Rose-
Hulman
Undergraduate
Mathematics
Journal

The Cayley-Hamilton Theorem via the $\mathcal{Z}$-Transform

Casey Tsai [a]

Volume 13, No. 2, Fall 2012

Sponsored by

Rose-Hulman Institute of Technology
Department of Mathematics
Terre Haute, IN 47803
Email: mathjournal@rose-hulman.edu
http://www.rose-hulman.edu/mathjournal

[a]Mathematics Department, Louisiana State University

Rose-Hulman Undergraduate Mathematics Journal
Volume 13, No. 2, Fall 2012

The Cayley-Hamilton Theorem via the $\mathcal{Z}$-Transform

Casey Tsai

Abstract. The $\mathcal{Z}$-transform is usually defined and developed in a typical course on Difference Equations. We extend the transform to matrix valued sequences. A couple of key observations leads to a rather novel and simple proof of the Cayley-Hamilton theorem.

Acknowledgements: This paper is a result of an undergraduate research project initiated in the Fall of 2011 under the direction of Professor Mark Davidson and can be regarded as an outgrowth of the SMILE@LSU program. I gratefully acknowledge his guidance. Research supported by NSF grant DMS 0739382: LSU-VIGRE proposal and the LA-STEM program at LSU.

1 Introduction

Suppose A is an $n \times n$ matrix. The Cayley-Hamilton Theorem states that A satisfies its own characteristic polynomial. In other words, if c_A is the characteristic polynomial of A then $c_A(A) = 0$. This result was first stated and proved by Cayley [2] for $n = 2$ and $n = 3$ in 1858. Hamilton [4] proved the result for $n = 4$ in 1862. The general case was proved by Frobenius [3] in 1878. The Cayley-Hamilton Theorem has many uses in mathematics. For example, it is used to compute the inverse of A; the matrix exponential e^{tA}, which is needed for solving systems of differential equations; powers of A^k, as in Putzer's algorithm; and many others.

There are many proofs of the Cayley-Hamilton Theorem. For example, in Lang's book [7], the proof involves a rather intricate induction argument that shows a linear map over the complex numbers is triangulable. The proofs found in Hoffman and Kunze [5] are for an arbitrary field and require more sophisticated tools.

In the summer of 2011, I participated in the SMILE program at LSU. This is a VIGRE-NSF supported program for undergraduates that involves group research projects on various topics. A short course that I took was on difference equations, taught by Professor Mark Davidson and based on Peterson and Kelley's textbook Difference Equations [6]. The group project I was assigned involved the Putzer algorithm, a method for finding the k^{th} power of an $n \times n$ matrix, A. The Cayley-Hamiliton theorem was central to this method but stated by Kelley and Petersen without proof. This is unfortunate since, as we will demonstrate here, the main tool, namely the $\mathcal{Z}$-transform, had already been developed in the text. Extending the transform to matrix valued sequences and a couple of key observations produces a proof that is quite different from the classical proofs mentioned above but parallels the proof found in Adkins and Davidson [1] where the Laplace transform plays the central role.

In [1] the proof of the Cayley-Hamilton Theorem is first done over an algebraically closed field of characteristic zero. Then it is extended to any field of characteristic zero. The ring of formal power series is used and the issue of convergence does not come up. We could likewise do the same here but have opted for simplicity to restrict to the complex field $\mathbb{C}$.

The organization of this paper is as follows. In Section 2 we provide all the necessary background for understanding the $\mathcal{Z}$-transform of complex-valued sequences. We introduce the fundamental sequence $\varphi_{n,a}(k)$, where n is a nonnegative integer and $a \in \mathbb{C}$, that is central to the proof of the Cayley-Hamilton Theorem. Proposition 5 gives its $\mathcal{Z}$-transform. In Section 3 we extend the definitions to matrix-valued sequences. Proposition 6 gives the $\mathcal{Z}$-transform of the sequence of powers of A. With this at hand, Proposition 9 gives, by $\mathcal{Z}$-transform inversion, a formula for A^k in terms of the fundamental $\mathbb{C}$-valued sequences φ_{n,a_i}, where $a_1, \ldots, a_R$ are the eigenvalues of A. In Section 4 we conclude with a rather straightforward proof of the Cayley-Hamilton Theorem based on Proposition 9.

2 Notation, Definitions, and some Basic Results

For the sake of a self contained presentation we provide all the necessary proofs of formulas concerning the $\mathcal{Z}$-transform instead of simply quoting them from [6].

Definition 1. *Let n be a nonnegative integer. The* ***falling factorial*** *is the sequence $k^{\underline{n}}$, with $k = 0, 1, 2, \ldots$, given by the following formula*

$$k^{\underline{n}} = k(k-1)(k-2)\cdots(k-n+1).$$

If k were allowed to be a real variable then $k^{\underline{n}}$ could be characterized as the unique monic polynomial of degree n that vanishes at $0, 1, \ldots, n-1$. Observe also that $k^{\underline{n}}|_{k=n} = n!$.

Definition 2. *Let a be a complex number and n a nonnegative integer. The following sequences $\varphi_{n,a}$ are fundamental to all that we do.*

$$\varphi_{n,a}(k) = \begin{cases} \dfrac{a^{k-n}k^{\underline{n}}}{n!} & a \neq 0 \\ \delta_n(k) & a = 0, \end{cases}$$

where $\delta_n(k)$ is the sequence which is 0 for all $k \neq n$ and $\delta_n(n) = 1$.

Example 1. *For example, the sequences $\varphi_{0,2}(k)$, $\varphi_{1,2}(k)$ and $\varphi_{2,0}(k)$ are*

$$\begin{aligned} \varphi_{0,2}(k) = 2^k &= (1, 2, 4, 8, 16, \ldots) \\ \varphi_{1,2}(k) = 2^{k-1}k &= (0, 1, 4, 12, 32, \ldots) \\ \varphi_{2,0}(k) = \delta_2(k) &= (0, 0, 1, 0, 0, 0, 0 \ldots) \end{aligned}$$

There is a close connection between the two sequences

$$\frac{a^{k-n}k^{\underline{n}}}{n!} \quad \text{and} \quad \delta_n(k).$$

It is clear that we cannot allow $a = 0$ in the formula $\frac{a^{k-n}k^{\underline{n}}}{n!}$ because of the presence of negative powers when $k < n$. However, for $k < n$ the term $k^{\underline{n}}$ is identically 0. This leads to the following observation:

$$\lim_{a\to 0} \frac{a^{k-n}k^{\underline{n}}}{n!} = \delta_n(k), \tag{1}$$

where the limit is understood in the pointwise sense. Specifically, assume $a \neq 0$. If $k < n$ then $\frac{a^{k-n}k^{\underline{n}}}{n!}$ is identically 0. If $k = n$ then $\frac{a^{k-n}k^{\underline{n}}}{n!} = \frac{a^0 n!}{n!} = 1$. Finally, if $k > n$ then $\lim_{a\to 0} a^{k-n} = 0$, thus verifying Equation (1).

Lemma 2. *Let $\mathbf{D}$ denote the ordinary derivative operator. Let n be a nonnegative integer and $a \in \mathbb{C}$. We then have*

$$\varphi_{n,a}(k) = \left.\frac{\mathbf{D}^n x^k}{n!}\right|_{x=a},$$

where the notation $|_{x=a}$, is to be understood in the limit sense, that is, $\lim_{x\to a}$, as in Equation (1).

Proof.

$$\begin{aligned}\frac{\mathbf{D}^n(x^k)}{n!} &= \frac{k(k-1)(k-2)(k-3)...(k-n+1)x^{k-n}}{n!} \\ &= \frac{k^{\underline{n}}x^{k-n}}{n!}.\end{aligned}$$

Evaluating at $x = a$ in the case $a \neq 0$ gives $\varphi_{n,a}(k)$. If $a = 0$ then as explained above, we get

$$\lim_{x\to 0}\frac{k^{\underline{n}}x^{k-n}}{n!} = \delta_n(k) = \varphi_{n,0}(k).$$

□

Definition 3. *Let $y(k)$ be a sequence of complex numbers. We define the $\mathcal{Z}$-**transform** of y to be the function $\mathcal{Z}\{y\}(z)$, where z is a complex variable, by the following formula:*

$$\mathcal{Z}\{y\}(z) = \sum_{k=0}^{\infty}\frac{y(k)}{z^k}.$$

Remark. The $\mathcal{Z}$-transform is the formal power series in the variable $w = 1/z$, with coefficients $y(k)$. The $\mathcal{Z}$-transform is said to **exist** if there is a number $R > 0$ such that $\sum_{k=0}^{\infty}\frac{y(k)}{z^k}$ converges for $|z| > R$. As a power series in $w = 1/z$ it is easy to compute R: if r is the radius of convergence of $\sum_{k=0}^{\infty} y(k)w^k$ and $r \neq 0$, then $R = 1/r$ (if $r = \infty$ we set $R = 0$). The number r can be computed by standard methods in Calculus. To simplify the notation we frequently use the corresponding capital letter to denote the $\mathcal{Z}$-transform. Thus, for example, we have $Y(z) = \mathcal{Z}\{y\}(z)$.

Proposition 3. *The $\mathcal{Z}$-transform is linear and one-to-one on the set of sequences for which the $\mathcal{Z}$-transform exists.*

Proof. It is easy to see that the $\mathcal{Z}$-transform is linear. Suppose $y_1(k)$ and $y_2(k)$ are two sequences for which the $\mathcal{Z}$-transform exists and suppose $Y_1(z) = Y_2(z)$ on a common domain $|z| > R$. Let $w = 1/z$. Then we have

$$\sum_{k=0}^{\infty} y_1(k)w^k = \sum_{k=0}^{\infty} y_2(k)w^k, \tag{2}$$

for all $|w| < 1/R$. Differentiating Equation (2) n times and evaluating at $w = 0$ gives $n!y_1(n) = n!y_2(n)$. Since n is arbitrary it follows that $y_1 = y_2$ as sequences. This implies the $\mathcal{Z}$-transform is one-to-one. □

Knowing the the $\mathcal{Z}$-transform is one-to-one allows us to define the inverse $\mathcal{Z}$-transform, $\mathcal{Z}^{-1}$. For a convergent series

$$A(z) = \sum_{k=0}^{\infty}\frac{a(k)}{z^k},$$

we define

$$\mathcal{Z}^{-1}\left\{A(z)\right\}(k) = a(k).$$

This is well defined since the $\mathcal{Z}$-transform is one to one.

In the following Proposition we list several $\mathcal{Z}$-transform formulas and principles that are useful for us.

Proposition 4. *Suppose a is a nonzero complex number, $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, and $y(k)$ is a sequence for which the $\mathcal{Z}$-transform exists. Then*

1. $\mathcal{Z}\left\{a^k\right\}(z) = \frac{z}{z-a}$
2. $\mathcal{Z}\left\{a^k y(k)\right\}(z) = Y\left(\frac{z}{a}\right)$
3. $\mathcal{Z}\left\{y(k+n)\right\}(z) = z^n Y(z) - \sum_{m=0}^{n-1} y(m) z^{n-m}$ *(translation principle)*
4. $\mathcal{Z}\left\{(k+n-1)^{\underline{n}} y(k)\right\}(z) = (-1)^n z^n \mathbf{D}^n Y(z)$
5. $\mathcal{Z}\left\{k^{\underline{n}}\right\}(z) = \frac{n! z}{(z-1)^{n+1}}$

Proof.

1. $\mathcal{Z}\left\{a^k\right\}(z) = \sum_{k=0}^{\infty} \frac{a^k}{z^k} = \sum_{k=0}^{\infty} \left(\frac{a}{z}\right)^k = \frac{1}{1-\frac{a}{z}} = \frac{z}{z-a}$
2. $\mathcal{Z}\left\{a^k y(k)\right\}(z) = \sum_{k=0}^{\infty} \frac{a^k y(k)}{z^k} = \sum_{k=0}^{\infty} \frac{y(k)}{(z/a)^k} = Y(\frac{z}{a})$
3. $$\begin{aligned} \mathcal{Z}\left\{y(k+n)\right\}(z) &= \sum_{k=0}^{\infty} \frac{y(k+n)}{z^k} = z^n \sum_{k=n}^{\infty} \frac{y(k)}{z^k} \\ &= z^n \left(\sum_{k=0}^{\infty} \frac{y(k)}{z^k} - \sum_{m=0}^{n-1} \frac{y(m)}{z^m} \right) \\ &= z^n Y(z) - \sum_{m=0}^{n-1} y(m) z^{n-m}. \end{aligned}$$
4. $$\begin{aligned} \frac{d^n Y(z)}{dz^n} &= (-1)^n \sum_{k=0}^{\infty} k(k+1)\ldots(k+n-1) y(k) z^{-k-n} \\ &= \frac{(-1)^n}{z^n} \sum_{k=0}^{\infty} \frac{(k+n-1)^{\underline{n}} y(k)}{z^k} \\ &= \frac{(-1)^n}{z^n} \mathcal{Z}\left\{(k+n-1)^{\underline{n}} y(k)\right\}(z). \end{aligned}$$

5. Let $y(k) = 1^k = 1$. Then $Y(z) = \mathcal{Z}\{1\}(z) = \frac{z}{z-1}$. A simple induction gives $\mathbf{D}^n Y(z) = \frac{(-1)^n n!}{(z-1)^{n+1}}$. Formulas (3) and (4) give

$$\begin{aligned}\frac{(-1)^n n!}{(z-1)^{n+1}} &= \frac{(-1)^n}{z^n}\mathcal{Z}\{(k+n-1)^{\underline{n}}\} \\ &= \frac{(-1)^n}{z^n}\left(z^{n-1}\mathcal{Z}\{k^{\underline{n}}\}(z) - \sum_{m=0}^{n-2} m^{\underline{n}} z^{n-m-1}\right) \\ &= \frac{(-1)^n}{z}\mathcal{Z}\{k^{\underline{n}}\}(z).\end{aligned}$$

Solving for $\mathcal{Z}\{k^{\underline{n}}\}(z)$ gives the result.

□

Proposition 5. *Let $a \in \mathbb{C}$ and $n \in \mathbb{N}$. With $\varphi_{n,a}$ given in Definition 2 we have*

$$\mathcal{Z}\{\varphi_{n,a}(k)\}(z) = \frac{z}{(z-a)^{n+1}}.$$

Proof. First assume $a \neq 0$. Then $\varphi_{n,a}(k) = \frac{a^{k-n}k^{\underline{n}}}{n!} = \frac{a^{-n}}{n!}a^k k^{\underline{n}}$. We use formula (2) and (5) in the previous proposition to get

$$\begin{aligned}\mathcal{Z}\{\varphi_{n,a}(k)\}(z) &= \frac{a^{-n}}{n!}\mathcal{Z}\{a^k k^{\underline{n}}\}(z) \\ &= \frac{a^{-n}}{n!}\mathcal{Z}\{k^{\underline{n}}\}(z/a) \\ &= \frac{a^{-n}}{n!}\frac{n!(z/a)}{\left(\frac{z}{a}-1\right)^{n+1}} \\ &= \frac{z}{(z-a)^{n+1}}.\end{aligned}$$

Now suppose $a = 0$. Then $\varphi_{n,0}(k) = \delta_n(k)$. From the definition of the $\mathcal{Z}$-transform we get

$$\begin{aligned}\mathcal{Z}\{\varphi_{n,0}(k)\}(z) &= \sum_{k=0}^{\infty}\frac{\delta_n(k)}{z^k} \\ &= \frac{1}{z^n} = \frac{z}{(z-0)^{n+1}}.\end{aligned}$$

This completes the proof. □

3 The Matrix-Valued Sequence A^k and its $\mathcal{Z}$-Transform

Suppose $\mathbf{y}(k)$ is a sequence of $n \times n$ matrices over $\mathbb{C}$. We can extend the $\mathcal{Z}$-transform to $\mathbf{y}(k)$ by applying it to each entry. The translation principle of Proposition 4 extends to this matrix valued case; the proof is verbatim the same.

Let A be an $n \times n$ matrix. Our next proposition is a description of the $\mathcal{Z}$-transform of the sequence of matrices A^k.

Proposition 6. *Let A be an $n \times n$ matrix with entries in the complex numbers. Then*

$$\mathcal{Z}\left\{A^k\right\}(z) = z(zI - A)^{-1},$$

where I is the $n \times n$ identity matrix.

Proof. Let $\mathbf{y}(k) = A^k$ with the understanding that $\mathbf{y}(0) = I$. Then $\mathbf{y}(k+1) = A\mathbf{y}(k)$, for all $k = 0, 1, 2, \ldots$. Now apply the $\mathcal{Z}$-transform to both sides of this equation and use the translation principle to get

$$z\mathcal{Z}\left\{\mathbf{y}(k)\right\}(z) - zI = A\mathcal{Z}\left\{\mathbf{y}(k)\right\}(z).$$

Solving for $\mathcal{Z}\left\{\mathbf{y}(k)\right\}(z)$ gives the result. □

Example 7. *Consider the following example. Let* $A = \begin{bmatrix} 0 & 1 \\ -4 & 4 \end{bmatrix}$. *Then* $c_A(z) = \det(zI - A) = (z-2)^2$. *Proposition 6 and a calculation involving partial fractions gives*

$$\mathcal{Z}\left\{A^k\right\}(z) = z(zI - A)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{z}{z-2} + \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix} \frac{z}{(z-2)^2}.$$

The following corollary is immediate.

Corollary 8. *Let A be an $n \times n$ matrix with entries in the complex numbers. Then*

$$A^k = \mathcal{Z}^{-1}\left\{z(zI - A)^{-1}\right\}(k).$$

Proposition 9. *Let A be an $n \times n$ matrix with complex entries. Let $c_A(z) = \det(zI - A)$ be the characteristic polynomial. Assume $a_1, \ldots, a_R$ are the distinct roots with corresponding multiplicities $M_1, \ldots, M_R$. Then for each r, $1 \le r \le R$, and m, $0 \le m \le M_r - 1$, there are $n \times n$ matrices $B_{r,m}$ such that*

$$A^k = \sum_{r=1}^{R} \sum_{m=0}^{M_r - 1} B_{r,m} \varphi_{m,a_r}(k)$$

Proof. Our assumptions imply that we can factor c_A in the following way:

$$c_A(z) = \prod_{r=1}^{R} (z - a_r)^{M_r}.$$

The (i,j) entry of $(zI - A)^{-1}$ is of the form $\frac{p_{i,j}(z)}{c_A(z)}$ where $p_{i,j}(z)$ is some polynomial with degree less than n. Using partial fractions, we can write

$$\frac{p_{i,j}(z)}{c_A(z)} = \sum_{r=1}^{R} \sum_{m=0}^{M_r - 1} \frac{b_{r,m}(i,j)}{(z - a_r)^{m+1}}$$

where $b_{r,m}(i,j) \in \mathbb{C}$. It follows then that

$$z(zI-A)^{-1} = \sum_{r=1}^{R} \sum_{m=0}^{M_r-1} \frac{zB_{r,m}}{(z-a_r)^{m+1}}$$

where $B_{r,m}$ is the $n \times n$ matrix whose (i,j) entry is $b_{r,m}(i,j)$ for each pair (r,m). By Corollary 8 and Proposition 5 we get

$$A^k = \sum_{r=1}^{R} \sum_{m=0}^{M_r-1} B_{r,m}\varphi_{m,a_r}(k).$$

□

Example 10. *To illustrate Proposition 9 let* $A = \begin{bmatrix} 0 & 1 \\ -4 & 4 \end{bmatrix}$. *By Example 7 and Corollary 8 we have*

$$\begin{aligned} A^k &= \mathcal{Z}^{-1}\left\{z(zI-A)^{-1}\right\}(k) \\ &= \mathcal{Z}^{-1}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\frac{z}{z-2} + \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix}\frac{z}{(z-2)^2}\right\}(k) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\varphi_{0,2}(k) + \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix}\varphi_{1,2}(k) \end{aligned}$$

Lemma 11. *Suppose $p(z)$ is a polynomial and $a \in \mathbb{C}$ is a root with multiplicity M. Let m be a nonnegative integer with $m < M$. Then*

$$\mathbf{D}^m p(z)|_{z=a} = 0.$$

Proof. By assumption we can write $p(z) = (z-a)^M q(z)$ for some polynomial $q(z)$. Let $0 \le m < M$. By induction

$$\mathbf{D}^m p(z) = (z-a)^{M-m} q_1(z),$$

for some polynomial $q_1(z)$. Evaluating at $z = a$ gives $\mathbf{D}^m p(a) = 0$. □

4 Proof of the Cayley-Hamilton Theorem

Theorem 12 (Cayley-Hamilton)**.** *Let A be an $n \times n$ complex matrix and $c_A(z) = \det(zI - A)$, the characteristic polynomial. Then A satisfies its own characteristic polynomial. In other words,*

$$c_A(A) = 0.$$

Proof. By Lemma 2 and Proposition 9

$$\begin{aligned} A^k &= \sum_{r=1}^{R}\sum_{m=0}^{M_r-1} B_{r,m}\varphi_{m,a_r}(k) \\ &= \sum_{r=1}^{R}\sum_{m=0}^{M_r-1} B_{r,m}\frac{\mathbf{D}^m z^k}{m!}\bigg|_{z=a_r}. \end{aligned}$$

Now let $p(z)$ be any polynomial. Then by linearity of the derivative operator we have

$$p(A) = \sum_{r=1}^{R}\sum_{m=0}^{M_r-1} B_{r,m}\frac{\mathbf{D}^m p(z)}{m!}\bigg|_{z=a_r}.$$

However, for each root a_r with multiplicity M_r, Lemma 11 implies $\mathbf{D}^m c_A(z)|_{z=a_r} = 0$, for $0 \le m < M_r$. If we let $p = c_A$ then it follows that

$$c_A(A) = 0.$$

□

Example 13. *To illustrate the method of proof given in Theorem 12 let* $A = \begin{bmatrix} 0 & 1 \\ -4 & 4 \end{bmatrix}$. *By Example 10 and Lemma 2 we have*

$$A^k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{D}^0 z^k|_{z=2} + \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix}\mathbf{D}^1 z^k|_{z=2}.$$

The characteristic polynomial of A *is* $c_A(z) = (z-2)^2$. *We therefore have*

$$\begin{aligned} c_A(A) &= (A-2I)^2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{D}^0(z-2)^2|_{z=2} + \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix}\mathbf{D}^1(z-2)^2|_{z=2} \\ &= 0, \end{aligned}$$

since $\mathbf{D}^0(z-2)^2|_{z=2} = 0$ *and* $\mathbf{D}^1(z-2)^2|_{z=2} = 2(z-2)|_{z=2} = 0$.

References

[1] W.A. Adkins and M. G. Davidson, *The Cayley-Hamilton and Frobenius theorems via the Laplace Transform*, Linear Algebra and its Applications, **371** (2003), 147-152.

[2] A. Cayley, *A memoir on the theory of matrices*, Philosophical Transactions of the Royal Society of London, **148**, (1858), 17-37.

[3] G. Frobenius, *Ueber lineare Substitutionen und blineare Formen* J. Reine Angew. Math. **84**, (1878), 1-63.

[4] W. Hamilton Lectures on Quaternions Hodges and Smith, Dublin, 1853.

[5] K. M. Hoffman and R. A. Kunze: Linear Algebra, second edition, Prentice Hall, 1971.

[6] W. G. Kelley and A. C. Peterson, Difference Equations: An Introduction with Applications, Harcourt/Academic Press, San Diego, CA, second edition, 2001.

[7] S. Lange, *L*inear Algebra, second edition, Addison-Wesly, 1971.