), etc.

It will be shown in general that each partial permutation of {1,…,n} of order q yields a signed term , which is one of he summands in the coefficient of x^{n-q}. Graph theory will help one visualize the partial permutations involved. Let $dom\ \sigma$ be vertices and the ordered pairs (i, σ(i)) ,where $i \in dom\ \sigma$, be directed edges. The bijective properties of σ mandate that this graph is a graph of adjoined cycles.

This transition from combinatorics to graph theory allows one to use an elegant set an example to prove the Cayley – Hamilton Theorem. To use this transiton $p_A(x)$ must be described in some detail.

3.1.2 Positive and negative parts of the charactristics polynomial

Let A = $(a_{ij}) \in M_n$ overC. As defined , the characteristics polynomial of A is :

$$P_A(x) = \det(xI - A) \sum_{\sigma \in s_n} sgn(\sigma) \prod_{i=1}^{n} b_{i,\sigma(i)}$$

Where, $b_{ij} = \begin{Bmatrix} -a_{ij} & if\ i \neq j \\ x - a_{ij} & if\ i = j \end{Bmatrix}$

and where the summation is overall complete permutations $\hat{\sigma}$ of {1,…,n}. It follows that the coefficient of x^{n-q} in $p_A(x)$ is

$$\sum_{|\sigma|=q} sgn(\sigma) \prod_{i \in dom\ \sigma} a_{i,\sigma(i).}$$

To see this, notice that from Equation 3.1 , the coefficient of x^{n-q} comes from the terms $b_{i,n(i)} = b_{ii} = x - a_{ii}$ for n-q of the indices, and $b_{i,n(i)} = -a_{i,n(i)}$ for the other q indices. Such n is completion of a partial permutation σ of order q. The q terms corresponding to the $a_{i,n(i)}$ are called a_{ij} corresponding to σ. For example, if $A \in M_3$ and q=2, then $x^{(n-q)} = x^1$, and σ permutes 2 different elements of {1,2,3} so the coefficients of x^1 are various products of to terms. This analysis gives a precise algorithm for finding the coefficients of a specific variable in the polynomial $p_A(x)$ for $A \in M_n$.

1. Note the variable's degree *n-q*
2. Find all partial permutations σ such that $|\sigma|$=q
3. Create the ordered pairs (i,j)= $(i, \sigma(i)),\ i \in dom\ \sigma$
4. Multiply all $a_{i,\sigma(i)}$ corresponding to a particular σ and attach the approriate sign.
5. The sum of these signed products is the coefficient of x^{n-q}

Referencing Table 3.1 , if n=3, q=2 , and σ is a partial permutation of order 2, then $|\sigma|$=2 and σ creates two ordered pairs or two a_{ij}*'s* . Thus , each term in the column represents the signed product of the two a_{ij}*'s* whichare the domain of each of the 6 σ 's .With the coefficients ofeach variable of p_A(x) properly defined , $p_A(x)$ is the sum of these terms. These coefficients are either positive or negative and the following definitions are created.

Definition 3.1.3. $p_A(x) = p_A^+(x) - p_A^-(x)$, Where

Definition 3.1.4. $p_A^+(x) = \sum_{q=0}^{n}\left(\sum_{\substack{|\sigma| \\ sgn\sigma=1}} \prod_{i\in dom\sigma} a_{i,\sigma(i)}\right) x^{n-q}$

Definition 3.1.5. $p_A^-(x) = \sum_{q=0}^{n}\left(\sum_{\substack{|\sigma| \\ sgn\sigma=-1}} \prod_{i\in dom\sigma} a_{i,\sigma(i)}\right) x^{n-q}$

Note that if $A \in M_3$, the variables of degree 1 qnd 0 have a combination of $p_A^+(x) and\ p_A^-(x)$ terms.

With the desciption of p_A(x), the cayley –Hamilton Theorem may be rewritten as follows:

Main Theorem =$p_A^+(x) = p_A^-(x)$

With the help of some basic graph theory definitions this theorem can be proved.

VI. CHARACTERSTICS

This theorem is one of the most powerful. It is also known classical matrix theory theorem. Various applications derive their results from C-H theorem. To understand the scope of this theorem , alternate proofs were used. Each proof helped to understand how intertwined areas of mathematics are with respect to matrices and the characteristics polynomial. Straubing's combinatorial proof of the C – H exploits three aspects of p_A(A). First, it elegantly explain the relationship between positvi and negative terms of p_A(A). Second , the proof illuminates the importance of n. It is cornerstone for the entire proof. P_A(A) is a sum of n products of elements of A. Third the proof introduces a cyclic property to p_A(A) are partial permuatins and therefore may be represented by adjoint cycles.

VII. APPLICATION OF CAYLEY– HAMILTON THEOREM

A very common application of the Cayley- Hamilton Theorem is to use it to find A^n usually for the large powers of n. However many of the techniques involved require the use of the eigen values of A.

REFERENCES

[1] William A. Adkins and Mark G. Davidson, *The Cayley Hamilton and Frobenious theorems via the Laplace Transformation*, Linear Algebra and its Applications 371(2003), 147-152.
[2] Arthur Cayley, *A memoir on the theory of Matrices*, available from http:// www.jstor.org, 1857.
[3] Jeffrey A. Rosoff, *A topological Proof of the Cayley- Hamilton Theorem*, Missouri J. Math .Sc. 7(1995), 63-67.
[4] Wikipedia , Arthur Cayley, Available from http:// en.wikipedia .org, 2004
[5] Wikipedia, William Rowan Hamilton, Available from http:// en. wikipedia .org, 2005
[6] D.R. Wilkins, *Linear Operators and the Cayley –Hamilton Theorem* , available from http://www.maths.tod.i.e/pub/histMath/people/Hamilton, 2005