MA322
Sp12

The Cayley-Hamilton Theorem **Use of similarity**

We give an outline of one of the most striking theorems of Linear Algebra.

1. Recall that a square matrix G is said to be similar to H if $G = PHP^{-1}$ for some invertible P. In symbols, we write $G \approx H$.

   By taking powers of the similarity equation, we see that for any positive integer m, we have:

   $$G^m = PHP^{-1}PHP^{-1} \cdots PHP^{-1} \text{ where we have written } m \text{ terms}$$

   and by multiplying out, we see that $G^m = PH^mP^{-1}$.

2. Now, if $f(X) = a_0 + a_1X + \cdots + a_nX^n$ is any polynomial and A is any square matrix, then we **define**

   $$f(A) = a_0I + a_1A + \cdots + a_nA^n.$$

   Thus, it makes sense to define A to be a **root of the polynomial f(X) if** $f(A) = 0$.

3. The Cayley Hamilton Theorem states that:

   If A is a square matrix and if $f(X)$ is its characteristic polynomial, then $f(A) = 0$. In other words, every square matrix is a root of its characteristic polynomial!!

4. **Idea of the proof.**

   The theorem is easily seen to be true for a diagonal matrix $diag(a_1, a_2, \cdots, a_n)$. We note that the characteristic polynomial is simply $f(X) = (a_1 - X)(a_2 - X) \cdots (a_n - X)$. Then we see that

   $$\begin{aligned} f(A) &= (a_1I - A)(a_2I - A) \cdots (a_nI - A) \\ &= diag(0, a_2, \cdots, a_n)diag(a_1, 0, a_3, \cdots, a_n) \cdots diag(a_1, a_2, \cdots, a_{n-1}, 0) \\ &= diag(0, 0, \cdots, 0) = 0 \end{aligned}$$

Next we observe that the theorem must be true for any matrix **similar to a diagonal matrix, i.e. for a diagonalizable matrix.**. To see this, note that If $G = PAP^{-1}$, then

$$(G - \lambda I) = (PAP^{-1} - \lambda I) = P(A - \lambda I)P^{-1}$$

and so taking determinants, G and A have the same characteristic polynomials, say $f(\lambda)$.

Then

$$f(G) = f(PAP^{-1}) = Pf(A)p^{-1} = P0P^{-1} = 0.$$

Thus the theorem is proved for all diagonalizable matrices.

Now we finish the proof for any square matrix. However, this depends on some general ideas which, at first, may seem too abstract. They will become clear after some thought.

- In this part, we may find it easier to replace λ by a variable X, so that the results seem more natural. These ideas are standard in a course in theory of equations.

  Suppose a square matrix A has characteristic polynomial $f(X) = \det(A - XI)$ and suppose that $f(X)$ and its derivative $f'(X)$ have no common factors. Then all roots of $f(X)$ are simple (i.e. of multiplicity 1).

  Moreover, if we replace our ground field by its algebraic closure, then $f(X)$ would have exactly n distinct roots, if A is $n \times n$.

- Thus, we may assume that our $f(X)$ has n distinct linear factors $X - a_1, X - a_2, \cdots X - a_n$, so that A has n distinct eigenvalues $a_1, a_2, \cdots, a_n$.

- Then we know that A is diagonalizable and hence $f(A) = 0$ as required.

- Now assume that A is a matrix with n^2 different variable entries $\{y_{ij}\}$ so that

$$A = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{nn} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix}.$$

  We shall prove that the theorem is true for such a matrix A and then it will follow for all special $n \times n$ matrices by specializing the variables to desired special values.

- We only have to prove that the characteristic polynomial $f(X) = \det(A - XI)$ has only simple roots. But the condition for this is that $GCD(f(X), f'(X))$ is not identically zero.

  But we note that this GCD is not zero when we specialize A to a diagonal matrix with distinct entries, so the GCD cannot be identically zero as a polynomial in $\{y_{ij}\}$. This proves our claim.