Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

October 15, 2021 (unfinished!)

Contents

Math 504 Lecture 11

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

1.1. Powers and the Jordan canonical form

Let $n \in \mathbb{N}$ and $A \in \mathbb{C}^{n \times n}$. Assume that we know the JCF J of A and an invertible matrix S such that

$$A = SJS^{-1}.$$

Then, it is fairly easy to compute all powers A^m of A. Indeed, recall that

- $\left(SJS^{-1}\right)^m = SJ^m S^{-1}$ for any $m \in \mathbb{N}$.
- $\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}^m = \begin{pmatrix} A_1^m & & & \\ & A_2^m & & \\ & & \ddots & \\ & & & A_k^m \end{pmatrix}$ for any $m \in \mathbb{N}$.

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

Thus, it suffices to compute the m-th power of any Jordan cell $J_k(\lambda)$.
So let us consider a Jordan cell

$$C := J_5(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}.$$

Then,

$$C^2 = \begin{pmatrix} \lambda^2 & 2\lambda & 1 & & \\ & \lambda^2 & 2\lambda & 1 & \\ & & \lambda^2 & 2\lambda & 1 \\ & & & \lambda^2 & 2\lambda \\ & & & & \lambda^2 \end{pmatrix}; \qquad C^3 = \begin{pmatrix} \lambda^3 & 3\lambda^2 & 3\lambda & 1 & \\ & \lambda^3 & 3\lambda^2 & 3\lambda & 1 \\ & & \lambda^3 & 3\lambda^2 & 3\lambda \\ & & & \lambda^3 & 3\lambda^2 \\ & & & & \lambda^3 \end{pmatrix};$$

$$C^4 = \begin{pmatrix} \lambda^4 & 4\lambda^3 & 6\lambda^2 & 4\lambda & 1 \\ & \lambda^4 & 4\lambda^3 & 6\lambda^2 & 4\lambda \\ & & \lambda^4 & 4\lambda^3 & 6\lambda^2 \\ & & & \lambda^4 & 4\lambda^3 \\ & & & & \lambda^4 \end{pmatrix}; \qquad C^5 = \begin{pmatrix} \lambda^5 & 5\lambda^4 & 10\lambda^3 & 10\lambda^2 & 5\lambda \\ & \lambda^5 & 5\lambda^4 & 10\lambda^3 & 10\lambda^2 \\ & & \lambda^5 & 5\lambda^4 & 10\lambda^3 \\ & & & \lambda^5 & 5\lambda^4 \\ & & & & \lambda^5 \end{pmatrix}.$$

In general:

Theorem 1.1.1. Let $k > 0$ and $\lambda \in \mathbb{C}$. Let $C = J_k(\lambda)$. Let $m \in \mathbb{N}$. Then, C^m is the upper-triangular $k \times k$-matrix whose (i,j)-th entry is $\dbinom{m}{j-i}\lambda^{m-j+i}$. (Here, we follow the convention that $\dbinom{n}{\ell} := 0$ when $\ell \notin \mathbb{N}$. Also, recall that $\dbinom{n}{\ell} = 0$ when $n \in \mathbb{N}$ and $\ell > n$.)

First proof. Induct on m and use $C^m = CC^{m-1}$ as well as Pascal's recursion

$$\binom{n}{\ell} = \binom{n-1}{\ell} + \binom{n-1}{\ell-1}.$$

$\square$

Second proof. Set $B := J_k(0) = \begin{pmatrix} & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ & & & & \end{pmatrix}$. We know all powers of B already: B^i has 1s i steps above the main diagonal, and 0s everywhere else.

However, $C = B + \lambda I_k$. The matrices λI_k and B commute (i.e., we have $B \cdot \lambda I_k = \lambda I_k \cdot B$). It is a general fact that if X and Y are two commuting $n \times n$-matrices, then the binomial formula

$$(X+Y)^m = \sum_{i=0}^{m} \binom{m}{i} X^i Y^{m-i}.$$

(This can be proved in the same way as for numbers, because the commutativity of X and Y lets you move any Xes past any Ys.) Applying this formula to $X = B$ and $Y = \lambda I_k$, we obtain

$$(B + \lambda I_k)^m = \sum_{i=0}^{m} \binom{m}{i} B^i \underbrace{(\lambda I_k)^{m-i}}_{=\lambda^{m-i} I_k} = \sum_{i=0}^{m} \binom{m}{i} \lambda^{m-i} B^i$$

$$= \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots & \cdots & \cdots \\ & \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots & \cdots \\ & & \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots \\ & & & \lambda^m & \binom{m}{1}\lambda^{m-1} & \cdots & \cdots \\ & & & & \lambda^m & \cdots & \vdots \\ & & & & & \ddots & \vdots \\ & & & & & & \lambda^m \end{pmatrix},$$

which is precisely the matrix claimed in the theorem. $\square$

Now we know how to take powers of Jordan cells, and therefore how to take powers of any matrix that we know how to bring to a Jordan canonical form.

Corollary 1.1.2. Let $A \in \mathbb{C}^{n \times n}$. Then, $\lim\limits_{m \to \infty} A^m = 0$ if and only if all eigenvalues of A have absolute value < 1.

Proof. $\Longrightarrow$: Suppose that $\lim\limits_{m \to \infty} A^m = 0$, but A has an eigenvalue λ of absolute value ≥ 1. We want a contradiction.

Consider a nonzero eigenvector x for eigenvalue λ. Thus, $Ax = \lambda x$. Then, $A^2 x = \lambda^2 x$ (since $A^2 x = A \underbrace{Ax}_{=\lambda x} = \lambda \underbrace{Ax}_{=\lambda x} = \lambda\lambda x = \lambda^2 x$) and similarly $A^3 x = \lambda^3 x$ and $A^4 x = \lambda^4 x$ and so on. Thus,

$$A^m x = \lambda^m x \qquad \text{for each } m \in \mathbb{N}.$$

Now, as $m \to \infty$, the vector $A^m x$ goes to 0 (since $A^m \to 0$), but the vector $\lambda^m x$ does not (since $x \neq 0$ and $|\lambda| \geq 1$). Contradiction.

$\Longleftarrow$: Suppose that all eigenvalues of A have absolute value < 1.

Let $A = SJS^{-1}$ be the Jordan canonical form of A. Write J as $\begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix}$, where each J_i is a Jordan cell.

It suffices to show that $\lim\limits_{n\to\infty} J_i^m = 0$.

Write J_i as $J_k(\lambda)$, with $|\lambda| < 1$. The preceding theorem then tells us that

$$J_i^m = \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots & \cdots & \cdots \\ & \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots & \cdots \\ & & \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \cdots \\ & & & \lambda^m & \binom{m}{1}\lambda^{m-1} & \cdots & \cdots \\ & & & & \lambda^m & \cdots & \vdots \\ & & & & & \ddots & \vdots \\ & & & & & & \lambda^m \end{pmatrix}.$$

Look closely at this matrix. We need to show that for each $i, j \in [k]$, we have

$$\lim_{m\to\infty} \binom{m}{j-i} \lambda^{m-j+i} = 0.$$

However, this is a standard asymptotics argument:

$$\underbrace{\binom{m}{j-i}}_{\substack{= \frac{m(m-1)(m-2)\cdots(m-j+i)}{(j-i)!} \\ \text{(for } i\leq j\text{; otherwise the claim is trivial)}}} \underbrace{\lambda^{m-j+i}}_{\substack{\text{exponential in } m, \\ \text{with quotient } \lambda \text{ having absolute value } |\lambda|<1}} \to 0$$

because exponential functions with a quotient of absolute value < 1 converge to 0 faster than polynomials can go to ∞. $\square$

1.2. The minimal polynomial

Recall: A polynomial $p(t) \in \mathbb{F}[t]$ (where $\mathbb{F}$ is any field, and t is an indeterminate) is said to be **monic** if its leading coefficient is 1 – that is, if it can be written in the form

$$p(t) = t^m + p_{m-1}t^{m-1} + p_{m-2}t^{m-2} + \cdots + p_0 t^0 \qquad \text{for some } m \in \mathbb{N} \text{ and } p_0, p_1, \ldots, p_{m-1} \in \mathbb{F}.$$

Definition 1.2.1. Given a matrix $A \in \mathbb{F}^{n\times n}$ and a polynomial $p(t) \in \mathbb{F}[t]$, we saay that $p(t)$ **annihilates** A if $p(A) = 0$.

The Cayley–Hamilton theorem says that the characteristic polynomial p_A of a square matrix A always annihilates A. However, often there are matrices that are annihilated by other – sometimes simpler – polynomials.

Example 1.2.2. The identity matrix I_n is annihilated by the polynomial $p(t) := t - 1$, because

$$p(I_n) = I_n - I_n = 0.$$

Example 1.2.3. The matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is annihilated by the polynomial $p(t) := t^2$, since its square is 0.

Example 1.2.4. The matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is annihilated by the polynomial $p(t) := t^2$, since its square is 0.

Example 1.2.5. The diagonal matrix $\begin{pmatrix} 2 & & \\ & 2 & \\ & & 3 \end{pmatrix}$ is annihilated by the polynomial $p(t) := (t-2)(t-3)$, since

$$\begin{aligned} &\left(\begin{pmatrix} 2 & & \\ & 2 & \\ & & 3 \end{pmatrix} - 2I_n\right)\left(\begin{pmatrix} 2 & & \\ & 2 & \\ & & 3 \end{pmatrix} - 3I_n\right) \\ &= \begin{pmatrix} & & \\ & & \\ & & 1 \end{pmatrix}\begin{pmatrix} -1 & & \\ & -1 & \\ & & \end{pmatrix} = 0. \end{aligned}$$

Theorem 1.2.6. Given an $n \times n$-matrix $A \in \mathbb{C}^{n\times n}$. Then, there is a **unique** monic polynomial $q_A(t)$ of minimum degree that annihilates A.

Proof. The Cayley–Hamilton theorem shows that p_A annihilates A. Since p_A is monic, we thus conclude that there exists **some** monic polynomial that annihilates A. Hence, there exists such a polynomial of minimum degree.

It remains to show that it is unique. To do so, we let q_A and $\widetilde{q}_A$ be two monic polynomials of minimum degree that annihilate A. Our goal then is to show that $q_A = \widetilde{q}_A$.

Indeed, if $q_A \neq \widetilde{q}_A$, then the difference $q_A - \widetilde{q}_A$ is a polynomial of smaller degree that annihilates A (indeed, it is of smaller degree because q_A and $\widetilde{q}_A$ are monic of the same degree and thus lose their leading terms upon subtraction; it annihilates A because $(q_A - \widetilde{q}_A)(A) = q_A(A) - \widetilde{q}_A(A) = 0 - 0 = 0$). By scaling this difference appropriately, we can make it monic (since it is nonzero), and then get a contradiction to the minimality of q_A's degree. This concludes the proof. $\square$

Definition 1.2.7. 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.

Example 1.2.8. Let A be the diagonal matrix $\begin{pmatrix} 2 & & \\ & 2 & \\ & & 3 \end{pmatrix}$. Then,

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

Indeed, we already know that the monic polynomial $(t-2)(t-3)$ annihilates A. If there was any monic polynomial of smaller degree that would annihilate A, then it would have the form $t - \lambda$ for some $\lambda \in \mathbb{C}$, but λ cannot be 2 and 3 at the same time.

For comparison: $p_A(t) = (t-2)^2(t-3)$.

Theorem 1.2.9. 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 $g(t) \in \mathbb{C}[t]$).

Proof. $\Longrightarrow$: Assume that f annihilates A. Thus, $f(A) = 0$. WLOG, assume that $f \neq 0$. Thus, we can make f monic by scaling it. Thus, $\deg f \geq \deg q_A$ (since q_A had minimum degree). Hence, we can divide f by q_A with remainder, obtaining

$$f(t) = q_A(t) \cdot g(t) + r(t),$$

where $g(t)$ and $r(t)$ are two polynomials with $\deg r < \deg q_A$. Substituting A for t in this equality, we obtain

$$f(A) = \underbrace{q_A(A)}_{\substack{=0 \\ (\text{since } q_A \text{ annihilates } A)}} \cdot g(A) + r(A) = r(A),$$

so that $r(A) = f(A) = 0$. In other words, r annihilates A. Since $\deg r < \deg q_A$, this entails that $r = 0$ (since otherwise, scaling r to make it monic would contradict the minimality of $\deg q_A$). Thus,

$$f(t) = q_A(t) \cdot g(t) + \underbrace{r(t)}_{=0} = q_A(t) \cdot g(t).$$

Thus, f is a multiple of q_A.

$\Longleftarrow$: Easy and LTTR. $\square$

Corollary 1.2.10. Let $A \in \mathbb{C}^{n \times n}$ be a matrix. Then, $q_A(t) \mid p_A(t)$.

Proof. Apply the previous theorem to $f = p_A$, recalling that p_A annihilates A. $\square$

The corollary yields that any root of q_A must be a root of p_A, that is, an eigenvalue of A. Conversely, we can show that any eigenvalue of A is a root of q_A (but we don't know with which multiplicity):

Proposition 1.2.11. If $\lambda \in \sigma(A)$, then $q_A(\lambda) = 0$.

Proof. Let $\lambda \in \sigma(A)$. Thus, there exists a nonzero eigenvector x for λ.

Then, $Ax = \lambda x$. As we have seen above, this entails $A^m x = \lambda^m x$ for each $m \in \mathbb{N}$. Therefore, $f(A)x = f(\lambda)x$ for each polynomial $f(t) \in \mathbb{C}[t]$ (because you can write $f(t)$ as $f_0 t^0 + f_1 t^1 + \cdots + f_p t^p$, and then apply $A^m x = \lambda^m x$ to each of $m = 0, 1, \ldots, p$). Hence, $q_A(A)x = q_A(\lambda)x$, so that

$$q_A(\lambda)x = \underbrace{q_A(A)}_{\substack{=0 \\ \text{(since } q_A \text{ annihilates } A)}} x = 0.$$

Since $x \neq 0$, this entails $q_A(\lambda) = 0$, qed. $\square$