2006

Understanding the Cayley-Hamilton Theorem

Tue Saxon Rust
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Rust, Tue Saxon, "Understanding the Cayley-Hamilton Theorem" (2006). *Master's Theses*. 3027.
DOI: <https://doi.org/10.31979/etd.a6v7-4cuy>
https://scholarworks.sjsu.edu/etd_theses/3027

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

UNDERSTANDING THE CAYLEY-HAMILTON THEOREM

A Thesis

Presented to

The Faculty of the Department of Mathematics

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Tue Saxon Rust

December 2006

UMI Number: 1441121

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 1441121

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2006

Tue Saxon Rust

ALL RIGHTS RESERVED

APPROVED FOR THE DEPARTMENT OF MATHEMATICS

Dr. Wasin So

Dr. Jane Day

Dr. Steven Crunk

APPROVED FOR THE UNIVERSITY

ABSTRACT

UNDERSTANDING THE CAYLEY-HAMILTON THEOREM

by Tue Saxon Rust

This thesis attempts to understand the Cayley-Hamilton theorem through its historical background, various derivations, and applications. It begins with a brief discussion about the origin of matrices, then provides biographies of Arthur Cayley and Sir William Rowan Hamilton, and concludes with the generally recognized proof of the theorem. The main focus of the thesis revolves around bringing new insight to the theorem through alternate derivations. It provides and analyzes proofs via Schur's Triangularization, combinatorics, a variation of a topological proof, and the Laplace Transformation. Properties of upper-triangular matrices are used in the proof via Schur's Triangularization. The proof through combinatorics explores constraints from the characteristic polynomial of a matrix with respect to graph theory. The topological proof uses continuity properties and matrix norms. The proof through the Laplace Transformation focuses on the matrix exponential. There are copious applications of the Cayley-Hamilton theorem; the thesis concludes with a recursive example for finding powers of a matrix without using eigenvalues. Appended is the relationship between the adjoint, the determinant, and the characteristic polynomial, and the permutation definition of a matrix.

DEDICATION

To my mom who monetarily and emotionally afforded me a second chance.

CONTENTS

CHAPTER

1	HISTORICAL BACKGROUND	1
1.1	Origin of Matrices	1
1.2	Sir William Rowan Hamilton	1
1.3	Arthur Cayley	3
1.4	Relationship Between Cayley and Hamilton	4
1.5	Origins of the Cayley-Hamilton Theorem	6
1.6	Generalization of the Cayley-Hamilton Theorem	7
2	ALTERNATE PROOFS OF THE CAYLEY-HAMILTON THEOREM	9
2.1	Proof Through Schur's Triangularization	9
2.2	Combinatorial Proof	12
2.2.1	Partial permutation σ	12
2.2.2	Positive and negative parts of the characteristic polynomial	13
2.2.3	Definitions from graph theory	15
2.2.4	Graphing $p_A(x)$ and resulting properties	17
2.3	Topological Proof	23
2.4	Z-Transform Proof	26
2.4.1	Proof of Equation 2.6.	26
3	ANALYSIS AND APPLICATIONS	33
3.1	Analysis of Alternate Proofs	33
3.1.1	Characteristics	33
3.1.2	Algorithms	34
3.2	Application of the Cayley-Hamilton Theorem	35

REFERENCES	37
-------------------	-----------

APPENDICES

A THE RELATIONSHIP BETWEEN THE ADJOINT, THE DETERMINANT, AND THE CHARACTERISTIC POLYNOMIAL	38
B PERMUTATION DEFINITION OF THE DETERMINANT OF A MATRIX	39
B.1 Characterization of the Determinant	39
B.2 Permutation Definition of the Determinant	42

FIGURES

Figure

2.1	Directed Graphs G_1 and G_2	18
2.2	η_{ij} applied to G_1 returns G_2	20
2.3	η_{ij} applied to G_3 returns G_4	21

CHAPTER 1

HISTORICAL BACKGROUND

1.1 Origin of Matrices

The origin of matrices stems from an endeavor to solve simultaneous systems of linear equations, which may be traced back to the Babylonians during the fourth century BC [OR96]. However, the first civilization to record a matrix was the Chinese, between 200 BC and 100 BC [OR96]. In the text Nine Chapters on the Mathematical Art written during the Han Dynasty, the author creates three systems of equations, records their coefficients in columns (instead of using the convention of rows, used today) and proceeds to explain how to solve the system using Gaussian elimination [OR96]. Over a millennium passes before an abundant amount of significant contributions to matrix theory appear again. In 1812, Cauchy creates the term “determinant” and subsequently invents eigenvalues, diagonalization, and introduces the idea of, but not the term, similar matrices [OR96]. Thereafter, many mathematicians contribute to the development of modern matrix theory.

Arthur Cayley was a prominent figure in the development of matrix theory. In a publication during 1841, “... he used two vertical lines on either side of the array to denote the determinant” [OR96]. He also created a matrix algebra in his 1858 publication of Memoir on the Theory of Matrices [Cay57]. It is in this publication where he conjectures the Cayley-Hamilton theorem.

1.2 Sir William Rowan Hamilton

Although Cayley was the first to conjecture the Cayley-Hamilton theorem, Hamilton was the first to derive it. Hamilton was also Cayley’s senior by 16 years. Therefore, Hamilton’s biography will be discussed first.

Sir William Rowan Hamilton was born August 4, 1805, in Dublin Ireland. He was

recognized as a child prodigy and "... before he was thirteen he had acquired, under the care of his uncle, who was a linguist, almost as many languages as he had years of age" [Wik05]. By age 16 he had mastered Euclid, Newton's *Arithmetica Universalis*, and a large part of *Principia* [Wik05].

His entire career, from college through professorship, centers at Trinity College, Dublin [Wik05]. He was awarded Professor of Astronomy in 1827 before he even graduated, but was given leave to study anything related to science [Wik05]. Regarding Hamilton's approach toward analysis,

It was one of the peculiar characteristics of Hamilton's mind never to be satisfied with a general understanding of a question; Hamilton pursued the problem until he knew it in all its details. [Wik05]

In 1837 he was awarded the president's chair in the Royal Irish Academy [Wik05]. He made many important contributions to optics and dynamics, of note being the principal of "Varying Action" or conical refraction. Although he is known for many mathematical contributions, his personal interest lies with quaternions (Latin for *by fours*). His interest arose from the search for a non-commutative extension of the complex numbers. While "walking along the Royal Canal in Dublin with his wife" [Wik05], he realized the relation

$$i^2 = j^2 = k^2 = ijk = -1$$

and "promptly carved this equation into the side of the nearby Brougham Bridge (now called Broom Bridge)" [Wik05].

Once the concept of a quaternion entered Hamilton's mind, true to his nature, he began painstakingly verifying every consequence of its creation. It became a life-long endeavor to revolutionize the tools used in applied mathematics. Sadly, his quaternions ultimately fell secondary to vector algebra and vector calculus, and have been replaced, in many situations, by the four-vector. However, Hamilton's creation still finds its place as "... the quaternions are in use by computer graphics, control theory, signal processing and orbital mechanics, mainly for representing rotations/orientations." [Wik05]

Hamilton's married life was fraught with difficulties and he eventually became addicted to alcohol. Sir William Rowan Hamilton died September 2, 1865, just after finishing his six-year work titled "Elements of Quaternions". He was a perfectionist and rarely published his works. However, after his death, volumes of unpublished mathematical papers and lectures were compiled and edited for the Royal Irish Academy in "The Mathematical Papers of Sir William Rowan Hamilton". As a special note, "2005 was the 200th anniversary of Hamilton's birth and the Irish government has designated this the Hamilton Year, celebrating Irish science" [Wik05].

1.3 Arthur Cayley

Arthur Cayley, son of Sir Henry Cayley and Maria Antonia Doughty, was born at Richmond in Surrey, England, on August 16, 1821 [Wik04]. Cayley spent the first 8 years of his life in St. Petersburg, Russia, and in 1829, immigrated to Blackheath, near London. At seventeen he entered Trinity College, Cambridge, where he began studying literature [Wik04]. He soon developed an interest in math which turned into "an insatiable hunger for the subject" [Mey00]. He published his first paper at age 20 [Mey00].

He continued to reside at Cambridge for four years; during which time he took some pupils, but his main work was the preparation of 28 memoirs to the *Mathematical Journal*. [Mey00]

However, his fellowship was limited, and he needed a means for earning money, so at age 25 he became a lawyer at Lincoln's Inn, London [Wik04]. His incredible memory [Mey00] helped create a very comfortable living. There, he listened to lectures by Hamilton [Wik04] and befriended James Joseph Sylvester. The two immediately enjoyed each other's company and were soon known as the "invariant twins" [Mey00]. "During this period of his life, extending over fourteen years, Cayley produced between two and three hundred papers" [Wik04].

Cayley's mind was ideal for a mathematician. He was undoubtedly brilliant, "and he read everything that he could lay his hands on" [Wik04]. Soon, his passion and his practice coalesced, as he became professor of Cambridge University. At age 42, Cayley left a lucra-

tive career as a lawyer to accept an elected position to the newly created Sadlerian Chair, the second chair of pure mathematics. The first chair is the ancient Lucasian chair which Isaac Newton occupied [Wik04]. He married and had a happy marriage. In 1883 he became President of the British Association for the Advancement of Science and gave a memorable President's address.

...the address is a valuable historical review of various mathematical theories, and is characterized by freshness, independence of view, suggestiveness, and learning.
[Wik04]

Arthur Cayley died in 1895, at age 74, while editing the volumes of all his mathematical papers. He published thirteen quarto volumes, containing 967 papers, a monumental achievement for mathematics.

He remained active in mathematics until his death and those who knew him say he never lost the gleam in his eye or his "peculiar boyish smile" (Arthur, 1996, 3)
[Ber98]

1.4 Relationship Between Cayley and Hamilton

When an important theorem is named after two people, one assumes that they collaborated on the effort, or at least directly influenced each other. This appears not the case with Hamilton and Cayley. In fact, save for Cayley's early interest in Hamilton's lectures, the amount of written material describing their relationship is sparse at best. They were definitely not adversaries, but they did not work together.

Perhaps the reason for this phenomenon may be due to age. Hamilton is Cayley's superior by 16 years. Cayley may have thought of Hamilton as a role-model for, "While still training to be a lawyer Cayley went to Dublin to hear Hamilton lecture on quaternions" [Wik04]. These excursions may have spirited on Cayley as a mathematician, for during his subsequent 14 years as a lawyer, he published 250 mathematical papers. These facts suggest that Hamilton may have been an indirect mentor to Cayley, but not his colleague.

Although Cayley and Hamilton may not have had a formal friendship, Cayley did have an interest in quaternions. However, his interest may not have been as pronounced as Hamilton's

devotion. In Tait's *Elementary Treatise on Quaternions*, Cayley describes quaternions,

I have the highest admiration for the notion of a quaternion; but, as I consider the full moon far more beautiful than any moonlit view, so I regard the notion of a quaternion as far more beautiful than any of its applications. As another illustration, I compare a quaternion formula to a pocket-map—a capital thing to put in one's pocket, but which for use must be unfolded: the formula, to be understood, must be translated into coordinates. [Wik04]

There is a certain temporal irony associated with the Cayley-Hamilton theorem. Hamilton, unknowing of its importance, proved it first. However, he proved it with quaternions, an arguably more difficult proof than Cayley's two and three dimensional proofs. Quaternions were also a more modern mathematical concept than the two and three dimensional matrices which Cayley used. Hamilton proved the theorem with modern techniques, yet earlier than Cayley, who was Hamilton's junior.

Further contrasts exist between the two mathematicians. It appears that Cayley and Hamilton, although both mathematicians, were pursuing their interest through completely different means. This is illustrated in mathematical upbringing, their comparative lifestyle and research technique. Hamilton was a mathematical prodigy and studied math, at the same university, his entire life. Cayley originally studied literature, and studied law for 14 years before devoting himself to math. Cayley had a wonderful marriage and enjoyed a happy life. Hamilton had a troublesome marriage and died, presumably [Mey00], from alcoholism. Cayley published an enormous amount of papers, while Hamilton, being a perfectionist, published very few. Only after his death were volumes of notes found and published.

Hamilton and Cayley are very different individuals. However, they are connected through a fundamental theorem of linear algebra. The Cayley-Hamilton theorem continues to unite differences through its various derivations. Classically, it is proved through linear algebra. However, it is also proved through combinatorics, the Laplace transformation, and topology. In understanding the proofs, some insight is offered toward the similarities between the different fields of mathematics, brought forth from the Cayley-Hamilton theorem.

1.5 Origins of the Cayley-Hamilton Theorem

Definition 1.5.1. If A is an $n \times n$ 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_0I in $q(A)$.

Theorem 1.5.2. (*Cayley-Hamilton Theorem*). If A is an $n \times n$ matrix, then $p_A(A) = 0$, the zero matrix.

D.R. Wilkins, School of Mathematics, Trinity College, Dublin, eloquently summarizes Hamilton's proof of the Cayley-Hamilton theorem.

Hamilton wrote three short papers on linear operators on quaternions, showing how the inverse of such an operator can be expressed as a polynomial in the operator, and proving a result for linear operators on the space of quaternions which is a special case of the general theorem which today is known as the 'Cayley-Hamilton Theorem', which states that every square matrix and every linear operator on a finite-dimensional vector space satisfies its characteristic equation. [Wil05]

D.R. Wilkins offers a web page with a link to Hamilton's three papers delineating his proof [Wil05]. Below is a concise version of the Cayley-Hamilton proof for a quaternion.

Theorem 1.5.3. If $q \neq 0$ is a quaternion of the form $q = a + bi + cj + dk$ with a, b, c, d being real, then $q^2 - 2aq + (a^2 + b^2 + c^2 + d^2) = 0$.

Proof.

$$\begin{aligned}
 q^{-1} &= \frac{\bar{q}}{|q|} \\
 &= \frac{a - bi - cj - dk}{a^2 + b^2 + c^2 + d^2} \\
 &= \frac{2a}{a^2 + b^2 + c^2 + d^2} - \frac{a + bi + cj + dk}{a^2 + b^2 + c^2 + d^2} \\
 &= \frac{1}{a^2 + b^2 + c^2 + d^2} (2a - q) \\
 \Rightarrow (a^2 + b^2 + c^2 + d^2) &= 2aq - q^2 \\
 \Rightarrow q^2 - 2aq + (a^2 + b^2 + c^2 + d^2) &= 0
 \end{aligned}$$

□

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

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

then $p_A(A) = A^2 - 2aA + (a^2 + b^2 + c^2 + d^2)I = 0$, and the polynomial given in Theorem 1.5.3 is the characteristic polynomial of A .

Although Hamilton first proves the result for a special kind of 2×2 matrix in 1853, Cayley enunciates the generalization in 1857 [Mey00].

...the determinant, having for its matrix a given matrix less the same matrix considered as a single quantity involving the matrix unity, is equal to zero. [Cay57]

His proof, however is incomplete. He proves it for 2×2 matrices, M_2 , states that he “has verified it” [Cay57] for M_3 matrices yet then states, “but I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case of a matrix of any degree”. For 20 years, the proof eludes mathematicians until Ferdinand Georg Frobenius proves it in 1878.

1.6 Generalization of the Cayley-Hamilton Theorem

Ferdinand Georg Frobenius (October 26, 1849 - August 3, 1917) generalized Cayley’s and Hamilton’s proof to include all n -dimensional matrices in his 1878 treatise [MPG98]. The generally recognized proof follows [Jac90].

Theorem 1.6.1. (*Cayley-Hamilton Theorem*). For any $n \times n$ matrix A , $p_A(A) = 0$.

Proof. Let $D(x)$ be the matrix with polynomial entries $D(x) = \text{adj}(xI_n - A)$, so $D(x)(xI - A) = \det(xI_n - A)I_n$ (see Appendix A). Since each entry in $D(x)$ is the determinant of an $(n - 1) \times (n - 1)$ submatrix of $(xI_n - A)$, each entry of $D(x)$ is a polynomial of degree less than or equal to $n - 1$. It follows that there exist matrices $D_0, D_1, \dots, D_{n-1}$ with entries from $\mathbb{C}$ such that $D(x) = D_{n-1}x^{n-1} + \dots + D_1x + D_0$. ($D_i x^i$ is an $n \times n$ matrix all of whose entries have the form dx^i for some $d \in \mathbb{C}$.) Then this matrix equation follows.

$$\det(xI_n - A)I_n = (xI_n - A)\text{adj}(xI_n - A) = (xI_n - A)D(x)$$

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

$$\begin{aligned}
x^n I_n &+ b_{n-1} x^{n-1} I_n + \cdots + b_1 x I_n + b_0 I_n \\
&= p_A(x) I_n = \det(xI_n - A) I_n \\
&= (xI_n - A) \operatorname{adj}(xI_n - A) \\
&= (xI_n - A)(x^{n-1} D_{n-1} + \cdots + x D_1 + D_0) \\
&= x^n D_{n-1} - x^{n-1} A D_{n-1} + x^{n-1} D_{n-2} \\
&\quad - x^{n-2} A D_{n-2} + \cdots + x D_0 - A D_0 \\
&= x^n D_{n-1} + x^{n-1} (-A D_{n-1} + D_{n-2}) \\
&\quad + \cdots + x(-A D_1 + D_0) - A D_0
\end{aligned} \tag{*}$$

Since two polynomials are equal if and only if their coefficients are equal, the coefficient matrices are equal; that is, $I_n = D_{n-1}$, $b_{n-1} I_n = (-A D_{n-1} + D_{n-2})$, $\dots$, $b_1 I_n = -(A D_1 + D_0)$, and $b_0 I_n = -A D_0$. This means that A may be substituted for the variable x in the equations (*) to conclude:

$$\begin{aligned}
p_A(A) &= A^n + b_{n-1} A^{n-1} + \cdots + b_1 A + b_0 I_n \\
&= A^n D_{n-1} + A^{n-1} (-A D_{n-1} + D_{n-2}) + \cdots \\
&\quad + A(-A D_1 + D_0) - A D_0 \\
&= A^n D_{n-1} - A^n D_{n-1} + A^{n-1} D_{n-2} - A^{n-1} D_{n-2} + \cdots \\
&\quad + A D_0 - A D_0 = 0
\end{aligned}$$

This proves the theorem. □

CHAPTER 2

ALTERNATE PROOFS OF THE CAYLEY-HAMILTON THEOREM

2.1 Proof Through Schur's Triangularization

This proof synthesizes work of Issai Schur, who was Frobenius's student, and Mirko Stojakovic [Sto71].

Lemma 2.1.1. *If $S_1, S_2, \dots, S_n$ are $n \times n$ upper-triangular matrices such that the (i, i) element of S_i is zero for all i , then*

$$S_1 S_2 \cdots S_n = 0.$$

Proof. (Induction on 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 $S_1 S_2 = 0$.

Assume now that the theorem is true for some integer m . Then

$$S_1 S_2 \cdots S_m S_{m+1} = T S_{m+1} \text{ where } S_1 = \begin{bmatrix} T_1 & a_1 \\ \bar{0} & x_1 \end{bmatrix}, \dots, S_m = \begin{bmatrix} T_m & a_m \\ \bar{0} & x_m \end{bmatrix}$$

and $T_1, \dots, T_m \in M_m$ are upper triangular matrices such that the (i, i) element of T_i is 0.

Then

$$T = S_1 S_2 \cdots S_m = \begin{bmatrix} T_1 T_2 \cdots T_m & u_m \\ \bar{0} & x \end{bmatrix} = \begin{bmatrix} 0_m & u_m \\ \bar{0} & x \end{bmatrix} \text{ and } S_{m+1} = \begin{bmatrix} A_m & t_m \\ \bar{0} & 0 \end{bmatrix}$$

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

Theorem 2.1.2. (*Schur's Theorem*). Given $A \in M_n$ with eigenvalues $\lambda_1, \dots, \lambda_n$ in any prescribed order, there is a unitary matrix $U \in M_n$ such that

$$U^*AU = T = [t_{ij}]$$

is upper triangular, with diagonal entries $t_{ii} = \lambda_i$, $i = 1, \dots, n$.

Proof. Let $x_{(1)}$ be a unit eigenvector of A associated with the eigenvalue λ_1 . The nonzero vector $x_{(1)}$ may be extended to a basis

$$x_{(1)}, y_{(2)}, y_{(3)}, \dots, y_{(n)}$$

of C^n . Apply the Gram-Schmidt orthonormalization procedure to this basis to produce an orthonormal basis

$$x_{(1)}, z_{(2)}, z_{(3)}, \dots, z_{(n)}$$

of C^n . Array these orthonormal vectors left to right as the columns of a unitary matrix U_1 .

Since the first column of AU_1 is $\lambda_1 x_{(1)}$, a calculation reveals that

$$U_1^*AU_1 = \left[\begin{array}{c|c} \lambda_1 & * \\ \hline 0 & A_1 \end{array} \right].$$

The matrix $A_1 \in M_{n-1}$ has eigenvalues $\lambda_2, \dots, \lambda_n$. Let $x_{(2)} \in C^{n-1}$ be a unit eigenvector of A_1 corresponding to λ_2 , and do it all over again. Determine a unitary $U_2 \in M_{n-1}$ such that

$$U_2^*A_1U_2 = \left[\begin{array}{c|c} \lambda_2 & * \\ \hline 0 & A_2 \end{array} \right]$$

and let

$$V_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & U_2 \end{array} \right].$$

The matrices V_2 and U_1V_2 are then unitary, and $V_2^*U_1^*AU_1V_2$ has the form

$$V_2^*U_1^*AU_1V_2 = \left[\begin{array}{cc|c} \lambda_1 & * & * \\ 0 & \lambda_2 & \\ \hline 0 & & A_2 \end{array} \right].$$

Continue this reduction to produce unitary matrices $U_i \in M_{n-i+1}$, $i = 1, \dots, n-1$ and unitary matrices $V_i \in M_n$, $i = 2, \dots, n-1$. The matrix

$$U = U_1 V_2 V_3 \cdots V_{n-1}$$

is unitary and $U^* A U$ yields the desired form. $\square$

With Schur's unitary triangularization theorem (Theorem 2.1.2) and Lemma 2.1.1, the Cayley-Hamilton Theorem may be easily proved [HJ85].

Main Theorem. (*Cayley-Hamilton Theorem*). *Let $p_A(t)$ be the characteristic polynomial of $A \in M_n$. 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 eigenvalues $\lambda_1, \dots, \lambda_n$ of A , counting multiplicities, factor $p_A(t)$ as

$$p_A(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$$

Using Schur's Theorem, write A as

$$A = U T U^*$$

where T is upper triangular with λ_i in the i th diagonal position, $i = 1, \dots, n$. The theorem follows.

$$\begin{aligned} p_A(A) &= p_A(UTU^*) = (UTU^* - \lambda_1 I)(UTU^* - \lambda_2 I) \cdots (UTU^* - \lambda_n I) \\ &= [U(T - \lambda_1 I)U^*][U(T - \lambda_2 I)U^*] \cdots [U(T - \lambda_n I)U^*] \\ &= U[(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I)]U^* \\ &= 0 \end{aligned}$$

The last equality follows from Lemma 2.1.1. $\square$

2.2 Combinatorial Proof

Straubing proves the Cayley-Hamilton Theorem using a combinatorial derivation of the determinant of a matrix in conjunction with elementary graph theory. First, he introduces notation for a particular partial permutation. Then he divides the matrix's characteristic polynomial into two groups composed of positive and negative terms. He concludes by asserting that these two groups are equal and opposite.

2.2.1 Partial permutation σ

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

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

$$\text{Definition 2.2.1. } \hat{\sigma}(i) = \begin{cases} \sigma(i) & \text{if } i \in \text{dom } \sigma, \\ i & \text{if } i \in \{1, \dots, n\} \setminus \text{dom } \sigma. \end{cases}$$

Definition 2.2.2. The *signature* of a complete permutation $\hat{\sigma}$, denoted $\text{sgn}(\hat{\sigma})$, is $+1$ if the total number of inversions in $\hat{\sigma}$ is even and -1 if that number is odd.

Definition 2.2.3. The *signature* of a partial permutation σ , denoted $\text{sgn}(\sigma)$, is defined by $\text{sgn}(\sigma) = (-1)^{|\sigma|} \text{sgn}(\hat{\sigma})$.

The characteristic polynomial, $p_A(x)$, of a matrix is the sum of certain products of elements of that matrix and powers of x (See Appendix B, B.0.7). It is shown below that the pairs of indices (i, j) appearing in one of these products can be described using partial permutations. There is a relation between the elements of a given product. Namely, their subscripts are ordered pairs $(i, \sigma(i))$ where σ is a partial permutation of $\{1, \dots, n\}$ and $i \in \text{dom } \sigma$. For example, if $A \in M_3$, the terms of $p_A(x)$ are:

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

Table 2.1: Terms of $p_A(x)$ of M_3

Note that here the coefficient 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 $\sigma_1 = (1)(2)$,

$a_{12}a_{21}$ corresponds to $\sigma_4 = (1, 2)$, etc.

It will be shown in general that each partial permutation of $\{1, \dots, n\}$ of order q yields a signed term, which is one of the summands in the coefficient of x^{n-q} . Graph theory will help one visualize the partial permutations involved. Let $\text{dom } \sigma$ be vertices and the ordered pairs $(i, \sigma(i))$, where $i \in \text{dom } \sigma$, 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 of lemmas to prove the Cayley-Hamilton Theorem. To use this transition $p_A(x)$ must be described in some detail.

2.2.2 Positive and negative parts of the characteristic polynomial

Let $A = (a_{ij}) \in M_n$ over $\mathbb{C}$. As defined in Appendix B, B.0.7, the characteristic polynomial of A is:

$$p_A(x) = \det(xI - A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n b_{i, \sigma(i)} \quad (2.1)$$

where

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

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

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

To see this, notice that from Equation 2.1, the coefficient of x^{n-q} comes from the terms $b_{i, \eta(i)} = b_{ii} = x - a_{ii}$ for $n - q$ of the indices, and $b_{i, \eta(i)} = -a_{i, \eta(i)}$ for the other q indices. Such η is the completion of a partial permutation σ of order q . The q terms corresponding to the $a_{i, \eta(i)}$ are called the 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 two 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 $(i, j) = (i, \sigma(i))$, $i \in \text{dom } \sigma$
- (4) Multiply all $a_{i, \sigma(i)}$ corresponding to a particular σ and attach the appropriate sign
- (5) The sum of these signed products is the coefficient of x^{n-q}

Referencing Table 2.1, if $n=3$, $q=2$, and σ is a partial permutation of order 2, then $|\sigma| = 2$ and each σ creates two ordered pairs, or two a_{ij} 's. Thus, each term in the third column represents the signed product of the two a_{ij} 's which are the domain of each of the 6 σ 's.

With the coefficients of each 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 2.2.4. $p_A(x) = p_A^+(x) - p_A^-(x)$,

where

Definition 2.2.5. $p_A^+(x) = \sum_{q=0}^n \left(\sum_{\substack{|\sigma|=q \\ \text{sgn } \sigma=1}} \prod_{i \in \text{dom } \sigma} a_{i,\sigma(i)} \right) x^{n-q}$

and

Definition 2.2.6. $p_A^-(x) = \sum_{q=0}^n \left(\sum_{\substack{|\sigma|=q \\ \text{sgn } \sigma=-1}} \prod_{i \in \text{dom } \sigma} a_{i,\sigma(i)} \right) x^{n-q}.$

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

With this description of $p_A(x)$, the Cayley-Hamilton Theorem may be rewritten as follows.

Main Theorem. $p_A^+(A) = p_A^-(A)$

The remainder of this section is devoted to proving this theorem. Some basic graph theory definitions are required.

2.2.3 Definitions from graph theory

A *graph* is a set of vertices connected by edges. A directed graph is a *graph* whose edges are ordered pairs of vertices. That is, each edge can be followed from one vertex to another vertex. A *path* π in $\{1, \dots, n\}$ is a sequence of ordered pairs $((i_0, i_1), \dots, (i_{q-1}, i_q))$ where each i_k belongs to $\{1, \dots, n\}$. The number q is called the *length* of π and is denoted $|\pi|$. The i_j are called the *vertices* of π and the pairs (i_j, i_{j+1}) are the *edges* of π . The vertices i_0 and i_q are called the *start* of π and the *finish* of π and are denoted $\alpha(\pi)$ and $\omega(\pi)$ respectively. A convention is made that for each $i \in \{1, \dots, n\}$ there is a unique path π_i such that $|\pi_i| = 0$, $\alpha(\pi_i) = \omega(\pi_i) = i$ and $\mu(\pi_i) = 1$. Observe that if $u < v$ and $i_u = i_v$ in π , one can remove $(i_u, i_{u+1}), \dots, (i_{v-1}, i_v)$ from π and obtain a shorter path π' . For example, if $\pi = ((1, 2), (2, 3), (3, 2), (2, 4))$, one can remove $(2, 3), (3, 2)$ to obtain the new path $\pi' = ((1, 2), (2, 4))$.

Use a_{ij} and A_{ij} to denote the (i, j) entry of a matrix $A \in M_n$. For a fixed matrix A , the *value* of π is defined by

Definition 2.2.7. $\mu(\pi) = \prod_{k=0}^{q-1} a_{i_k i_{k+1}}.$

If σ is a partial permutation, then the *value* of σ is defined by

Definition 2.2.8. $\mu(\sigma) = \prod_{i \in \text{dom } \sigma} a_{i, \sigma(i)}.$

By the convention made in the previous section, the value of the partial permutation of degree zero is 1.

Let $1 \leq i, j \leq n$ and let T_{ij}^+ be the set of all pairs (σ, π) where σ is a partial permutation and π is a path which satisfy

$$|\sigma| + |\pi| = n, \quad \alpha(\pi) = i, \quad \omega(\pi) = j, \quad \text{sgn } \sigma = 1. \quad (2.2)$$

The set T_{ij}^- is defined identically, except that condition ‘ $\text{sgn } \sigma = 1$ ’ is replaced by ‘ $\text{sgn } \sigma = -1$ ’. T_{ij} denotes the union $T_{ij}^+ \cup T_{ij}^-$.

Theorem 2.2.9. For $k \geq 1$,

$$(A^k)_{ij} = \sum_{\substack{|\pi|=k, \alpha(\pi)=i, \\ \omega(\pi)=j}} \mu(\pi).$$

Proof. (Induction on n)

$(A^1)_{ij} = a_{ij}$ so $\pi = (i, j)$ is the only path from i to j and the formula is valid.

Assume $k \geq 1$ and

$$(A^k)_{ij} = \sum_{\substack{|\pi|=k, \alpha(\pi)=i, \\ \omega(\pi)=j}} \mu(\pi).$$

Then,

$(A^{k+1})_{ij} = ((A^k) A)_{ij} = \sum_{p=0}^{n-1} (A^k)_{ip} A_{pj}$ and each term has $k+1$ products of elements of A , so $|\pi| = k+1$. Then, by the Inductive Hypothesis,

$$(A^{k+1})_{ij} = \sum_{p=0}^{n-1} \sum_{\substack{|\pi|=k, \alpha(\pi)=i \\ \omega(\pi)=p}} \mu(\pi) \cdot A_{pj} = \sum_{\substack{p=0 \\ |\pi|=k+1, \alpha(\pi)=i \\ \omega(\pi)=j}}^{n-1} \mu(\pi).$$

□

The convention regarding paths of length zero makes this formula valid for $k = 0$ as well. It follows that

$$\sum_{q=0}^{n-1} \left(\sum_{\substack{|\sigma|=q \\ \text{sgn } \sigma=1}} \mu(\sigma) \right) (A^{n-q})_{ij} = \sum_{q=0}^{n-1} \left(\sum_{\substack{|\sigma|=q \\ \text{sgn } \sigma=1}} \mu(\sigma) \right) \sum_{\substack{p=0 \\ |\pi|=n-q, \alpha(\pi)=i \\ \omega(\pi)=j}}^{n-1} \mu(\pi), \text{ and}$$

$$p_A^+(A)_{ij} = \sum_{(\sigma, \pi) \in T_{ij}^+} \mu(\sigma) \mu(\pi), \quad p_A^-(A)_{ij} = \sum_{(\sigma, \pi) \in T_{ij}^-} \mu(\sigma) \mu(\pi). \quad (2.3)$$

The Main Theorem will follow after the results below, which show that these two sums are equal for all $i, j \in \{1, \dots, n\}$.

2.2.4 Graphing $p_A(x)$ and resulting properties

Definition 2.2.10. Given a particular (σ, π) , create a pair of directed graphs G_1 and G_2 on the vertices $\{1, \dots, n\}$ with two classes of edges. Let σ be a partial permutation of degree q and let π be a path of length $n - q$ so that σ and π satisfy Equation 2.2. Define the edges of G_1 and G_2 as

$$U = \{(i, \sigma(i)) \mid i \in \text{dom } \sigma\} \text{ and}$$

$$V = \{(k, l) \mid (k, l) \text{ is an edge of } \pi\}.$$

Graph G_2 will, in general, contain multiple edges, since π may traverse the same edge more than once (i.e., V is a *multiset*). Also, the same edge may appear in both U and V . The expression $\mu(\sigma) \mu(\pi)$ is the product, with multiplicities taken into account, of all the a_{kl} where (k, l) is an edge of G_1 or G_2 .

Two examples are illustrated (see Figure 2.1). Denote the edges in U by dashed lines, and those in V by solid lines. The order of traversal of the edges of the path π is indicated by the small Roman numerals.

Figure 2.1: Directed Graphs G_1 and G_2

Lemma 2.2.11. *Let $\pi = ((i_0, i_1), \dots, (i_{q-1}, i_q))$ be a path of length q and σ be a partial permutation such that $|\sigma| = n - q$. There is a smallest integer $v \geq 0$ such that exactly one of the following must be true: $i_u = i_v$ for some u less than v or $i_v \in \text{dom } \sigma$.*

Proof. Let $S = \{j \in \{0, \dots, n\} : i_j \in \text{dom } \sigma \text{ or there exists } u < j \text{ with } i_u = i_j\}$. To see $S \neq \emptyset$, suppose $i_u = i_j$ is not true for all $u < j$. Then π does not cross itself and it must span $q + 1$ vertices. Also, σ must span $n - q$ vertices for it is bijective. However, there are only n vertices in the graph (Equation 2.2), and there must be some $i_j \in \text{dom } \sigma$, hence $j \in S$.

Let $v = \min(S)$. Suppose there exists $u < v$ with $i_u = i_v$. Then $i_v \notin \text{dom } \sigma$. For if $i_v \in \text{dom } \sigma$, then $i_u \in \text{dom } \sigma$ also, but then $u \in S$, which contradicts $v = \min(S)$. $\square$

For the examples above, $v = 0$ in G_1 , and $v = 2$ in G_2 .

Definition 2.2.12. When $\begin{pmatrix} i_1 & \dots & i_k \\ \sigma(i_1) & \dots & \sigma(i_k) \end{pmatrix}$ is a partial permutation of $\{i_1, \dots, i_k\}$ such that $\sigma(i_k) = i_1$, this is called a *cycle*. For example, $\begin{pmatrix} 2 & 4 & 5 \\ 4 & 5 & 2 \end{pmatrix}$ is a cycle for $\{2, 4, 5\}$.

Theorem 2.2.13. For each $(i, j) \in \{1, \dots, n\} \times \{1, \dots, n\}$ there is a bijective function $\eta_{ij} : T_{ij}^+ \rightarrow T_{ij}^-$ such that $\eta_{ij}(\sigma, \pi) = (\sigma', \pi')$ implies $\mu(\sigma)\mu(\pi) = \mu(\sigma')\mu(\pi')$.

Proof. Let $(\sigma, \pi) \in T_{ij}^+$ and define v as in Lemma 2.2.11. If $i_v \in \text{dom } \sigma$, then define $\eta_{ij}(\sigma, \pi) = (\sigma', \pi')$ where σ' is formed by removing the cycle containing i_v from σ , and where π' is formed by inserting this cycle into π . Note that if $i_v \in \text{dom } \sigma$, then it is always part of a cycle because σ is bijective. If $i_v \notin \text{dom } (\sigma)$, then $i_u = i_v$ for some $u < v$. In this case, remove the loop $((i_u, i_{u+1}), \dots, (i_{v-1}, i_v))$ from π and adjoin it to σ . This yields a new partial permutation σ' because the vertices in this loop are distinct and none already belong to $\text{dom } \sigma$ from Lemma 2.2.11. In this instance define $\eta_{ij}(\sigma, \pi) = (\sigma', \pi')$, where σ' is the resulting longer partial permutation and π' is the resulting shorter path. These two cases are disjoint and are the only alternatives, so η_{ij} is well defined. The results of applying η_{ij} to the graphs in Figure 2.1 are illustrated in Figures 2.2 and 2.3. The other assertions in the theorem are proved by the next three Lemmas. $\square$

$$n = 6, \sigma = \begin{pmatrix} 1 & 3 & 5 & 6 \\ 3 & 5 & 1 & 6 \end{pmatrix} \pi = ((1, 3), (3, 2))$$

$$U = \{(1, 3), (3, 5), (5, 1), (6, 6)\}, \quad V = \{(1, 3), (3, 2)\}$$

$$n = 6, \sigma' = \begin{pmatrix} 6 \\ 6 \end{pmatrix} \pi' = ((1, 3), (3, 5), (5, 1), (1, 3), (3, 2))$$

$$U' = \{(6, 6)\}, \quad V' = \{(1, 3), (3, 5), (5, 1), (1, 3), (3, 2)\}$$

Figure 2.2: η_{ij} applied to G_1 returns G_2

Figure 2.3: η_{ij} applied to G_3 returns G_4

Lemma 2.2.14. $\text{sgn } \sigma = -\text{sgn } \sigma'$, hence η_{ij} does map T_{ij}^+ into T_{ij}^- .

Proof. The function η_{ij} adjoins or removes exactly one cycle to or from σ . Adjoining or removing a cycle of length k makes $|\sigma'| = |\sigma| \pm k$. If k is odd, then $\text{sgn}(\hat{\sigma})$ does not change but $(-1)^{|\sigma|}$ does, hence $\text{sgn}(\sigma)$ changes (see Definition 2.2.3). If k is even, then $\text{sgn}(\hat{\sigma})$ changes but $(-1)^{|\sigma|}$ does not change, hence $\text{sgn}(\sigma)$ changes. Either way, when $\eta_{ij}(\sigma, \pi) = (\sigma', \pi')$, $\text{sgn}(\sigma') = -\text{sgn}(\sigma)$, so $\eta_{ij} : T_{ij}^+ \rightarrow T_{ij}^-$. $\square$

Lemma 2.2.15. η_{ij} is bijective.

Proof. η_{ij} either adds or removes a cycle from σ . If it adds a cycle to σ , then it has removed it from π . This happens when this cycle is in π and contains the lowest index v such that i_v is the finish vertex of one of V 's cycles. Because V is connected, i_v is connected to another vertex in V . Also, there is no other u such that $i_u \in \text{dom } \sigma$ where $u < v$. Thus, after η_{ij}

removed the cycle from V and adjoined it to U , it necessarily makes $i_v \in \text{dom } \sigma$, and applying η_{ij} to (σ', π') will take the exact same cycle from σ' and adjoin it to π' , yielding (σ, π) . A similar argument proves the same result if η_{ij} removes a cycle from σ . $\square$

Definition 2.2.16. In graph theory, *graph coloring* is an assignment of colors, red, blue, and so on (but consecutive integers starting from 1 can be used without loss of generality), to certain objects in a graph so that adjacent objects have different colors. Such objects can be vertices, edges, faces, or a mixture of those. In Figures 2.2 and 2.3, the edges corresponding to σ and π can be colored with two different colors.

Theorem 2.2.17. $\mu(\sigma)\mu(\pi) = \mu(\sigma')\mu(\pi')$

Proof. This is obviously true because $\mathbb{C}$ is commutative and η_{ij} only changes the coloring of the graph's edges. $\square$

The previous three lemmas assert that, given $(\sigma, \pi) \in T_{ij}^+$, there exists a $(\sigma', \pi') \in T_{ij}^-$ such that $\mu(\sigma)\mu(\pi) = \mu(\sigma')\mu(\pi')$, which completes the proof of Theorem 2.2.13. Equation 2.3 together with Theorem 2.2.17 imply $p_A^+(A)_{ij} = p_A^-(A)_{ij}$ and $p_A(A)_{ij} = p_A^+(A)_{ij} - p_A^-(A)_{ij} = 0$ and this completes the proof of the Main Theorem.

2.3 Topological Proof

The following proof is inspired by Jeffrey A. Rosoff's topological proof [Ros95]. The proof has been shortened considerably by offering an alternate proof of the density of diagonalizable matrices. The following is the most concise alternate proof to Frobenius's proof I have found or created and uses the matrix norm and its properties.

Theorem 2.3.1. *The set $D_n = \{A \in M_n \mid A \text{ is diagonalizable}\}$ of diagonalizable matrices is dense in M_n .*

Proof. Fix $\epsilon > 0$. It is sufficient to show that, given any matrix $A \in M_n$, there exists a diagonalizable matrix B such that

$$\|A - B\|_2 < \epsilon$$

where $\|\bullet\|_2$ is the *Frobenius norm* given by

$$\|\bullet\|_2 \equiv \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}.$$

Given $A \in M_n$, let $A = UTU^*$ where U is unitary and T is upper-triangular, possible by Schur's Triangularization theorem (Theorem 2.1.2). Define $B = A + UCU^*$, where

$$C_{fg} = \begin{cases} 0 & \text{if } f \neq g \\ \frac{\epsilon\delta}{2f\sqrt{n}} & \text{if } f = g. \end{cases}$$

Choose $\delta > 0$ so that B will have distinct eigenvalues, hence B will be diagonalizable. Let $\sigma(A) = \{\lambda_f, f = 1, 2, \dots, n\}$, and $\lambda_{i+1} \geq \lambda_i$. Define δ as

$$\delta < \min_{\lambda_f \neq \lambda_g} \left\{ \frac{(\lambda_g - \lambda_f)2\sqrt{n}}{\epsilon} \left(\frac{1}{f} - \frac{1}{g} \right) \right\} \text{ and } \delta < 1.$$

If $\lambda_h = \lambda_i$

$$\begin{aligned} \lambda_h + \frac{\epsilon\delta}{2h\sqrt{n}} &= \lambda_i + \frac{\epsilon\delta}{2h\sqrt{n}} \\ &\neq \lambda_i + \frac{\epsilon\delta}{2i\sqrt{n}}. \end{aligned}$$

If $\lambda_h \neq \lambda_i$

$$\begin{aligned}
\frac{(\lambda_i - \lambda_h)2\sqrt{n}}{\epsilon \left(\frac{1}{h} - \frac{1}{i}\right)} &\geq \min_{\lambda_g \geq \lambda_f} \left\{ \frac{(\lambda_g - \lambda_f)2\sqrt{n}}{\epsilon} \left(\frac{1}{f} - \frac{1}{g}\right) \right\} > \delta \\
\Rightarrow \lambda_i - \lambda_h &> \frac{\epsilon\delta}{2\sqrt{n}} \left(\frac{1}{h} - \frac{1}{i}\right) \\
\Rightarrow \lambda_i + \frac{\epsilon\delta}{2i\sqrt{n}} &> \lambda_h + \frac{\epsilon\delta}{2h\sqrt{n}}.
\end{aligned}$$

The diagonal entries of $T + C$ are distinct from the choice of δ , so B has distinct eigenvalues and is diagonalizable. Then

$$\begin{aligned}
|||A - B|||_2 &= |||A - (A + UCU^*)|||_2 \\
&= |||C|||_2 && \text{because } |||\bullet|||_2 \text{ is unitarily invariant} \\
&= \sqrt{\sum_{f=1}^n \left(\frac{\epsilon\delta}{2f\sqrt{n}}\right)^2} && \text{because } C \text{ is diagonal} \\
&\leq \sqrt{\sum_{f=1}^n \frac{1}{n} \left(\frac{\epsilon\delta}{2}\right)^2} \\
&= \frac{\delta\epsilon}{2} \\
&< \epsilon && \text{since every } \delta < 1.
\end{aligned}$$

□

Lemma 2.3.2. *If A is diagonalizable, then $p_A(A) = 0$.*

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

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

and λ_i are eigenvalues of A . Then,

$$\begin{aligned}
 p_A(A) &= (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I) \\
 &= P(D - \lambda_1 I)(D - \lambda_2 I) \cdots (D - \lambda_n I)P^{-1} \\
 &= P \cdot 0 \cdot P^{-1} \\
 &= 0
 \end{aligned}$$

□

Main Theorem. (*Cayley-Hamilton*). If $p_A(x)$ is the characteristic 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 Lemma 2.3.2, the Cayley-Hamilton Theorem is proved for all diagonalizable matrices. The mapping $\Omega : M_n \times P_n \rightarrow M_n$ given by $\Omega(A, f(x)) = f(A)$ is continuous, and the mapping $\Phi : M_n \rightarrow M_n \times P_n$ given by $\Phi(A) = (A, p_A(x))$ is continuous. Hence, the composition

$$\Omega \circ \Phi : M_n \rightarrow M_n,$$

which is given by $\Omega \circ \Phi(A) = p_A(A)$, is continuous. By Theorem 2.3.1 and Lemma 2.3.2, this mapping is identically zero on a dense subset of M_n , so by continuity vanishes everywhere. □

2.4 Z-Transform Proof

This proof is based on the article by Adkins and Davidson [AD03]. Although an alternate proof is not offered, aspects of the article will be clarified to elicit understanding by any person with some knowledge of power series, differentiation, abstract algebra, and the characteristic polynomial of a matrix. The article proves the theorem through understanding the inverse Laplace transform of a matrix exponential.

The proof is quite involved and a brief summary may help in its understanding. The Z-transform allows us to reduce a polynomial over matrices to a polynomial over complex numbers. The definition of the characteristic polynomial of $A \in M_n$, $p_A(s)$, is

$$p_A(s) = \prod_{i=1}^m (s - a_i)^{n_i}, \quad (2.4)$$

where a_i are the eigenvalues of A , n_i are the eigenvalue's multiplicities and $0 < m \leq n$ is the number of distinct eigenvalues of A . Using a power series representation of the matrix exponential e^{At} , and the inverse Laplace transform with respect to matrices, rewrite $p_A(s)$ as

$$p_A(A) = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} p_A^{(j)}(a_i). \quad (2.5)$$

Proving this and defining the elements of M represents the bulk of the paper. Now, Equation 2.4 shows that if $j \leq n_i - 1$, then $p_A^{(j)}(a_i) = 0$, where $p_A^{(j)}(a_i)$ is the j^{th} derivative of p_A , evaluated at a_i . However, Equation 2.5 shows that $j \leq n_i - 1$. Thus, $p_A(A) = 0$ and the Cayley-Hamilton theorem may be formally proved.

Equation 2.6 explains how to write the matrix exponential as a summation of matrices.

$$e^{At} = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} t^j e^{a_i t} = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} \varphi_{j,a_i}(t) \quad (2.6)$$

where $M_{ij} \in M_n(k)$.

2.4.1 Proof of Equation 2.6.

To prove Equation 2.6 the Laplace transform formula with respect to matrices must be understood,

$$L(e^{At}) = (sI - A)^{-1}, \quad (2.7)$$

or more specifically, its inverse,

$$e^{At} = L^{-1} ((sI - A)^{-1}) .$$

Equation 2.7 creates a bijective function not only over matrix exponentials but also over any algebraically closed field, k , of characteristic 0. A field has characteristic 0 if multiple additions of its multiplicative identity never yield the additive identity. For example, $Z_3 = Z(\text{mod } 3)$ is a field but has characteristic 3 because 1, added to itself 3 times, mod 3, equals 0. However, the complex field does have characteristic 0. Thus, the identity matrix, added multiple times, never equals the 0 matrix.

If k is the complex field, let $k(s)$ be the vector space of complex rational functions of indeterminate s , $k^0(s)$ be the vector space of proper rational functions (functions whose degree of the denominator is larger than the degree of the numerator), and $k[[t]]$ be the vector space of formal power series over k in the indeterminate t . The formal power series of an infinitely differentiable function, $f(x)$, about a is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a) (x-a)^n}{n!} .$$

Let $E[[t]]$ be the k -linear subspace of $k[[t]]$ generated by the formal power series $\varphi_{n,a}(t) = t^n e^{at}$ for $n = 0, 1, \dots$, and $a \in k$, where $e^{at} = \sum_{m=0}^{\infty} \frac{1}{m!} a^m t^m$. First observe that the additional family of all $\varphi_{n,a}(t)$ is linearly independent, and thus is a basis for $E[[t]]$. To see the independence, assume $\sum_{i=1}^p c_i t^{n_i} e^{a_i t} = 0$ for all t . Reorder if necessary so that $n_1 \leq n_2 \leq \dots \leq n_m \leq n_{m+1} \leq \dots \leq n_p$. Suppose there is a least subscript m such that $c_m \neq 0$. Then

$$\begin{aligned} c_m t^{n_m} e^{a_m t} + c_{m+1} t^{n_{m+1}} e^{a_{m+1} t} + \dots + c_p t^{n_p} e^{a_p t} &= 0 \quad \forall t \\ \Rightarrow c_m e^{a_m t} + c_2 t^{n_{m+1}-n_m} + \dots + c_p t^{n_p-n_m} e^{a_p t} &= 0 \quad \forall t \end{aligned}$$

Now let $t = 0$. Then, $c_m = 0$, a contradiction. This proves the independence of the set $\{\varphi_{n,a}(t)\}$.

Next, observe that $B = \left\{ \frac{1}{(s-a)^{n+1}} : a \in k \text{ and } n = 0, 1, \dots \right\}$ is a basis for $k^0(s)$. Since

each element of $k^0(s)$ is a proper rational function, they may be expanded into partial fractions. In detail, let $\frac{a_0s^p+\dots+a_p}{b_0s^q+\dots+b_q}$ be a proper fraction with $p < q$. Then

$$\begin{aligned} \frac{a_0s^p}{b_0s^q+\dots+b_q} &= \frac{a_0s^p}{(s-a_1)^{n_1}\dots(s-a_l)^{n_l}} && \text{Fundamental Theorem of Algebra} \\ &= \sum_{j=1}^l \sum_{i=1}^{n_i} \frac{d_i}{(s-a_j)^{n_i}} && \text{Method of Partial Fractions} \end{aligned}$$

Thus, B spans $k^0(s)$, and B is easily seen to be independent. Finally, observe that the bijective correspondence, $\varphi_{n,a}(t) \longleftrightarrow \frac{1}{(s-a)^{n+1}}$ where $a \in k$ and $n = 0, 1, \dots$, induces a linear isomorphism between $E[[t]]$ and $k^0(s)$. This linear isomorphism creates a Laplace transform which can be used to prove Equation 2.6.

Definition 2.4.1. Define the *formal Laplace transform* $L : E[[t]] \longrightarrow k^0(s)$ to be the k -linear map determined by the formula

$$L(t^n e^{at}) = \frac{n!}{(s-a)^{n+1}}.$$

The *formal inverse Laplace transform* L^{-1} is defined on the basis B by

$$L^{-1}\left(\frac{1}{(s-a)^{n+1}}\right) = \frac{t^n}{n!} e^{at}.$$

If $A(t) \in M_n(E[[t]])$ is an $n \times n$ matrix of formal power series, then $L(A(t)) \in M_n(k^0(s))$ is the $n \times n$ matrix of proper rational functions obtained by applying L to each entry of the matrix $A(t)$. There is a similar interpretation of $L^{-1}(B(s))$ for $B(s) \in M_n(k^0(s))$.

Note that if $A \in M_n(k)$ then $e^{At} = \sum_{m=0}^{\infty} \frac{A^m}{m!} t^m$ is the unique matrix $\Phi(t) \in M_n(k[[t]])$ which solves the initial value problem

$$\Phi'(t) = A\Phi(t), \quad \Phi(0) = I_n.$$

To prove Equation 2.6, the individual entries of e^{At} must be shown to actually lie in the subspace $E[[t]]$. This is the most important equation of the paper, for it turns a matrix exponential function into an exponential function with matrix coefficients. This transformation is key toward applying the characteristic polynomial to a matrix. To prove it, three corollaries of the formal Laplace and inverse Laplace transformation must be proved.

Lemma 2.4.2. $L(y'(t)) = sL(y(t)) - y(0)$ for all $y(t) \in E[[t]]$.

Lemma 2.4.3. If $r(s) = q(s)/p(s) \in k^0(s)$ is a rational function with $p(s)$ monic of degree n , then $L^{-1}(r(s))(0)$ is the coefficient of s^{n-1} in $q(s)$.

Lemma 2.4.4. If $A \in M_n(k)$ then

$$L^{-1}((sI - A)^{-1}) = e^{At}.$$

Proof. Lemma 2.4.2

Because L is linear, it is sufficient to check this for $y(t) = \frac{t^n}{n!}e^{at}$. If $n \geq 1$, then

$$\begin{aligned} y'(t) &= \frac{t^{n-1}}{(n-1)!}e^{at} + a\frac{t^n}{n!}e^{at} \\ \Rightarrow L(y'(t)) &= \frac{1}{(s-a)^n} + \frac{a}{(s-a)^{n+1}} \\ &= \frac{s}{(s-a)^{n+1}} \\ &= sL(y(t)) - y(0). \end{aligned} \tag{2.8}$$

If $n = 0$ then using Equation 2.8, $L(y'(t)) = \frac{s}{s-a} = sL(y(t)) - y(0)$. □

Proof. Lemma 2.4.3

From Definition 2.4.1 the constant term ($t = 0$) of the series $L^{-1}\left(\frac{1}{(s-a)^{n+1}}\right)$ is 0 unless $n = 0$, in which case the constant term is 1. Also note that, by the Fundamental Theorem of Algebra, $p(s) = (s-a_1)^{r_1}(s-a_2)^{r_2}\dots(s-a_m)^{r_m}$. Decomposing $r(s)$ into partial fractions gives

$$r(s) = \frac{c_{11}}{(s-a_1)^1} + \frac{c_{12}}{(s-a_1)^2} + \dots + \frac{c_{mr_m}}{(s-a_m)^{r_m}} = \sum_{i=1}^m \sum_{j=1}^{r_j} \frac{c_{ij}}{(s-a_i)^j}.$$

The constant term of $L^{-1}(r(s))$ may now be calculated.

$$\begin{aligned} L^{-1}(r(s))(0) &= L^{-1}\left(\frac{c_{11}}{(s-a)^1}\right)(0) + L^{-1}\left(\frac{c_{12}}{(s-a)^2}\right)(0) + \dots \\ &+ L^{-1}\left(\frac{c_{m1}}{(s-a)^1}\right)(0) + \dots + L^{-1}\left(\frac{c_{mr_m}}{(s-a)^{r_m}}\right)(0) \\ &= c_{11} \cdot 1 + c_{12} \cdot 0 + \dots + c_{m1} \cdot 1 + \dots + c_{mr_m} \cdot 0 \\ &= c_{11} + c_{21} + \dots + c_{m1} \end{aligned}$$

Using the partial fraction expansion of $r(s)$ above, combine terms to get

$$r(s) = \frac{c_{11}O(s^{n-1}) + c_{12}O(s^{n-2}) + \dots + c_{mr_j}O(s^{n-r_j})}{(s-a_1)^{r_1}(s-a_2)^{r_2}\dots(s-a_m)^{r_m}}.$$

Therefore, the coefficient of s^{n-1} in $q(s)$ is $c_{11} + c_{21} + \dots + c_{m1}$. This proves Lemma 2.4.3. $\square$

Proof. Lemma 2.4.4

Let $p(s) = \det(sI - A)$ be the characteristic polynomial of A . If $B(s) = (sI - A)^{-1}$ then the adjoint formula $\text{adj}(X) = (X^{-1})\det(X)$ applied to $sI - A$ gives

$$B(s)_{uv} = \frac{A'_{uv}}{\det(sI - A)} = \frac{c_{uv}}{p(s)}$$

where, A' is the matrix of cofactors of $sI - A$ so A'_{uv} is the cofactor A_{vu} of A .

Note that if $u = v$, then c_{uv} is a monic polynomial of degree $n - 1$ and if $u \neq v$ then the degree of c_{uv} is less than $n - 1$. Then Lemma 2.4.3 shows that $\Phi(t) = L^{-1}(B(s))$ has $\Phi(0) = I_n$. This is easily proved by realizing that all off-diagonal components of $\Phi(0)$ are polynomials of degree less than $n - 1$. For these elements, the $(n - 1)^{\text{th}}$ degree term is 0. If $u = v$, then from Definition 2.4.1, $c_{uv} = 1$.

Since $(sI - A)B(s) = I$, it follows that $sB(s) - I = AB(s)$. Since $B(s) \in k^0(s)$, it follows that $L^{-1}(B(s)) \in E[[t]]$ and $y(t) = L^{-1}(B(s)) = \Phi(t)$. Then, $L^{-1}(B(s))(0) = \Phi(0) = y(0)$ and $y'(t) = \Phi'(t)$. Now Lemma 2.4.2 applied to $B(s)$ yields

$$L(\Phi'(t)) = sB(s) - I_n = AB(s) = AL(\Phi(t)), \text{ and } \Phi'(t) = A\Phi(t).$$

From Definition 2.4.1, e^{At} is the unique solution to the matrix $\Phi(t)$ and

$$L^{-1}((sI - A)^{-1}) = \Phi(t) = e^{At}.$$

$\square$

To prove Equation 2.6, apply the above results to matrices.

$$(sI - A)^{-1} = \frac{\text{adj}(sI - A)}{(s - a_1)^{n_1} \dots (s - a_m)^{n_m}} = M$$

$$\text{where, } \text{adj}(sI - A)_{f,g} = \begin{cases} O(n-1) & \text{if } f = g \\ O(n-2) & \text{if } f \neq g \end{cases}.$$

Then $M(i, j) = \frac{q(s)}{p(s)}$, where $O(q) < O(p)$, and $M(f, g)$ may be written in partial fractions.

$$M(f, g) = \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{c_{ij}(f, g)}{(s - a_i)^j}$$

Now, M may be written as a sum of matrices

$$M = \sum_{i=1}^m \sum_{j=1}^{n_i} \begin{bmatrix} \frac{c_{ij}(1,1)}{(s-a_i)^{n_j}} & \cdots & \frac{c_{ij}(1,n)}{(s-a_i)^{n_j}} \\ \vdots & \ddots & \vdots \\ \frac{c_{ij}(1,n)}{(s-a_i)^{n_j}} & \cdots & \frac{c_{ij}(n,n)}{(s-a_i)^{n_j}} \end{bmatrix} = \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{M_{ij}}{(s - a_i)^{n_j}},$$

and Equation 2.6 may now be proved:

$$\begin{aligned} L^{-1}((sI - A)^{-1}) &= L^{-1} \left(\sum_{i=1}^m \sum_{j=1}^{n_i} \frac{M_{ij}}{(s - a_i)^{n_j}} \right) = \sum_{i=1}^m \sum_{j=1}^{n_i} M_{ij} \left(\frac{t^{j-1}}{(j-1)!} e^{a_i t} \right) \\ &= \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} t^j e^{a_i t} = \sum_{i=1}^m \sum_{j=1}^{n_i} M_{ij} \varphi_{j, a_i}(t) \\ &= e^{At} \text{ (from Corollary 2.4.4)} \end{aligned}$$

Main Theorem. (Cayley-Hamilton). If $p_A(s)$ is the characteristic polynomial of A then $p_A(A) = 0$.

Proof. The characteristic polynomial of A , $p_A(s) = \prod_{i=1}^m (s - a_i)^{n_i}$, determines the series $\varphi_{j, a_i}(t) = t^j e^{a_i t}$ which appear in Equation 2.6 since $p_A(s)$ is the denominator of every entry of $(sI - A)^{-1}$, where a_i are the eigenvalues of A and n_i are the respective multiplicities of a_i . Differentiate Equation 2.6 l times and evaluate at 0 to get

$$A^l = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} (t^j e^{a_i t})^{(l)}(0). \quad (2.9)$$

This creates a set of tools for use with the characteristic polynomial of A . To create A^l , evaluate the series at $t = 0$. Expand $\varphi_{j, a_i}(t) = t^j e^{a_i t}$ in a series.

$$\varphi_{j, a_i}(t) = t^j \sum_{n=0}^{\infty} \frac{(a_i t)^n}{n!} = \sum_{n=0}^{\infty} \frac{a_i^n t^{n+j}}{n!}$$

Differentiate φ_{j,a_i} l times.

$$\varphi_{j,a_i}^{(l)}(t) = \sum_{n=0}^{\infty} a_i^n \frac{n! (n+j)!}{(n+j-l)!} t^{n+j-l}$$

Evaluate at $t = 0$ and observe that the expression is non-zero only if $n+j-l = 0$ and $l \geq j$.

Note also that $\varphi_{l,0}(t) = t^l$. Therefore,

$$\varphi_{j,a_i}^{(l)}(0) = \frac{l!}{(l-j)!} a_i^{l-j} = \varphi_{l,0}^{(j)}(a_i). \quad (2.10)$$

Let $q(t) = c_0 + c_1 t + \dots + c_N t^N$ be an arbitrary polynomial. Then

$$\begin{aligned} q(t) &= \sum_{l=0}^N c_l t^l, \text{ and} \\ q^{(j)}(a_i) &= \sum_{l=0}^N c_l \frac{l!}{(l-j)!} a_i^{l-j} \\ &= \sum_{l=0}^N c_l \varphi_{l,0}^{(j)}(a_i). \end{aligned} \quad (2.11)$$

Equations 2.9, 2.10 and 2.11 give

$$\begin{aligned} q(A) = \sum_{l=0}^N c_l A^l &= \sum_{i=1}^m \sum_{j=0}^{n_i-1} \sum_{l=0}^N c_l M_{ij} \varphi_{j,a_i}^{(l)}(0) \\ &= \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} \sum_{l=0}^N c_l \varphi_{l,0}^{(j)}(a_i) = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} q^{(j)}(a_i). \end{aligned} \quad (2.12)$$

Since the characteristic polynomial $p_A(s) = \prod_{i=1}^m (s - a_i)^{n_i}$ has $p^{(j)}(a_i) = 0$ for $0 \leq j \leq n_i - 1$,

Equation 2.12 gives

$$p_A(A) = \sum_{i=1}^m \sum_{j=0}^{n_i-1} M_{ij} p^{(j)}(a_i) = 0.$$

□

CHAPTER 3

ANALYSIS AND APPLICATIONS

3.1 Analysis of Alternate Proofs

3.1.1 Characteristics

The Cayley-Hamilton theorem is one of the most powerful and classic matrix theory theorems. Many applications derive their results from this 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 characteristic polynomial.

Straubing's combinatorial proof of the Cayley-Hamilton theorem exploits three aspects of $p_A(A)$. First, it elegantly explains the relationship between positive and negative terms of $p_A(A)$. A two-color graph representing the coefficients and variables of the polynomial reveals that the graph of $p_A^+(A_{ij})$ and $p_A^-(A_{ij})$ differ only in coloring. This symmetry presents a visual understanding of why $p_A(A) = 0$. Second, the proof illuminates the importance of n . It is the cornerstone for the entire proof. $p_A(A)$ is a sum of n products of elements of A . This constraint is the main force behind the theorem that $p_A^+(A_{ij})$ and $p_A^-(A_{ij})$ differ only in coloring. The necessity and sufficiency of n illuminates another aspect of why A^k may be represented by a function of degree at most n . Third, the proof introduces a cyclic property to $p_A(A)$. The coefficients of $p_A(A)$ are partial permutations and therefore may be represented by adjoint cycles.

This version of Rosoff's topological proof is almost identical to Horn's proof [HJ85]. However, whereas Horn uses induction to prove $p_A(T) = 0$, Rosoff's version exploits the continuity of $p_A(A)$. This alteration increases the elegance of Horn's proof by using the density of diagonal matrices with respect to $p_A(A)$.

Whereas the previous two proofs use cycles and continuity to prove the Cayley-Hamilton theorem, Adkins and Davidson's Z-transform proof exploits the fact that $p_A(A)$ has only

finitely many different powers of A , coupled with the multiplicity of eigenvalues, to prove $p_A(A) = 0$.

3.1.2 Algorithms

The analysis of alternate proofs of the Cayley-Hamilton theorem helps shed new light on the ramifications of the result, especially in regard to algorithms. Each proof offers insight into more efficient algorithms for generating matrix powers.

In proving the Cayley-Hamilton theorem through combinatorics and graph theory, Straubing actually creates an algorithm for finding the coefficients of $p_A(x)$. He proves that a coefficient a_i of $p_A(x)$ is the sum of a function of all partial permutations with certain constraints. Using this algorithm, individual coefficients of $p_A(x)$ may be generated without having to generate all the coefficients and without ever knowing the eigenvalues of A .

The version of Rosoff's topological proof illustrated in this thesis suggests a possible way to find A^n using diagonal matrices. It explains how to approximate any matrix A by a diagonalizable matrix A_d arbitrarily close to A . A_d may then be used to find A_d^n and to compare the distance between A_d^n and A^n .

Matrix multiplication is an inefficient process. Adkins and Davidson's Z-transform proof offer an alternative to matrix multiplication and thus offer a possibly more efficient way to compute A^n . They reduce $p_A(A)$, a polynomial in matrices, to a polynomial in real numbers multiplied by matrix. Although the computations to find this one matrix are time-consuming, they need only be computed once. Thereafter, the polynomial in real numbers is all that is required to find $p_A(A)$ and thus A^i for all i .

3.2 Application of the Cayley-Hamilton Theorem

A very common application of the Cayley-Hamilton Theorem is to use it to find A^n , usually for very large powers of n . However, many of the techniques involved require the use of the eigenvalues of A . Raghieb Abu-Saris and Wajdi Ahmad [ASA05] use the Cayley-Hamilton theorem to show this is not necessary.

Consider a $k \times k$ matrix A ($k \geq 2$), and let

$$p(\lambda) = \det(\lambda I - A) = \lambda^k + \sum_{j=0}^{k-1} a_j \lambda^j \quad (3.1)$$

be the characteristic polynomial of A . By the Cayley-Hamilton theorem $p(A) = 0$ and by the division algorithm,

$$A^n = q_n(A)p(A) + r_n(A) = r_n(A) \quad (n \geq k) \quad (3.2)$$

where the remainder polynomial r_n is of degree at most $k-1$. Therefore A^n can be computed once the coefficients of r_n are determined. To do this, write

$$r_n(\lambda) = \sum_{j=0}^{k-1} b_j(n) \lambda^j.$$

From Equation 3.1, $A^k = -\sum_{j=0}^{k-1} a_j A^j$, and for $n = k$

$$b_j(k) = -a_j \quad (j = 0, \dots, k-1) \quad (3.3)$$

from Equation 3.2. Furthermore,

$$\begin{aligned} \sum_{j=0}^{k-1} b_j(n+1) A^j &= A^{n+1} = A A^n = \sum_{j=0}^{k-1} b_j(n) A^{j+1} = \sum_{j=1}^k b_{j-1}(n) A^j \\ &= b_{k-1}(n) A^k + \sum_{j=1}^{k-1} b_{j-1}(n) A^j \\ &= -b_{k-1}(n) \sum_{j=0}^{k-1} a_j A^j + \sum_{j=1}^{k-1} b_{j-1}(n) A^j \\ &= -b_{k-1}(n) \left[a_0 A^0 + \sum_{j=1}^{k-1} a_j A^j \right] + \sum_{j=1}^{k-1} b_{j-1}(n) A^j \\ &= -a_0 b_{k-1}(n) I + \sum_{j=1}^{k-1} [-a_j b_{k-1}(n) + b_{j-1}(n)] A^j. \end{aligned}$$

Hence, comparing coefficients,

$$b_j(n+1) = -a_j b_{k-1}(n) + b_{j-1}(n) \quad (j = 0, \dots, k-1; \quad b_{-1}(n) = 0). \quad (3.4)$$

Observe that Equations 3.3 and 3.4 provide a recursive algorithm for computing A^n that does not require determination of the eigenvalues of A . Moreover, these recursive formulas can be solved explicitly in terms of the coefficients of the associated characteristic polynomial.

Example 3.2.1. If A is the 3×3 matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ -2 & 3 & 1 \\ -3 & 1 & 4 \end{bmatrix}$$

then the characteristic polynomial of A is

$$p(\lambda) = \lambda^3 - 7\lambda + 16\lambda - 12.$$

Therefore, $a_0 = -12$, $a_1 = 16$, and $a_2 = -7$, so by Equations 3.3 and 3.4 the coefficients $b_j(n)$ of A^4 are computed as follows:

$$\begin{aligned} A^4 &= b_0(4)A^0 + b_1(4)A^1 + b_2(4)A^2 \\ b_0(3) &= -a_0 = 12 \\ b_1(3) &= -a_1 = -16 \\ b_2(3) &= -a_2 = 7 \\ b_0(4) &= -a_0 b_2(3) + b_{-1}(3) = 12 \cdot 7 = 84 \\ b_1(4) &= -a_1 b_2(3) + b_0(3) = -16 \cdot 7 + 12 = -100 \\ b_2(4) &= -a_2 b_2(3) + b_1(3) = 7 \cdot 7 + -16 = 33 \\ A^4 &= 84I - 100A + 33A^2 \end{aligned}$$

REFERENCES

- [AD03] William A. Adkins and Mark G. Davidson, The Cayley Hamilton and Frobenius Theorems Via the Laplace Transformation, Linear Algebra and its Applications **371** (2003), 147–152.
- [ASA05] Raghib Abu-Saris and Wajdi Ahmad, Avoiding Eigenvalues in Computing Matrix Powers, The Mathematical Association of America **112** (2005), 450–454.
- [Ber98] Cindy Bervig, Arthur Cayley’s Remarkable Theorem, The Adams State College Mathematics, Computer Science, and Physics Student Journal **1** (1997-1998), 15–18, available from <http://slv.adams.edu/journal/journal.rap>.
- [Cay57] Arthur Cayley, A Memoir on the Theory of Matrices, available from <http://www.jstor.org>, 1857.
- [HJ85] Roger A. Horn and Charles R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
- [Jac90] Bill Jacob, Linear Algebra, Freeman, 1990.
- [Kni05] Oliver Knill, Determinants, available from <http://www.math.harvard.edu>, 2005.
- [Mey00] Carl D. Meyer, Matrix Analysis, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000, available from <http://www.matrixanalysis.com/Chapter3.pdf> and [Chapter7.pdf](http://www.matrixanalysis.com/Chapter7.pdf).
- [MPG98] Koch Marc Pollefeys, Reinhard and Luc Van Gool, Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters, available from <http://www.lems.brown.edu>, 1998.
- [OR96] J. J. O’Connor and E. F. Robertson, Matrices and determinants, available from <http://www-groups.dcs.st-and.ac.uk/~history/HistTopics>, 1996.
- [Ros95] Jeffrey A. Rosoff, A Topological Proof of the Cayley-Hamilton Theorem, Missouri J. Math. Sc. **7** (1995), 63–67.
- [Sto71] Mirko Stojakovic, On a Theorem of G.P. barker on Triangular Matrices, Mathematics Magazine **44** (1971), no. 3.
- [Str83] Howard Straubing, A Combinatorial Proof of the Cayley-Hamilton Theorem, Discrete Mathematics **43** (1983), 273–279.
- [Wik04] Wikipedia, Arthur Cayley, available from <http://en.wikipedia.org>, 2004.
- [Wik05] ———, William Rowan Hamilton, available from <http://en.wikipedia.org>, 2005.
- [Wil05] D.R. Wilkins, Linear Operators and the Cayley-Hamilton Theorem, available from <http://www.maths.tcd.ie/pub/HistMath/People/Hamilton>, 2005.

APPENDIX A

THE RELATIONSHIP BETWEEN THE ADJOINT, THE DETERMINANT, AND THE CHARACTERISTIC POLYNOMIAL

Definition A.0.2. The i, j minor matrix M_{ij} of B is the $(n-1) \times (n-1)$ sub-matrix that results from deleting the i^{th} row and the j^{th} column of B .

Definition A.0.3. The i, j cofactor matrix of B is $C_{ij} = (-1)^{i+j} \det(M_{ij})$.

Definition A.0.4. If $A = (a_{ij})$ is an $n \times n$ matrix and $i, j \in \{1, 2, \dots, n\}$, then the determinant, computed by the Laplace expansion of the first row or first column, is

$$\begin{aligned} \det(A) &= a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in} \\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + \dots + a_{nj}C_{nj} \end{aligned}$$

Definition A.0.5. The ij^{th} entry of the adjugate or classical adjoint of an $n \times n$ matrix A is $\text{adj}(A)_{ij} = C_{ij}$.

Theorem A.0.6. If A is an $n \times n$ matrix and I is an $n \times n$ identity matrix, then $(\det(A))I_{kl} = (A \cdot \text{adj}(A))_{kl}$, where $k, l \in \{1, 2, \dots, n\}$ represents the kl^{th} entry of a matrix.

Proof.

$$\begin{aligned} (A \cdot \text{adj}(A))_{kl} &= \sum_{m=1}^n a_{km} \text{adj}(A)_{ml} \\ &= \sum_{m=1}^n a_{km} C_{lm} \\ &= \begin{cases} \det(A) & \text{for } k = l, \\ 0 & \text{for } k \neq l. \end{cases} \end{aligned}$$

The last equality comes realizing that the sum is $\det(B)$ where B is matrix A with row l replaced by row k . If B has two equal rows, then $\det(B) = 0$; otherwise, $B = A$ and $\det(B) = \det(A)$. □

APPENDIX B

PERMUTATION DEFINITION OF THE DETERMINANT OF A MATRIX

The determinant of an $n \times n$ matrix A can be defined using permutations [Jac90], as follows.

Definition B.0.7. $\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots a_{n\sigma_n}$

To see this, the determinant must be characterized. Then, it must be shown that Definition B.0.7 satisfies the resulting criteria.

B.1 Characterization of the Determinant

Definition B.1.1. For the complex field, $\mathbb{C}$, denote by $M_{n \times n}(\mathbb{C})$ the set of $n \times n$ matrices with entries from $\mathbb{C}$. Suppose

$$\Delta : M_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$$

is a scalar-valued function on $M_{n \times n}(\mathbb{C})$. Δ is *n-linear* if

(i) Whenever $A \in M_{n \times n}(\mathbb{C})$ and B is obtained from A by multiplying a single row of A by $k \in \mathbb{C}$, $\Delta(B) = k\Delta(A)$, and

(ii) Whenever $A, B, C \in M_{n \times n}(\mathbb{C})$ are identical except for the i th rows and the i th row of C is the sum of the i th rows of A and B , then $\Delta(C) = \Delta(A) + \Delta(B)$.

In addition, an n -linear function Δ is called *alternating* if whenever $A \in M_{n \times n}(\mathbb{C})$ has two identical rows, $\Delta(A) = 0$. Any function $\Delta : M_{n \times n} \rightarrow \mathbb{C}$ which is n -linear and alternating and for which $\Delta(I_n) = 1$ is called a *determinant function*.

To prove existence, employ the Laplace expansion. To prove uniqueness, the fundamental relationship between the determinant functions and row operations in the following theorem must be defined.

Theorem B.1.2. Suppose that $\Delta : M_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$ is a determinant function. Let $A \in M_{n \times n}(\mathbb{C})$. Then

- (i) If B is obtained from A by multiplying a single row of A by $k \in \mathbb{C}$, then $\Delta(B) = k\Delta(A)$.
- (ii) If B is obtained from A by adding a multiple of one row to another, then $\Delta(B) = \Delta(A)$.
- (iii) If B is obtained from A by interchanging two rows, then $\Delta(B) = -\Delta(A)$.

Proof. (i) This is true by definition (it is the first condition in Definition B.1.1).

(ii) Suppose that $A = (a_{ij})$ and $B = (b_{ij})$ where $a_{ij} = b_{ij}$ for all j with $i \neq t$ and $b_{tj} = a_{tj} + ka_{sj}$ where $t \neq s$. If A' is the matrix obtained from A by replacing the t th row of A by the s th row of A , then A' has two identical rows. It follows from the alternating property of Δ that $\Delta(A') = 0$. Since Δ is n -linear, condition (i) of Definition B.1.1 shows that the matrix A'' obtained from A' by multiplying the t th row by the scalar k also satisfies $\Delta(A'') = 0$. Since the matrix B is identical to the matrices A and A'' at each row except the t th, which is the sum of the t th rows of A and A'' , condition (ii) of n -linearity gives $\Delta(B) = \Delta(A) + \Delta(A'') = \Delta(A)$, as desired.

(iii) Suppose that the i th and j th rows are interchanged. Let R_i denote the i th row of A , and let R_j denote the j th row of A . Consider the three matrices

$$A = \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \end{pmatrix}, B = \begin{pmatrix} \vdots \\ R_j \\ \vdots \\ R_i \\ \vdots \end{pmatrix}, C = \begin{pmatrix} \vdots \\ R_i + R_j \\ \vdots \\ R_i + R_j \\ \vdots \end{pmatrix}$$

Since Δ is alternating, Definition B.1.1 gives that $\Delta(C) = 0$. Repeated applications of the

second condition of n -linearity yields

$$\begin{aligned}
 0 = \Delta(C) &= \Delta \begin{pmatrix} \vdots \\ R_i + R_j \\ \vdots \\ R_i + R_j \\ \vdots \end{pmatrix} = \Delta \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_i + R_j \\ \vdots \end{pmatrix} + \Delta \begin{pmatrix} \vdots \\ R_j \\ \vdots \\ R_i + R_j \\ \vdots \end{pmatrix} \\
 &= \Delta \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_i \\ \vdots \end{pmatrix} + \Delta \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \end{pmatrix} + \Delta \begin{pmatrix} \vdots \\ R_j \\ \vdots \\ R_i \\ \vdots \end{pmatrix} + \Delta \begin{pmatrix} \vdots \\ R_j \\ \vdots \\ R_j \\ \vdots \end{pmatrix} \\
 &= \Delta \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \end{pmatrix} + \Delta \begin{pmatrix} \vdots \\ R_j \\ \vdots \\ R_i \\ \vdots \end{pmatrix} = \Delta(A) + \Delta(B)
 \end{aligned}$$

It follows that $\Delta(B) = -\Delta(A)$. □

An additional definition of *upper* and *lower triangular* matrices and their determinants will allow us to prove uniqueness of the determinant function.

Definition B.1.3. A square matrix $A = (a_{ij})$ is called *upper triangular* if $a_{ij} = 0$ whenever $i > j$. Analogously, A is called *lower triangular* if $a_{ij} = 0$ whenever $i < j$. A is called *diagonal* if $a_{ij} = 0$ whenever $i \neq j$.

Lemma B.1.4. Suppose that Δ is a determinant function. If $A = (a_{ij})$ is triangular (upper or lower), then $\Delta(A) = a_{11}a_{22} \cdots a_{nn}$.

Proof. Assume that A is triangular (upper or lower) and each diagonal entry a_{ii} of A is nonzero. Applying condition (i) of n -linearity n times shows that $\Delta(A) = a_{11}a_{22} \cdots a_{nn}\Delta(B)$

where B is a triangular matrix with all diagonal entries 1. (To produce the matrix B simply multiply the i th row of B by a_{ii}^{-1} .) Repeated applications of Definition B.1.1(ii), together with Theorem B.1.2(ii), shows that $\Delta(B) = \Delta(R)$ where R is the reduced row-echelon form of B . Since B has rank n , necessarily $R = I_n$. Thus, $\Delta(A) = a_{11}a_{22} \cdots a_{nn}\Delta(I_n) = a_{11}a_{22} \cdots a_{nn}$ since $\Delta(I_n) = 1$ for any determinant function Δ .

In case some diagonal entry a_{ii} of A is zero, then the reduced row-echelon form R of A has an entire row zero. Therefore, $\Delta(R) = 0$. Since $\Delta(A)$ differs from $\Delta(R)$ by a scalar multiple, it follows that $\Delta(A) = 0 = a_{11}a_{22} \cdots a_{nn}$. $\square$

Theorem B.1.5. (*The Characterization of the Determinant*) *There is a unique function $\Delta : M_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$ which is n -linear and alternating and for which $\Delta(I_n) = 1$.*

Proof. Suppose that Δ and Δ' are both determinant functions on $M_{n \times n}(\mathbb{C})$. Let $A \in M_{n \times n}(\mathbb{C})$ and suppose that R is the reduced row-echelon form of A . The matrix R is upper triangular, so applying Lemma B.1.4, $\Delta(R) = \Delta'(R)$ because Δ and Δ' only depend upon the diagonal elements of R . However, R can be obtained from A by a sequence of elementary row operations, so repeatedly applying Theorem B.1.2 for this sequence of elementary operations shows that $\Delta(A) = \Delta'(A)$ as well. $\square$

For the remainder of this section, use the standard notation $\det(A)$ for the determinant of an $n \times n$ matrix A , defined by the Laplace expansion of the determinant.

B.2 Permutation Definition of the Determinant

The determinant has been characterized and proved unique. It remains to be shown that Definition B.0.7 satisfies the uniqueness criteria detailed in Theorem B.1.5.

Definition B.2.1. A *permutation of n* is an n -tuple $\sigma = (i_1, i_2, \dots, i_n)$ where $i_1, i_2, \dots, i_n$ are the n distinct integers $1, 2, \dots, n$ (but not necessarily listed in the usual order). Denote the j th entry in a permutation σ by σ_j . The collection of all permutations of n is denoted by S_n .

Definition B.2.2. Suppose that $\sigma = (i_1, i_2, \dots, i_n) \in S_n$. If $1 \leq j < k \leq n$ and $i_j > i_k$, let $i_j > i_k$ be an inversion of σ . For any permutation σ define the *sign* of σ , denoted $\text{sgn}(\sigma)$, by $\text{sgn}(\sigma) = +1$ if the total number of inversions of σ is even and by $\text{sgn}(\sigma) = -1$ if the total number of inversions of σ is odd. When $\text{sgn}(\sigma) = +1$, σ is called *even*, and when $\text{sgn}(\sigma) = -1$, σ is called *odd*.

First prove that Definition B.0.7 satisfies Definition B.1.1, or the condition of n -linearity, through two lemmas.

Lemma B.2.3. Suppose that A is a square matrix and B is obtained from A by multiplying the s th row by k . Then $\det(B) = k \cdot \det(A)$.

Proof. Assume that $A = (a_{ij})$ and $B = (b_{ij})$. Suppose that B is obtained from A by multiplying the s th row by k . Then $b_{ij} = a_{ij}$ for all $i \neq s$ and all j , while $b_{sj} = ka_{sj}$ for all j . By Definition B.0.7

$$\begin{aligned} \det(B) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) b_{1\sigma_1} b_{2\sigma_2} \cdots b_{s\sigma_s} \cdots b_{n\sigma_n} \\ &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots k a_{s\sigma_s} \cdots a_{n\sigma_n} \\ &= k \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots a_{s\sigma_s} \cdots a_{n\sigma_n} \\ &= k \cdot \det(A) \end{aligned}$$

□

Lemma B.2.4. Suppose that the three matrices $A = (a_{ij})$, $B = (b_{ij})$, and $C = (c_{ij})$ are identical except for the k th row, where there is an assumption that for all j that $a_{kj} + b_{kj} = c_{kj}$. Then $\det(A) + \det(B) = \det(C)$.

Proof. The hypotheses give that for all j $a_{ij} = b_{ij} = c_{ij}$ whenever $i \neq k$ and $a_{kj} + b_{kj} = c_{kj}$ otherwise. It then follows from Definition B.0.7 that

$$\begin{aligned} \det(C) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) c_{1\sigma_1} c_{2\sigma_2} \cdots (a_{k\sigma_k} + b_{k\sigma_k}) \cdots c_{n\sigma_n} \\ &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots a_{k\sigma_k} \cdots a_{n\sigma_n} + \sum_{\sigma \in S_n} \text{sgn}(\sigma) b_{1\sigma_1} b_{2\sigma_2} \cdots b_{k\sigma_k} \cdots b_{n\sigma_n} \\ &= \det(A) + \det(B) \end{aligned}$$

□

Now prove the alternating property of the determinant. First a technical result about permutations is needed.

Lemma B.2.5. *Suppose that $1 \leq s < t \leq n$, $\sigma = (i_1, i_2, \dots, i_n) \in S_n$, and σ' is obtained from σ by interchanging i_s and i_t . Then $\text{sgn}(\sigma') = -\text{sgn}(\sigma)$.*

Proof. Consider the permutation $\sigma = (\dots, i_s, \dots, i_t, \dots)$ where $s < t$. Without loss of generality, assume $i_s < i_t$. Then σ' is the permutation $(\dots, i_t, \dots, i_s, \dots)$ which is identical to σ at every place, except that the t th and s th entries are reversed. Consider the collection of entries $i_{s+1}, i_{s+2}, \dots, i_{t-1}$ of σ . Assume that among this collection there are e numbers less than i_s , f numbers larger than i_s but less than i_t , and g numbers larger than i_t . When i_s and i_t are interchanged, the number of inversions passing from σ to σ' is:

- (i) Decreased by e and increased by $f + g$ (those inversions involving i_s but not i_t)
- (ii) Decreased by g and increased by $e + f$ (those inversions involving i_t but not i_s)
- (iii) Increased by 1 (the inversion involving both i_t and i_s)

The total change in the number of inversions is

$$-e + (f + g) + -g + (e + f) + 1 = 2f + 1$$

Since $2f + 1$ is always odd, $\text{sgn}(\sigma') = -\text{sgn}(\sigma)$. □

Now establish the alternating property of the determinant function as defined in Definition B.0.7.

Lemma B.2.6. *If a square matrix A has two equal rows, then $\det(A) = 0$.*

Proof. Suppose that $A = (a_{ij})$ and $a_{sj} = a_{tj}$ for all j where $1 \leq s < t \leq n$. Partition S_n into two subsets C and C' , where C consists of those permutations in S_n with $i_s < i_t$ and C' consists of those permutations in S_n with $i_t < i_s$. Note that $C \cup C' = S_n$ and $C \cap C' = \emptyset$. It follows from the definitions that whenever $\sigma \in C$, then $\sigma' \in C'$. Moreover, each element of C' is σ' for some $\sigma \in C$, where σ' is as described in Lemma B.2.5. Since $a_{sj} = a_{tj}$ for all j ,

$a_{s\sigma_s} = a_{t\sigma'_t}$ and $a_{t\sigma_t} = a_{s\sigma'_s}$. Using Lemma B.2.5 together with the fact that $a_{i\sigma_i} = a_{i\sigma'_i}$ for all $i \neq s, t$,

$$\begin{aligned}
 \det(A) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &= \sum_{\sigma \in C} \operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &\quad + \sum_{\sigma \in C'} \operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &= \sum_{\sigma \in C} \operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &\quad + \sum_{\sigma \in C} \operatorname{sgn}(\sigma') a_{1\sigma'_1} \cdots a_{s\sigma'_s} \cdots a_{t\sigma'_t} \cdots a_{n\sigma'_n} \\
 &= \sum_{\sigma \in C} \operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &\quad + \sum_{\sigma \in C'} -\operatorname{sgn}(\sigma) a_{1\sigma_1} \cdots a_{s\sigma_s} \cdots a_{t\sigma_t} \cdots a_{n\sigma_n} \\
 &= 0
 \end{aligned}$$

□

As an immediate consequence of Lemmas B.2.3, B.2.4, and B.2.6, the following theorem is obtained.

Theorem B.2.7. *The function $\det : M_{n \times n}(\mathbb{C}) \rightarrow \mathbb{C}$ defined for $A = (a_{ij})$ by*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots a_{n\sigma_n}$$

is a determinant function in the sense of Definition B.1.1.