Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

October 18, 2021 (unfinished!)

Contents

Math 504 Lecture 12

1. Jordan canonical (aka normal) form (cont'd)

1.1. The minimal polynomial (cont'd)

Recall from last lecture:

Definition 1.1.1. Let $A \in \mathbb{C}^{n\times n}$ be an $n \times n$-matrix. The preceding theorem shows that there is a **unique** monic polynomial $q_A(t)$ of minimum degree that annihilates A. This unique polynomial will be denoted $q_A(t)$ and will be called the **minimal polynomial** of A.

Theorem 1.1.2. Let $A \in \mathbb{C}^{n\times n}$ be an $n \times n$-matrix. Let $f(t) \in \mathbb{C}[t]$ be any polynomial. Then, f annihilates A if and only if f is a multiple of q_A (that is, $f(t) = q_A(t) \cdot g(t)$ for some polynomial $g(t) \in \mathbb{C}[t]$).

*Drexel University, Korman Center, 15 S 33rd Street, Philadelphia PA, 19104, USA

Corollary 1.1.3. Let $A \in \mathbb{C}^{n\times n}$ be a matrix. Then, $q_A(t) \mid p_A(t)$.

Proposition 1.1.4. If $\lambda \in \sigma(A)$, then $q_A(\lambda) = 0$.

Combining the Corollary with the Proposition, we see that the roots of $q_A(t)$ are precisely the eigenvalues of A; we just don't know yet with which multiplicities they appear as roots. In other words, we have

$$q_A(t) = (t-\lambda_1)^{k_1}(t-\lambda_2)^{k_2}\cdots\left(t-\lambda_p\right)^{k_p},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the distinct eigenvalues of A, and the $k_1, k_2, \ldots, k_p$ are positive integers; but we don't know these $k_1, k_2, \ldots, k_p$ yet. So let us find them. We will use some lemmas for this.

Lemma 1.1.5. Let A and B be two similar $n\times n$-matrices. Then, $q_A(t) = q_B(t)$.

Proof. This is obvious from the viewpoint of endomorphisms. For a pedestrian proof, you can just argue that a polynomial f annihilates A if and only if it annihilates B. But this is easy: We have $A = SBS^{-1}$ for some invertible S (since A and B are similar), and therefore every polynomial f satisfies

$$f(A) = f\left(SBS^{-1}\right) = Sf(B)\,S^{-1}$$

and therefore $f(A) = 0$ holds if and only if $f(B) = 0$. □

We recall the notion of the lcm (= least common multiple) of several polynomials. It is defined as one would expect: If $p_1, p_2, \ldots, p_m$ are m nonzero polynomials (in a single indeterminate t), then $\operatorname{lcm}(p_1, p_2, \ldots, p_m)$ is the monic polynomial of smallest degree that is a common multiple of $p_1, p_2, \ldots, p_m$. For example,

$$\begin{aligned}\operatorname{lcm}\left(t^2-1,\ t^3-1\right) &= \operatorname{lcm}\left((t-1)(t+1),\ (t-1)\left(t^2+t+1\right)\right)\\ &= (t-1)(t+1)\left(t+t^2+1\right) = t^4+t^3-t-1.\end{aligned}$$

(Again, the lcm of several polynomials is unique. This can be shown in the same way that we used to prove uniqueness of the minimal polynomial.)

Lemma 1.1.6. Let $A_1, A_2, \ldots, A_m$ be m square matrices. Let

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix}.$$

Then,

$$q_A = \operatorname{lcm}\left(q_{A_1},\ q_{A_2},\ \ldots,\ q_{A_m}\right).$$

Proof. For any polynomial $f \in \mathbb{C}[t]$, we have

$$f(A) = f\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix} = \begin{pmatrix} f(A_1) & & & \\ & f(A_2) & & \\ & & \ddots & \\ & & & f(A_m) \end{pmatrix}$$

(indeed, the last equality follows from

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix}^k = \begin{pmatrix} A_1^k & & & \\ & A_2^k & & \\ & & \ddots & \\ & & & A_m^k \end{pmatrix}$$

and from the fact that a polynomial f is just a $\mathbb{C}$-linear combination of t^ks). Thus, $f(A) = 0$ holds if and only if

$$f(A_1) = 0 \text{ and } f(A_2) = 0 \text{ and } \cdots \text{ and } f(A_m) = 0.$$

However, $f(A) = 0$ holds if and only if f is a multiple of q_A, whereas $f(A_i) = 0$ holds if and only if f is a multiple of q_{A_i}. Thus, the previous sentence says that f is a multiple of q_A if and only if f is a multiple of all of the q_{A_i}s. In other words, the multiples of q_A are precisely the common multiple of all the q_{A_i}s. But this is the universal property of the lcm. So q_A is the lcm of the q_{A_i}s. □

Lemma 1.1.7. Let $k > 0$ and $\lambda \in \mathbb{C}$. Let $A = J_k(\lambda)$. Then,

$$q_A = (t - \lambda)^k.$$

Proof. It is easy to see that $q_A = q_{A-\lambda I_k}(t - \lambda)$, because for a polynomial f to annihilate $A - \lambda I_k$ is the same as for the polynomial $f(t - \lambda)$ to annihilate A. So we need to find $q_{A-\lambda I_k}$. Recall that

$$A - \lambda I_k = J_k(0) = \begin{pmatrix} & 1 & & \\ & & 1 & \\ & & & \ddots \\ & & & \end{pmatrix}.$$

Therefore, for any polynomial $f = f_0 t^0 + f_1 t^1 + f_2 t^2 + \cdots$, we have

$$f(A - \lambda I_k) = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{k-1} \\ & f_0 & f_1 & \cdots & f_{k-2} \\ & & f_0 & \cdots & f_{k-3} \\ & & & \ddots & \vdots \\ & & & & f_0 \end{pmatrix}.$$

So $f(A-\lambda I_k)=0$ if and only if $f_0=f_1=\cdots=f_{k-1}=0$, i.e., if and only if the first k coefficients of f are 0. Now, the monic polynomial of smallest degree whose first k coefficients are 0 is the polynomial t^k. So the monic polynomial f of smallest degree that satisfies $f(A-\lambda I_k)=0$ is t^k. In other words, $q_{A-\lambda I_k}=t^k$.

Now, recall that $q_A=q_{A-\lambda I_k}(t-\lambda)=(t-\lambda)^k$ (since $q_{A-\lambda I_k}=t^k$). Qed. □

Theorem 1.1.8. Let $A\in\mathbb{C}^{n\times n}$ be an $n\times n$-matrix. Let J be the Jordan canonical form of A. Let $\lambda_1,\lambda_2,\ldots,\lambda_p$ be the distinct eigenvalues of A. Then,

$$q_A=(t-\lambda_1)^{k_1}(t-\lambda_2)^{k_2}\cdots(t-\lambda_p)^{k_p},$$

where k_i is the size of the largest Jordan cell at eigenvalue λ_i in J.

Example 1.1.9. Let A have Jordan canonical form

$$J=\begin{pmatrix} 5 & 1 &&&&&& \\ & 5 & 1 &&&&& \\ && 5 &&&&& \\ &&& 5 & 1 &&& \\ &&&& 5 &&& \\ &&&&& 2 && \\ &&&&&& 2 & 1 \\ &&&&&&& 2 \end{pmatrix}.$$

Then,

$$q_A=(t-5)^3(t-2)^2.$$

Proof of the Theorem. We have $A\sim J$, so that $q_A=q_J$ (by our first lemma).

Recall that J is a Jordan matrix, i.e., a block-diagonal matrix whose diagonal blocks are Jordan cells $J_1,J_2,\ldots,J_m$. Thus, by our second lemma, we have

$$\begin{aligned} q_J &= \operatorname{lcm}(q_{J_1},\ q_{J_2},\ \ldots,\ q_{J_m}) \\ &= \operatorname{lcm}\left((t-\lambda_{J_1})^{k_{J_1}},\ (t-\lambda_{J_2})^{k_{J_2}},\ \ldots,\ (t-\lambda_{J_m})^{k_{J_m}}\right), \end{aligned}$$

where each J_i has eigenvalue λ_{J_i} and size k_{J_i} (by our third lemma). This lcm must be divisible by each $t-\lambda$ at least as often as each of the $(t-\lambda_{J_i})^{k_{J_i}}$s is; i.e., it must be divisible by $(t-\lambda)^k$, where k is the largest size of a Jordan cell of J at eigenvalue λ. So the lcm is the product of these $(t-\lambda)^k$s. But this is precisely our claim. □

1.2. Application of functions to matrices

Consider a square matrix $A\in\mathbb{C}^{n\times n}$. We have already defined what it means to apply a polynomial f to A: We just write f as $\sum_i f_i t^i$, and substitute A for t.

Can we do the same with non-polynomial functions f ? For example, can we define $\exp A$ or $\sin A$?

One option to do so is to follow the same rule as for polynomials, but using the Taylor series for f. For example, since exp has Taylor series $\exp t = \sum_{i \in \mathbb{N}} \frac{t^i}{i!}$, we can set

$$\exp A = \sum_{i \in \mathbb{N}} \frac{A^i}{i!}.$$

This indeed works for exp and for sin, as the sums you get always converge. But it doesn't generally work, e.g., for $f = \tan t$, since its Taylor series only converges in a certain neighborhood of 0. Is this the best we can do?

There is a different approach that gives a more general definition.

Lemma 1.2.1. Let $k > 0$ and $\lambda \in \mathbb{C}$. Let $A = J_k(\lambda)$. Then, for any polynomial $f \in \mathbb{C}[t]$, we have

$$f(A) = \begin{pmatrix} \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(k-1)}(\lambda)}{(k-1)!} \\ & \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \cdots & \frac{f^{(k-2)}(\lambda)}{(k-2)!} \\ & & \frac{f(\lambda)}{0!} & \cdots & \frac{f^{(k-3)}(\lambda)}{(k-3)!} \\ & & & \ddots & \vdots \\ & & & & \frac{f(\lambda)}{0!} \end{pmatrix}$$

Proof. Exercise. $\square$

Now, we aim to define $f(A)$ by the above formula, at least when A is a Jordan cell. This only requires f to be $(k-1)$-times differentiable at λ.

Definition 1.2.2. Let $A \in \mathbb{C}^{n \times n}$ be an $n \times n$-matrix that has minimal polynomial

$$q_A(t) = (t - \lambda_1)^{k_1} (t - \lambda_2)^{k_2} \cdots (t - \lambda_p)^{k_p},$$

where the $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the distinct eigenvalues of A.

Let f be a function from $\mathbb{C}$ to $\mathbb{C}$ that is defined at each of the numbers $\lambda_1, \lambda_2, \ldots, \lambda_p$ and is holomorphic at each of them, or at least $(k_i - 1)$-times differentiable at each λ_i if λ_i is real. Then, we can define an $n \times n$-matrix $f(A) \in \mathbb{C}^{n \times n}$ as follows: Write $A = SJS^{-1}$, where J is a Jordan matrix and S is invertible. Write

J as $\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix}$, where the $J_1, J_2, \ldots, J_m$ are Jordan cells. Then, we set

$$f(A) := Sf(J)S^{-1}, \qquad \text{where}$$

$$f(J) := \begin{pmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_m) \end{pmatrix}, \qquad \text{where}$$

$$f(J_k(\lambda)) := \begin{pmatrix} \dfrac{f(\lambda)}{0!} & \dfrac{f'(\lambda)}{1!} & \dfrac{f''(\lambda)}{2!} & \cdots & \dfrac{f^{(k-1)}(\lambda)}{(k-1)!} \\ & \dfrac{f(\lambda)}{0!} & \dfrac{f'(\lambda)}{1!} & \cdots & \dfrac{f^{(k-2)}(\lambda)}{(k-2)!} \\ & & \dfrac{f(\lambda)}{0!} & \cdots & \dfrac{f^{(k-3)}(\lambda)}{(k-3)!} \\ & & & \ddots & \vdots \\ & & & & \dfrac{f(\lambda)}{0!} \end{pmatrix}.$$

Theorem 1.2.3. This definition is actually well-defined. That is, the value $f(A)$ does not depend on the choice of S and J.

Exercise 1.2.1. Prove this.

1.3. The companion matrix

For each $n \times n$-matrix A, we have defined its characteristic polynomial p_A and its minimal polynomial q_A. What variety of polynomials do we get this way? Do all characteristic polynomials share some property, or can any monic polynomial be a characteristic polynomial? The same question for minimal polynomials?