

The Frame Algorithm and the Cayley-Hamilton Theorem

The following excerpt is from **Matrix Theory, From Generalized Inverses to Jordan Forms**, by Robert Piziak and P. L. Odell.

In 1949, **J. Sutherland Frame** (24 December 1907 - 27 February 1997) published an abstract in the *Bulletin of the American Mathematical Society* indicating a recursive algorithm for computing the inverse of a matrix and, as a by-product, getting additional information, including the famous Cayley-Hamilton theorem. (Hamilton is the Irish mathematician **William Rowan Hamilton** (4 August 1805 - 2 September 1895), and Cayley is **Aurthur Cayley** (16 August 1821 - 26 January 1895).) We have not been able to find an actual paper with a detailed account of these claims. Perhaps the author thought the abstract sufficient and went on with his work in group representations. Perhaps he was told this algorithm had been rediscovered many times (see [House, 1964, p. 72]). Whatever the case, in this section, we will expand on and expose the details of Frame's algorithm. Suppose $A \in \mathbf{C}^{n \times n}$. The *characteristic matrix* of A is $xI - A \in \mathbf{C}[x]^{n \times n}$, the collection of n -by- n matrices with polynomial entries. We must open our minds to accepting matrices with polynomial entries. For example, $\begin{bmatrix} x^2 + 1 & x - 3 \\ 4x + 2 & x^3 - 7 \end{bmatrix} \in \mathbf{C}[x]^{2 \times 2}$. Determinants work just fine for these kinds of matrices. The determinant of $xI - A$, $\det(xI - A) \in \mathbf{C}[x]$, the polynomials in x , and is what we call the *characteristic polynomial* of A:

$$\mathcal{X}_A(x) = \det(xI - A) = x^n + c_1x^{n-1} + \cdots + c_{n-1}x + c_n$$

For example, if $A = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} \in \mathbf{C}[x]^{3 \times 3}$, then $xI_3 - A =$

$$\begin{bmatrix} x-1 & -2 & -2 \\ -3 & x-4 & -5 \\ -6 & -7 & x-8 \end{bmatrix} \in \mathbf{C}[x]^{3 \times 3}.$$

Thus $\mathcal{X}_A(x) = \begin{vmatrix} x-1 & -2 & -2 \\ -3 & x-4 & -5 \\ -6 & -7 & x-8 \end{vmatrix} = x^3 - 13x^2 - 9x - 3$. This is computed

using the usual familiar rules for expanding a determinant.

You may recall that the roots of the characteristic polynomial are quite important, being the *eigenvalues* of the matrix. We will turn to this topic later. For now, we focus on the coefficients of the characteristic polynomial.

First, we consider the constant term c_n . You may already know the answer here, but let's make an argument. Now $\det(A) = (-1)^n \det(-1) = (-1)^n \det(0I - A) = (-1)^n (X)_A(0) = (-1)^n c_n$. Therefore,

$$\det(A) = (-1)^n c_n.$$

As a consequence, we see immediately that A is invertible iff $c_n \neq 0$, in which case

$$A^{-1} = \frac{(-1)^n}{c_n} \text{adj}(A),$$

where $\text{adj}(A)$ is the adjugate matrix of A introduced previously. Also recall the important relationship, $B \text{adj}(B) = \det(B) I$. We conclude that

$$(xI - A) \text{adj}(xI - A) = \chi_A(x) I$$

To illustrate with the example above,

$$\begin{aligned} (xI - A) &= \begin{bmatrix} x-1 & -2 & -2 \\ -3 & x-4 & -5 \\ -6 & -7 & x-8 \end{bmatrix} \cdot \begin{bmatrix} x^2-12x-3 & 2x-2 & 2x+2 \\ 3x+6 & x^2-9x-4 & 5x+1 \\ 6x-3 & 7x+5 & x^2-5x-2 \end{bmatrix} \\ &= \begin{bmatrix} x^3-13x^2-9x-3 & 0 & 0 \\ 0 & x^3-13x^2-9x-3 & 0 \\ 0 & 0 & x^3-13x^2-9x-3 \end{bmatrix} \\ &= x^3-13x^2-9x-3 \cdots \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Next, let $C(x) = \text{adj}(xI - A) \in \mathbf{C}[x]^{n \times n}$. We note that the elements of $\text{adj}(xI - A)$ are computed as $(n-1)$ -by- $(n-1)$ subdeterminants of $(xI - A)$, so the highest power that can occur in $C(x)$ is x^{n-1} . Also, note that we can identify $\mathbf{C}[x]^{n \times n}$, the n -by- n matrices with polynomial entries with $\mathbf{C}^{n \times n}[x]$, the polynomials with matrix coefficients, so we can view $C(x)$ as a polynomial in x with scalar matrices as coefficients. For example,

$$\begin{bmatrix} x^2+1 & x-3 \\ 4x+2 & x^3-7 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} x^3 + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x^2 + \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} x + \begin{bmatrix} 1 & -3 \\ 2 & -7 \end{bmatrix} x.$$

All you do is gather the coefficients of each power of x and make a matrix of scalar coefficients of that power of x . Note that what we have thusly created is an element of $\mathbf{C}[x]^{n \times n}$, the polynomials in x whose coefficients come from the n -by- n matrices over \mathbf{C} . Also note, $xB = Bx$ for all $B \in \mathbf{C}[x]^{n \times n}$, so it does not matter which side we put the x on. We now view $C(x)$ as such an expression in $\mathbf{C}^{n \times n}[x]$:

$$C(x) = B_0 x^{n-1} + B_1 x^{n-2} + \cdots + B_{n-2} x + B_{n-1}.$$

These coefficient matrices turn out to be of interest. For example, $\text{adj}(A) = (-1)^{n-1} \text{adj}(-A) = (-1)^{n-1} C(0) = (-1)^{n-1} B_{n-1}$, so

$$\text{adj}(A) = (-1)^{n-1} B_{n-1}.$$

Thus, if $c_n \neq 0$, A is invertible and we have

$$A^{-1} = \frac{-1}{c_n} B_{n-1}.$$

But now we compute

$$\begin{aligned} (xI - A) \text{adj}(xI - A) &= (xI - A) C(x) \\ &= (xI - A) (B_0 x^{n-1} + B_1 x^{n-2} + \cdots + B_{n-2} x + B_{n-1}) \\ &= x^n I + x^{n-1} c_1 I + \cdots + x c_{n-1} I + c_n I \end{aligned}$$

and we compare coefficients using the following table:

Compare Coefficients	Multiply by	on the Left	on the Right
$B_0 = I$	A^n	$A^n B_0$	A^n
$B_1 - AB_0 = C_1 I$	A^{n-1}	$A^{n-1} B_1 - A^n B_0$	$c_1 A^{n-1}$
$B_2 - AB_1 = C_2 I$	A^{n-2}	$A^{n-2} B_2 - A^{n-1} B_1$	$c_2 A^{n-2}$
\vdots			
$B_k - AB_{k-1} = c_k I$			
\vdots			
$B_{n-2} - AB_{n-3} = C_{n-2} I$	A^2	$A^2 B_{n-2} - A^3 B_{n-3}$	$c_{n-2} A^2$
$B_{n-1} - AB_{n-2} = C_{n-1} I$	A	$AB_{n-1} - A^2 B_{n-2}$	$c_{n-1} A$
$-AB_{n-1} = C_n I$		$-AB_{n-1}$	$c_n I$
column	sum =	$\mathbf{0} =$	$\mathcal{X}_A(A)$

So, the first consequence we get from these observations is that the Cayley-Hamilton theorem just falls out as an easy consequence. (Actually, Liebler [2003] reports that Cayley and Hamilton only established the result for matrices up to size 4-by-4. He says it was Frobenius (**Ferdinand Georg Frobenius** [26 October 1849 - 3 August 1917] who gave the first complete proof in 1878.)

THEOREM 2.18 (Cayley-Hamilton theorem)

For any n -by- n matrix A over \mathbf{C} , $\mathcal{X}_A(A) = \mathbf{O}$.

What we are doing in the Cayley-Hamilton theorem is plugging a matrix into a polynomial. Plugging numbers into a polynomial seems reasonable, almost inevitable, but matrices? Given a polynomial

$$p(A) = 4I + 3A - 9A^3 = 4 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 3 \begin{bmatrix} 1 & 2 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} - 9 \begin{bmatrix} 1 & 2 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}^3 = \begin{bmatrix} -2324 & -2964 & -3432 \\ -5499 & -7004 & -8112 \\ -9243 & -11778 & -13634 \end{bmatrix},$$

The Cayley-Hamilton theorem says that any square matrix is a “root” of its characteristic polynomial.

But there is much more information packed in those equations on the left of the table, so let’s push a little harder. Notice we can rewrite these equations as

$$\begin{aligned} B_0 &= I \\ B_1 &= AB_0 + c_1 I \\ B_2 &= AB_1 + c_2 I \\ &\vdots \\ B_{n-1} &= AB_{n-2} + c_{n-1} I \\ \mathbf{O} &= AB_{n-1} + c_n I. \end{aligned}$$

By setting $B_n := \mathbf{O}$, we have the following recursive scheme clear from above: for $k = 1, 2, \dots, n$,

$$\begin{aligned} B_0 &= I \\ B_k &= AB_{k-1} + c_k I. \end{aligned}$$

In other words, the matrix coefficients, the B_k s are given recursively in terms of the B_{k-1} s and c_k s. If we can get a formula for the B_k and c_k in terms of B_{k-1} , we will get a complete set of recurrence formulas for the B_k and c_k . In particular, if we know B_{n-1} and c_n , we have A^{-1} , provided, of course, A^{-1} exists (i.e. provided $c_n \neq 0$). For this, let’s exploit the recursion given below:

$$\begin{aligned}
 B_0 &= I \\
 B_1 &= AB_0 + c_1 I = AI + c_1 I = A + c_1 I \\
 B_2 &= AB_1 + c_2 I = A(A + c_1 I) + c_2 I = A^2 + c_1 A + c_2 I \\
 B_3 &= A^3 + c_1 A^2 + c_2 A + c_3 I \\
 &\vdots
 \end{aligned}$$

Inductively, we see for $k = 1, 2, \dots, n$,

$$B_k = A^k + c_1 A^{k-1} + \dots + c_{k-1} A + c_k I.$$

Indeed, when $k = n$, this is just the Cayley-Hamilton theorem all over again. Now we have for $k = 2, 3, \dots, n+1$,

$$B_{k-1} = A^{k-1} + c_1 A^{k-2} + \dots + c_{k-2} A + c_{k-1} I.$$

If we multiply through by A , we get for $k = 2, 3, \dots, n+1$,

$$AB_{k-1} = A^k + c_1 A^{k-1} + \dots + c_{k-2} A^2 + c_{k-1} A.$$

Now we pull a trick out of the mathematician's hat. Take the trace of both sides of the equation using linearity of the trace functional.

$$\text{tr}(AB_{k-1}) = \text{tr}(A^k) + \text{tr}(c_1 A^{k-1}) + \dots + \text{tr}(c_{k-2} A^2) + \text{tr}(c_{k-1} A).$$

$$\text{for } k = 2, 3, \dots, n+1.$$

Why would anybody think to do such a thing? Well, the appearance of the coefficients of the characteristic polynomial on the right is very suggestive. Those who know a little matrix theory realize that the trace of A^r is the sum of the r^{th} powers of the roots of the characteristic polynomial and so Newton's identities leap to mind. Let s_r denote the sum of the r^{th} powers of roots of the characteristic polynomial. Thus, for $k = 2, 3, \dots, n+1$,

$$\text{tr}(AB_{k-1}) = s_k + c_1 s_{k-1} + \dots + c_{k-2} s_2 + c_{k-1} s_1.$$

Digression on Newton's Identities

Newton's identities go back aways. They relate the sums of powers of the roots of a polynomial recursively to the coefficients of the polynomial. Many proofs are available. Some involve the algebra of symmetric functions, but we do not want to take the time to go there. Instead, we will use a calculus-baed argument following the ideas of [Eidswick 1968]. First, we need to recall some facts about polynomials. Let $p(x) = a_0 + a_1x + \cdots + a_nx^n$. Then the coefficients of p can be expressed in terms of derivatives of p evaluated at zero (remember Taylor polynomials?):

$$p(x) = p(0) + p'(0)x + \frac{p''(0)}{2!}x^2 + \cdots + \frac{p^{(n)}(0)}{n!}x^n.$$

Now here is something really slick. Let's illustrate a general fact. Suppose $p(x) = (x-1)(x-2)(x-3) = -6 + 11x - 6x^2 + x^3$. Do a wild and crazy thing. Reverse the rolls [*sic roles*] of the coefficients and form the new reversed polynomial $q(x) = -6x^3 + 11x^2 - 6x + 1$. Clearly $q(1) = 0$ but, more amazingly, $q(\frac{1}{2}) = -\frac{6}{8} + \frac{11}{4} - \frac{6}{2} + 1 = \frac{-6+22-24+8}{8} = 0$. You can also check $q(\frac{1}{3}) = 0$. So the reversed polynomial has as roots the reciprocals of the roots of the original polynomial. Of course, the roots are not zero for this to work. This fact is generally true. Suppose $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and the reversed polynomial is $q(x) = a_n + a_{n-1}x + \cdots + a_0x^n$. Note

$$q(0) = a_n, q'(0) = a_{n-1}, \dots, \frac{q^{(n)}(0)}{n!} = a_0.$$

Then $r \neq 0$ is a root of p iff $\frac{1}{r}$ is a root of q .

Suppose $p(x) = a_0 + a_1x + \cdots + a_nx^n = a_n(x-r_1)(x-r_2)\cdots(x-r_n)$. The r_i s are, of course, the roots of p , which we assume to be nonzero but not necessarily distinct. Then the reversed polynomial $q(x) = a_n + a_{n-1}x + \cdots + a_0x^n = a_0\left(x - \frac{1}{r_1}\right)\left(x - \frac{1}{r_2}\right)\cdots\left(x - \frac{1}{r_n}\right)$. For the sake of illustration, suppose $n = 3$.

Then form

$$\begin{aligned} f(x) &= \frac{q'(x)}{q(x)} = \frac{(x-r_1^{-1})[(x-r_2^{-1})+(x-r_3^{-1})][(x-r_2^{-1})+(x-r_3^{-1})]}{(r-r_1)(r-r_2)(r-r_3)} \\ &= \frac{1}{x-r_1^{-1}} + \frac{1}{x-r_2^{-1}} + \frac{1}{x-r_3^{-1}} \end{aligned}$$

. Generally then,

$$f(x) = \sum_{k=1}^n \frac{1}{(x-r_k^{-1})}.$$

Let's introduce more notation. Let $s_m = \sum_{k=1}^m$ for $m = 1, 2, 3, \dots$. Thus, s_m is the sum of the m^{th} powers of roots of p . The derivatives of f are intimately related to the s s. Basi differentiation yields

$$\begin{aligned} f(0) &= -s_1 \\ f'(x) &= \sum_{k=1}^n \frac{-1}{(x-r_k^{-1})^2} & f'(0) &= -s_2 \\ f''(x) &= \sum_{k=1}^n \frac{-2}{(x-r_k^{-1})^3} & f''(0) &= -2s_3 \\ &\vdots \\ f^{(k)}(x) &= \sum_{k=1}^n \frac{-k!}{(x-r_k^{-1})^{k+1}} & f^{(k)}(0) &= -k!s_{k+1} \end{aligned}$$

The last piece of the puzzle is the rule of taking the derivative of a product; this is the so-called *Leibnitz rule* for differentiating a product:

$$D^n(F(x)G(x)) = \sum_{j=0}^n \binom{n}{j} F^{(j)} G^{(n-j)}(x).$$

All right, let's do the argument. We have $f(x) = \frac{q'(x)}{q(x)}$, so $q'(x) = f(x)q(x)$.

Therefore, using Leibnitz rule

$$q^{(m)}(x) = [f(x)q(x)]^{(m-1)} = \sum_{k=0}^{m-1} \binom{m-1}{k} f^{(k)}(x) q^{(m-1-k)}(x).$$

Plugging in zero, we get

$$\begin{aligned} q^{(m)}(0) &= \sum_{k=0}^{m-1} \binom{m-1}{k} f^{(k)}(0) q^{(m-1-k)}(0) \\ &= \sum_{k=0}^{m-1} \frac{(m-1)!}{k! (m-1-k)!} (-k!) s_{k+1} q^{(m-1-k)}(0). \end{aligned}$$

Therefore,

$$\frac{q^{(m)}(0)}{m!} = -\frac{1}{m} \sum_{k=0}^{m-1} \frac{q^{(m-1-k)}(0)}{(m-1-k)!} s_{k+1}.$$

One more substitution and we have the *Newton identities*

$$\begin{aligned} 0 = ma_{n-m} &= \sum_{k=0}^{m-1} a_{n-m+k+1} s_{k+1} \text{ if } 1 \leq m \leq n \\ 0 &= \sum_{k=m-n-1}^{m-1} a_{n-m+k+1} s_{k+1} \text{ if } m > n. \end{aligned}$$