UChicago REU 2013 Apprentice Program **Spring 2013**

Lecture 18: July 25th, 2013

Prof. Babai Scribe: David Kim

1 Cayley-Hamilton Theorem

Exercise 1.1 *Let* $f(t) = \frac{at^2+bt+c}{dt^2+e}$, *where* $e \neq 0$. *Prove: If* $(\forall t)(f(0) \geq f(t))$, *then* $b = 0$.

Hint: calculus. Evaluate $f'(0)$.

Exercise 1.2 *Let* $f(t) = a_0 + a_1 t + \cdots + a_k t^k$ *and* $D \in M_n(\mathbb{C})$ *be diagonal, s.t.* $D_{ii} = \lambda_i$. *What is* $f(D) = a_0 I + a_1 D + \cdots + a_k D^k$*?*

Let's try a few examples first.

Example 1.3 *Let* $f(t) = t^2 + 1 \Rightarrow f(D) = D^2 + I = \begin{pmatrix} \lambda_1^2 + 1 & & \\ & \ddots & \\ & & \lambda_n^2 + 1 \end{pmatrix}$.

Example 1.4 *If* D_1, D_2 *are diagonal s.t.* $(D_1)_{ii} = \lambda_i, (D_2)_{ii} = \mu_i$

$$\Rightarrow D_1 D_2 = \begin{pmatrix} \lambda_1\mu_1 & & \\ & \ddots & \\ & & \lambda_n\mu_n \end{pmatrix}, D_1 + D_2 = \begin{pmatrix} \lambda_1 + \mu_1 & & \\ & \ddots & \\ & & \lambda_n + \mu_n \end{pmatrix}$$

For exercise 1.2, it is easy to see from the examples that if D is diagonal, then $f(D)$ is also a diagonal matrix:

$$f(D) = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}$$

Theorem 1.5 (Cayley-Hamilton Theorem) *Let* $A \in M_n(\mathbb{C})$ *and* $f_A(t) = \det(tI - A)$. *Then* $f_A(A) = 0$.

Sanity Check: If $f_A(t) = \det(tI - A)$, then isn't $f_A(A) = \det(A - A) = 0$ obvious? What's the big deal? $\Rightarrow$ The statement $f_A(A) = \det(A - A)$ does not make any sense. A polynomial function of a matrix is also a matrix: $f_A(A) = \mathbf{0}$ means the $\mathbf{0}$ matrix, while $\det(A - A) = 0$ is just a number.

We are not evaluating the value of the determinant for a single matrix, but looking at an *identity* relation which holds for every n^2 entry of $f_A(A)$, that they all evaluate to 0 - which is very surprising.

Let's try an example for a 2×2 matrix.

Example 1.6 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then*

$$f_A(t) = \det \begin{pmatrix} t-a & -b \\ -c & t-d \end{pmatrix} = (t-a)(t-d) - bc = t^2 - (a+d)t + ad - bc$$

and $f_A(A) = A^2 - (a+d)A + (ad-bc)I$. We have four identities in four variables, a, b, c, d. We just check one of them, for the top left entry of $f_A(A)$. $(A^2)_{11} = a^2 + bc$, so $(f_A(A))_{11} = (a^2 + bc) - (a+d)a + (ad - bc) = 0$. Check that all other entries also evaluate to 0.

We will prove the Cayley-Hamilton Theorem in the following steps:

Cayley-Hamilton Proof Outline

1. C-H true for diagonal matrices.
2. C-H true for diagonalizable matrices.
3. Diagonalizable matrices are dense $\Rightarrow$ C-H true for all $A \in M_n(\mathbb{C})$.

1.1 Cayley-Hamilton for Diagonal Matrices

Proof: [C-H for Diagonal Matrices] Let $A \in M_n(\mathbb{C})$ be diagonal s.t. $A_{ii} = \lambda_i$. Then

$$f_A(t) = \det(tI - A) = \det \begin{pmatrix} t-\lambda_1 & & \\ & \ddots & \\ & & t-\lambda_n \end{pmatrix} = \prod_{i=1}^{n} (t - \lambda_i)$$

and $f_A(A) = \prod_{i=1}^{n} (A - \lambda_i I)$, a product of diagonal matrices. As in the previous examples, since A is diagonal, we have

$$f_A(A) = \begin{pmatrix} f_A(\lambda_1) & & \\ & \ddots & \\ & & f_A(\lambda_n) \end{pmatrix} = \begin{pmatrix} \prod_{i=1}^{n}(\lambda_1 - \lambda_i) & & \\ & \ddots & \\ & & \prod_{i=1}^{n}(\lambda_n - \lambda_i) \end{pmatrix} = \mathbf{0}$$

■

1.2 Cayley-Hamilton for Diagonalizable Matrices

Recall A is similar to B, or $A \sim B$, if $\exists C$ such that $B = C^{-1}AC$ (B is the conjugate of A by C).

Exercise 1.7 *Let* $E, F \in M_n(\mathbb{C}) \Rightarrow \det(EF) = \det(E) \cdot \det(F)$.

Proof. Perform Guassian elimination: do elementary column operations on F. Such an operation will change EF by exactly the same column operation. So the determinant stays the same on each side after each column operation, and eventually, F will reach I - giving us the desired equality.

Exercise 1.8 $\det(C^{-1}) = \dfrac{1}{\det(C)}$.

Exercise 1.9 *If* $A \sim B \Rightarrow \det(A) = \det(B)$.

Proof. If $A = C^{-1}BC$, $\det(A) = \det(C^{-1}BC) = \det(C^{-1})\det(B)\det(C) = \det(C^{-1})\det(C)\det(B) = \det(B)$ (multiplication of numbers is commutative). □

Exercise 1.10 *If* $A \sim B \Rightarrow (tI - A) \sim (tI - B)$.

Proof. $C^{-1}(tI - A)C = tC^{-1}IC - C^{-1}AC = tI - B$. □

Lemma 1.11 *If* $A \sim B \Rightarrow f_A(t) = f_B(t)$.

Proof. $A \sim B \Rightarrow (tI - A) \sim (tI - B) \Rightarrow \det(tI - A) = \det(tI - B) \Rightarrow f_A(t) = f_B(t)$. □

The advantage of using $f_A(t) = \det(tI - A)$ to using $\det(A - tI)$
$\Rightarrow$ The leading coefficient is 1, whereas for $\det(A - tI)$, the leading coefficient is $(-1)^n$. Note: $\det(tI - A) = t^n - \mathrm{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A)$,

Lemma 1.12 *If* g *is a polynomial and* $A \sim B$ $(B = C^{-1}AC)$*, then* $g(A) \sim g(B)$*, in particular,* $g(B) = C^{-1}f(A)C$.

Proof. Let's try for $g(t) = t^k$. Then

$$g(B) = B^k = (C^{-1}AC)^k = (C^{-1}AC)(C^{-1}AC)\ldots(C^{-1}AC) = C^{-1}A^kC = C^{-1}g(A)C$$

So given $g(t) = \alpha_k t^k + \cdots + \alpha_1 t + \alpha_0$, each $\alpha_i B^i = \alpha_i C^{-1}A^iC$. Use distributivity of matrix multiplication over addition to pull out C^{-1} and C to get $g(B) = C^{-1}g(A)C$. □

Proof: [Cayley-Hamilton for Diagonalizable Matrices]

Let $A \sim D$, where D is diagonal. Then $f_D(D) = 0$, as C-H holds for diagonal matrices. By lemma 1.11, $f_A(t) = f_D(t)$, so $f_A(D) = f_D(D) = 0$. By lemma 1.12, $f_A(A) \sim f_A(D)$, and $f_A(A) = C^{-1}f_A(D)C = C^{-1}0C = 0$. ■

1.3 Diagonalizable Matrices Dense, Cayley-Hamilton for All Matrices

Exercise 1.13 *Can you give an example of a matrix that is not diagonalizable?*
example: $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Why is the above example not diagonalizable? Suppose $A \sim D$. Then $f_A(t) = (t-1)^2 = f_D(t)$, meaning $D = I$. But $A = C^{-1}DC = C^{-1}IC = I$. (contradiction) (This shows that the only matrix similar to I is I)

Exercise 1.14 $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$?

Hint: distinct eigenvalues (see below)

Exercise 1.15 *$A \in M_n(\mathbb{C})$ is diagonalizable $\Leftrightarrow$ A has an eigenbasis, i. e., $\mathbb{C}^n$ has a basis consisting of eigenvectors of A.*

Exercise 1.16 *$C^{-1}AC = D \Leftrightarrow$columns of C form an eigenbasis of A.*

Exercise 1.17 *If $v_1, \ldots, v_k$ are eigenvectors of the matrix $A \in M_n(\mathbb{C})$ to distinct eigenvalues $\Rightarrow$ $v_1, \ldots, v_k$ are linearly independent.*

Corollary 1.18 *If f_A for $A \in M_n(\mathbb{C})$ has no multiple roots, i.e., all n eigenvalues of A are distinct $\Rightarrow$ A is diagonalizable. ("corollary" - a statement that follows immediately)*

Now we want to prove the Cayley-Hamilton Theorem for all matrices. Let's first try an example: a non-diagonalizable triangular matrix.

Example 1.19 $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. *We showed this is not diagonalizable. Find a sequence of diagonalizable matrices that converges to it.*

Proof. Consider $\{B_m\}$, where $B_m = \begin{pmatrix} 1 & 1 \\ 0 & 1 + \frac{1}{m} \end{pmatrix}$. Then by the previous corollary, each B_m is diagonalizable, hence $f_{B_m}(B_m) = 0$. Since $B_m \longrightarrow A$, $f_{B_m} \longrightarrow f_A$, and the determinant is continuous, $f_A(A) = \lim_{m\to\infty} f_{B_m}(B_m) = 0$. $\square$

If we could show that diagonalizable matrices are dense, i.e., every neighborhood of $A \in M_n(\mathbb{C})$ contains a diagonalizable matrix, then we can make a similar argument as the above exercise to show that $f_A(A) = 0$.

Example 1.20 *Find a diagonalizable matrix where each matrix entry is within ε of the corresponding entry of A.*

$$A = \begin{pmatrix} 5 & * & * & * & * \\ & 5 & * & * & * \\ & & \pi & * & * \\ & & & \pi & * \\ & & & & \pi \end{pmatrix}$$

Solution. make small changes to the diagonal entries such that all the diagonal entries become distinct. Example:

$$\begin{pmatrix} 5 & * & * & * & * \\ & 5+\frac{\varepsilon}{2} & * & * & * \\ & & \pi & * & * \\ & & & \pi+\frac{\varepsilon}{2} & * \\ & & & & \pi-\frac{\varepsilon}{2} \end{pmatrix}$$

So how can we extend this idea to all matrices to show that diagonaliazable matrices are dense? Use Schur's Theorem - $\forall A, \exists T$, upper triangular, $\exists U$, unitary s.t. $A = U^*TU$.

Lemma 1.21 *Diagonalizable matrices are dense in $M_n(\mathbb{C})$.*

Proof. Given $A \in M_n(\mathbb{C})$, by Schur's Theorem, $\exists T, U$ s.t. U is unitary ($U^* = U^{-1}$) and $A = U^*TU$, so $A = U^{-1}TU$. Then $\forall \varepsilon > 0$, $\exists T'$, diagonalizable, obtainable by changing some of the diagonal entries of T by less than ε so that they all become distinct. Then as $\varepsilon \longrightarrow 0$, $U^*T'U \longrightarrow U^*TU = A$. □

Theorem 1.22 (Cayley-Hamilton Theorem) *If $A \in M_n(\mathbb{C})$, then $f_A(A) = 0$.*

Proof: It follows from Lemma 1.21 the same way as we solved Example 1.19. ■

2 The Minimal Polynomial

Exercise 2.1 *Prove without using Cayley-Hamilton: $(\forall A \in M_n(\mathbb{C}))(\exists f \in \mathbb{C}[x]$, a polynomial s.t. $f \neq 0$ but $f(A) = 0)$.*

What does it mean that A is a root of a polynomial? It is equivalent to saying $\exists$ a non-trivial linear combination of the powers of A that evaluates to 0, i.e., the powers of A involved are linearly dependent. Question: Are the A^k vectors? What is the vector space?

Exercise 2.2 *$M_n(\mathbb{C})$ is a vector space of dimension n^2. What is a basis?*

So $\{I, A, \ldots, A^{n^2}\}$ are $n^2 + 1$ vectors in an n^2-dimensional vector space $\Rightarrow \exists$ non-trivial linear combination of them evaluating to $0 \Rightarrow \exists$ a polynomial of degree n^2 of which A is a root.

Definition 2.3 (minimal polynomial) *The smallest degree polynomial, $m_A(t) \neq 0$, s.t. $m_A(A) = 0$ is called the minimal polynomial of A.*

Example 2.4 *What is the minimal polynomial of the diagonal matrix* $D = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 2 & \\ & & & & 2 \end{pmatrix}$?

Answer: $m_D(t) = (t-1)(t-2)$

The above example shows that if A is diagonal, then $m_A(t) = \prod(t - \lambda_i)$, for all distinct values of λ_i (no multiple roots).

Exercise 2.5 *A is diagonalizable $\Leftrightarrow$ $m_A(t)$ has no multiple roots.*

Example 2.6 *What is the minimal polynomial of* $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$? *Note that $t-1$ doesn't work.*

Exercise 2.7 $(\forall g \in \mathbb{C}[t])(g(A) = 0 \Leftrightarrow m_A | g$, *i.e.* $(\exists h)(g = h \cdot m_A))$.

Corollary 2.8 (Cayley-Hamilton) $\forall A \in M_n(\mathbb{C})$, $m_A | f_A$.

That is, $f_A(A) = 0$. Going back to example 2.6, $f_A(t) = (t-1)^2$. Since $m_A | f_A$, $m_A(t) = (t-1)$ or $(t-1)^2$. We already know $t-1$ does not work. Therefore, it must be that $m_A(t) = (t-1)^2$.

Given a polynomial, how can one determine if it has multiple roots? Given two polynomials with integer coefficients, how can one decide if any two have a common complex root using rational operations?

Exercise 2.9 *Let $f(x) = x^3 - 1$, $g(x) = x^5 + x^4 + x^3 + x^2 + x + 1$. What are the common complex roots of f, g?*

Solution. Factor the polynomials to get $f(x) = (x-1)(x^2+x+1)$, $g(x) = (x^3+1)(x^2+x+1)$. Clearly, the complex roots of $x^2 + x + 1$ are the common complex roots of f and g. Note that this is analogous to finding the gcd of two integers.

Recall the Division Theorem: $(\forall a, b, b \neq 0)(\exists q, r$ s.t. $a = bq + r$, $0 \leq r < |b|)$. So if $e|a, e|b \Rightarrow e|r$. Euclid's Algorithm works by recursively using the division theorem, $\gcd(a, b) = \gcd(b, r)$ where $a = bq + r$, until one reaches 0. The last positive term is the gcd. One can do this for polynomials as well (the reminader has smaller degree than the divisor), and the coefficients would alwasy stay in $\mathbb{Q}$.

Exercise 2.10 *Let $f \in \mathbb{C}[x]$. f has multiple roots $\Leftrightarrow$ $\gcd(f, f')$ is not constant.*

3 Linear Transformations in Euclidean Space

Recall that a Euclidean space is a vector space over $\mathbb{R}$ with a positive definite inner product. Let V be a Euclidean space, and let $\varphi : V \to V$ be a linear transformation from V to itself.

Definition 3.1 *φ is an orthogonal transformation if* $(\forall x, y \in V)(\langle x, y\rangle = \langle \varphi(x), \varphi(y)\rangle)$.

Definition 3.2 *φ is a symmetric transformation if* $(\forall x, y \in V)(\langle \varphi(x), y\rangle = \langle x, \varphi(y)\rangle)$.

Given any linear transformation, we can associate a matrix to it, given that we have a fixed basis. If $\underline{\mathrm{e}} = \{e_1, \ldots, e_n\}$ and $\underline{\mathrm{f}} = \{f_1, \ldots, f_n\}$ are two bases, what is the relationship between $[\varphi]_{\underline{\mathrm{e}}}$ and $[\varphi]_{\underline{\mathrm{f}}}$? $C^{-1}[\varphi]_{\underline{\mathrm{e}}}C = [\varphi]_{\underline{\mathrm{f}}}$, where $C = [\underline{\mathrm{f}}]_{\underline{\mathrm{e}}}$ (change of basis). This is the motivation behind similarity - two similar matrices denote the same linear transformation under different bases.

Exercise 3.3 *Let $\underline{e}$ be an orthonormal basis of the Euclidean space V. Then*

(i) φ is an orthogonal transformation $\Leftrightarrow$ $[\varphi]_{\underline{e}}$ is an orthogonal matrix.

(ii) φ is a symmetric transformation $\Leftrightarrow$ $[\varphi]_{\underline{e}}$ is a symmetric matrix.

Equivalent statement of the Spectral Theorem:

If φ is a symmetric linear transformation in a finite dimensional Euclidean space then φ has an orthonormal eigenbasis.

4 Invariant Subspaces

Definition 4.1 *Let $\varphi : V \to V$. Then $U \subseteq V$ is φ-invariant if $\forall u \in U \Rightarrow \varphi(u) \in U$.*

Definition 4.2 *If $U \subseteq V$, then define $U^{\perp} = \{v \in V : (\forall u \in U)(v \perp u)\}$. (Say "U-perp.")*

Exercise 4.3 *If U is a subspace of a finite-dimansional Euclidean space V then* $\dim(U)+\dim(U^{\perp}) = \dim(V)$.

Exercise 4.4 *If φ is either orthogonal or symmetric and U is a φ-invariant subspace $\Rightarrow$ $U^{\perp}$ is also φ-invariant.*