Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

October 6, 2021 (unfinished!)

Contents

Math 504 Lecture 8

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

1.1. The Jordan canonical form (aka Jordan normal form) (cont'd)

Recall from last lecture:

Definition 1.1.1. A **Jordan cell** is a $m \times m$-matrix of the form

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix} \qquad \text{for some } m > 0 \text{ and some } \lambda \in \mathbb{C}.$$

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

So its diagonal entries are λ; its entries directly above the diagonal are 1; all its other entries are 0. In formal terms, it is the $m \times m$-matrix A whose entries are given by the rule

$$A_{i,j} = \begin{cases} \lambda, & \text{if } i = j; \\ 1, & \text{if } i = j - 1; \\ 0, & \text{otherwise.} \end{cases}$$

Specifically, this matrix is called the **Jordan cell of size m at eigenvalue λ**. It is denoted by $J_m(\lambda)$.

Definition 1.1.2. A **Jordan matrix** is a block-diagonal matrix whose diagonal blocks are Jordan cells. In other words, it is a matrix of the form

$$\begin{pmatrix} J_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k) \end{pmatrix},$$

where $n_1, n_2, \ldots, n_k$ are positive integers and $\lambda_1, \lambda_2, \ldots, \lambda_k$ are scalars in $\mathbb{C}$ (not necessarily distinct, but not necessarily equal either).

We claimed:

Theorem 1.1.3 (Jordan canonical form theorem)**.** Let A be an $n \times n$-matrix over $\mathbb{C}$. Then, there exists a Jordan matrix J such that $A \sim J$. Furthermore, this J is unique up to the order of the diagonal blocks.

Definition 1.1.4. The matrix J in this theorem is called a **Jordan canonical form** of A (or a **Jordan normal form** of A).

Proposition 1.1.5. Let A be an $n \times n$-matrix, and let $J = \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k) \end{pmatrix}$ be its Jordan canonical form. Then:

(a) We have $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$.

(b) The geometric multiplicity of an eigenvalue λ of A is the number of Jordan cells of A at eigenvalue λ. In other words, it is the number of $i \in [k]$ satisfying $\lambda_i = \lambda$.

(c) The algebraic multiplicity of an eigenvalue λ of A is the **sum** of the sizes of all Jordan cells of A at eigenvalue λ. In other words, it is $\sum\limits_{\substack{i \in [k]; \\ \lambda_i = \lambda}} n_i$.

Now, let us improve on this result somewhat:

Proposition 1.1.6. Let A be an $n \times n$-matrix, and let $J = \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k) \end{pmatrix}$ be its Jordan canonical form. Let $\lambda \in \mathbb{C}$ and $m \geq 1$. Then,

$$\begin{aligned} &\left(\text{the number of } i \in [k] \text{ such that } \lambda_i = \lambda \text{ and } n_i \geq m\right) \\ &= \dim\left(\operatorname{Ker}\left((A - \lambda I_n)^m\right)\right) - \dim\left(\operatorname{Ker}\left((A - \lambda I_n)^{m-1}\right)\right). \end{aligned}$$

This proposition gives us a way to tell how many Jordan blocks of the Jordan form of A have eigenvalue λ and size $\geq m$. As a consequence, this number is uniquely determined by A and λ. Hence, the whole structure of J is determined uniquely by A, up to the order of the Jordan blocks. This shows the "uniqueness" part of the Jordan canonical form theorem, as long as we can prove the proposition.

Example 1.1.7. Let A be an 8×8-matrix. Assume that we know that A has

- 1 Jordan block of size ≥ 1 for eigenvalue 17;
- 0 Jordan blocks of size ≥ 2 for eigenvalue 17;
- 3 Jordan blocks of size ≥ 1 for eigenvalue 35;
- 1 Jordan block of size ≥ 2 for eigenvalue 35;
- 0 Jordan blocks of size ≥ 3 for eigenvalue 35;
- 1 Jordan block of size ≥ 1 for eigenvalue 59;
- 1 Jordan block of size ≥ 2 for eigenvalue 59;
- 1 Jordan block of size ≥ 3 for eigenvalue 59;
- 0 Jordan blocks of size ≥ 4 for eigenvalue 59;
- 0 Jordan blocks of size ≥ 1 for eigenvalue λ whenever $\lambda \notin \{17, 35, 59\}$.

(This is the sort of information you can obtain from A using the preceding proposition.)

How does the Jordan canonical form of A look like? It is the block-diagonal matrix

$$\begin{pmatrix} J_1(17) & & & & \\ & J_2(35) & & & \\ & & J_1(35) & & \\ & & & J_1(35) & \\ & & & & J_3(59) \end{pmatrix}$$

(where all invisible entries are 0s), or one that is obtained from it by permuting the diagonal blocks.

Let us prove the proposition:

Proof of Proposition. We have $A \sim J$, so that $A - \lambda I_n \sim J - \lambda I_n$, so that $(A - \lambda I_n)^m \sim (J - \lambda I_n)^m$ and therefore

$$\dim\left(\operatorname{Ker}\left((A - \lambda I_n)^m\right)\right) = \dim\left(\operatorname{Ker}\left((J - \lambda I_n)^m\right)\right).$$

However

$$\begin{aligned} J - \lambda I_n &= \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k) \end{pmatrix} - \lambda I_n \\ &= \begin{pmatrix} J_{n_1}(\lambda_1 - \lambda) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2 - \lambda) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k - \lambda) \end{pmatrix}, \end{aligned}$$

so that

$$\begin{aligned} (J - \lambda I_n)^m &= \begin{pmatrix} J_{n_1}(\lambda_1 - \lambda) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\lambda_2 - \lambda) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k}(\lambda_k - \lambda) \end{pmatrix}^m \\ &= \begin{pmatrix} (J_{n_1}(\lambda_1 - \lambda))^m & 0 & \cdots & 0 \\ 0 & (J_{n_2}(\lambda_2 - \lambda))^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (J_{n_k}(\lambda_k - \lambda))^m \end{pmatrix} \\ &\qquad \left(\begin{array}{c} \text{because multiplication of block-diagonal matrices} \\ \text{means that respective blocks get multiplied} \end{array}\right). \end{aligned}$$

Thus,

$$\begin{aligned}
&\dim\left(\operatorname{Ker}\left((J-\lambda I_n)^m\right)\right)\\
&=\dim\left(\operatorname{Ker}\begin{pmatrix}(J_{n_1}(\lambda_1-\lambda))^m & 0 & \cdots & 0\\ 0 & (J_{n_2}(\lambda_2-\lambda))^m & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & (J_{n_k}(\lambda_k-\lambda))^m\end{pmatrix}\right)\\
&=\dim\left(\operatorname{Ker}\left((J_{n_1}(\lambda_1-\lambda))^m\right)\right)+\dim\left(\operatorname{Ker}\left((J_{n_2}(\lambda_2-\lambda))^m\right)\right)+\cdots+\dim\left(\operatorname{Ker}\left((J_{n_k}(\lambda_k-\lambda))^m\right)\right)\\
&=\sum_{i=1}^{k}\dim\left(\operatorname{Ker}\left((J_{n_i}(\lambda_i-\lambda))^m\right)\right).
\end{aligned}$$

Now, fix $i \in [k]$. If $\lambda_i \neq \lambda$, then $J_{n_i}(\lambda_i-\lambda)$ is a triangular matrix with nonzero entries on the diagonal (in fact, the diagonal entries are all $\lambda_i - \lambda \neq 0$), and thus has determinant $\neq 0$ and therefore is invertible. Hence, in this case, its power $(J_{n_i}(\lambda_i-\lambda))^m$ is also invertible, so it has nullity 0. Thus,

$$\text{if } \lambda_i \neq \lambda, \text{ then } \dim\left(\operatorname{Ker}\left((J_{n_i}(\lambda_i-\lambda))^m\right)\right)=0.$$

Now consider the case when $\lambda_i = \lambda$. Then, $J_{n_i}(\lambda_i-\lambda) = J_{n_i}(0) = \begin{pmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\0&0&0&\cdots&0\\ \vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&0\end{pmatrix}$.

Let me call this matrix B. From last time, we know that

$$\dim\left(\operatorname{Ker}\left(B^m\right)\right)=\begin{cases}m, & \text{if } m \leq n_i;\\ n_i, & \text{if } m > n_i.\end{cases}$$

Thus,

$$\dim\left(\operatorname{Ker}\left((J_{n_i}(\lambda_i-\lambda))^m\right)\right)=\begin{cases}m, & \text{if } m \leq n_i;\\ n_i, & \text{if } m > n_i\end{cases} \qquad \text{in this case.}$$

Now let us see what this means for our sum:

$$\begin{aligned}
&\sum_{i=1}^{k}\dim\left(\operatorname{Ker}\left((J_{n_i}(\lambda_i-\lambda))^m\right)\right)\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases}m, & \text{if } m \leq n_i;\\ n_i, & \text{if } m > n_i.\end{cases}
\end{aligned}$$

So

$$\begin{aligned}
&\dim\left(\operatorname{Ker}\left((A-\lambda I_n)^m\right)\right)\\
&=\dim\left(\operatorname{Ker}\left((J-\lambda I_n)^m\right)\right)\\
&=\sum_{i=1}^{k}\dim\left(\operatorname{Ker}\left((J_{n_i}(\lambda_i-\lambda))^m\right)\right)\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases} m, & \text{if } m\le n_i;\\ n_i, & \text{if } m>n_i.\end{cases}
\end{aligned}$$

The same argument, applied to $m-1$ instead of m, yields

$$\begin{aligned}
&\dim\left(\operatorname{Ker}\left((A-\lambda I_n)^{m-1}\right)\right)\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases} m-1; & \text{if } m-1\le n_i;\\ n_i, & \text{if } m-1>n_i.\end{cases}
\end{aligned}$$

Subtracting these two equalities, we obtain

$$\begin{aligned}
&\dim\left(\operatorname{Ker}\left((A-\lambda I_n)^m\right)\right)-\dim\left(\operatorname{Ker}\left((A-\lambda I_n)^{m-1}\right)\right)\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases} m, & \text{if } m\le n_i;\\ n_i, & \text{if } m>n_i\end{cases}-\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases} m-1; & \text{if } m-1\le n_i;\\ n_i, & \text{if } m-1>n_i\end{cases}\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\underbrace{\left(\begin{cases} m, & \text{if } m\le n_i;\\ n_i, & \text{if } m>n_i\end{cases}-\begin{cases} m-1; & \text{if } m-1\le n_i;\\ n_i, & \text{if } m-1>n_i\end{cases}\right)}_{\substack{=\begin{cases} 1, & \text{if } m\le n_i;\\ 0, & \text{if } m>n_i\end{cases}\\ \text{(this can be directly checked in each}\\ \text{of the three cases } m\le n_i \text{ and } m=n_i+1 \text{ and } m>n_i+1)}}\\
&=\sum_{\substack{i\in[k];\\ \lambda_i=\lambda}}\begin{cases} 1, & \text{if } m\le n_i;\\ 0, & \text{if } m>n_i\end{cases}\\
&=(\text{the number of } i\in[k] \text{ such that } \lambda_i=\lambda \text{ and } m\le n_i)\\
&=(\text{the number of } i\in[k] \text{ such that } \lambda_i=\lambda \text{ and } n_i\ge m)\,.
\end{aligned}$$

Thus, the proposition follows. □

So we are done proving the uniqueness part of the JCF theorem. ("JCF" is short for "Jordan Canonical Form".)

Let us now approach the existence part.

1.2. Proof of the existence part of the JCF

1.2.1. Step 1: Schur triangularization

We have an $n \times n$-matrix $A \in \mathbb{C}^{n\times n}$, and we want to make it into a Jordan matrix by conjugation.

First, we make it upper-triangular by conjugation. We know that this is possible by the Schur triangularization theorem, but we are a bit pickier now. To wit, we want the triangular matrix T to have the property that equal eigenvalues form contiguous blocks on the main diagonal. For instance, we don't want

$$T = \begin{pmatrix} 1 & * & * & * \\ & 2 & * & * \\ & & 1 & * \\ & & & 2 \end{pmatrix}.$$

Instead, we want

$$T = \begin{pmatrix} 1 & * & * & * \\ & 1 & * & * \\ & & 2 & * \\ & & & 2 \end{pmatrix}.$$

To this purpose, we need a stronger version of Schur triangularization:

> **Theorem 1.2.1** (Schur triangularization with perscribed diagonal)**.** Let $A \in \mathbb{C}^{n\times n}$ be an $n \times n$-matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues (listed with their algebraic multiplicities). Then, there exists an upper-triangular matrix T such that $A \overset{\text{us}}{\sim} T$ (this means "A is unitarily similar to T") and such that the diagonal entries of T are $\lambda_1, \lambda_2, \ldots, \lambda_n$ in this order.

Proof. We proceed as in the proof of the original Schur triangularization theorem, but we pay some attention to the eigenvectors that we pick. That proof constructed T recursively, starting by picking an eigenvalue λ of A and a corresponding λ-eigenvector $x \neq 0$, and then finding a unitary matrix U such that

$$U^* A U = \begin{pmatrix} \lambda & p \\ 0 & B \end{pmatrix} \qquad \text{for some } p \in \mathbb{C}^{1\times(n-1)} \text{ and } B \in \mathbb{C}^{(n-1)\times(n-1)}.$$

Then, the algorithm was applied recursively to B.

Now, we perform this algorithm, but we make sure to pick $\lambda = \lambda_1$. Thus,

$$U^* A U = \begin{pmatrix} \lambda_1 & p \\ 0 & B \end{pmatrix},$$

which is a good start. Now we want to apply the IH (= induction hypothesis) to B. To that purpose, we need to know that the eigenvalues of B are $\lambda_2, \lambda_3, \ldots, \lambda_n$ (with algebraic multiplicities). In other words, we need to know that

$$p_B(t) = (t - \lambda_2)(t - \lambda_3) \cdots (t - \lambda_n).$$

Now, I need a lemma about determinants:

> **Lemma 1.2.2.** Let $\mathbb{F}$ be a field. Let $X \in \mathbb{F}^{n\times n}$, $Y \in \mathbb{F}^{n\times m}$ and $Z \in \mathbb{F}^{m\times m}$ be three matrices. Then,
> $$\det \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix} = \det X \cdot \det Z.$$

For example,

$$\det \begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \\ 0 & 0 & c'' & d'' \\ 0 & 0 & c''' & d''' \end{pmatrix} = \det \begin{pmatrix} a & b \\ a' & b' \end{pmatrix} \cdot \det \begin{pmatrix} c'' & d'' \\ c''' & d''' \end{pmatrix}.$$

Now, since U is unitary, we have $A \overset{\text{us}}{\sim} U^*AU$ and thus $A \sim U^*AU$. Thus,

$$\begin{aligned} p_A = p_{U^*AU} &= \det(tI_n - U^*AU) = \det\left(tI_n - \begin{pmatrix} \lambda_1 & p \\ 0 & B \end{pmatrix}\right) \\ &\qquad \left(\text{since } U^*AU = \begin{pmatrix} \lambda_1 & p \\ 0 & B \end{pmatrix}\right) \\ &= \det \begin{pmatrix} t-\lambda_1 & -p \\ 0 & tI_{n-1} - B \end{pmatrix} \\ &= \underbrace{\det\left(t-\lambda_1 \right)}_{t-\lambda_1} \cdot \underbrace{\det(tI_{n-1} - B)}_{=p_B} \qquad \left(\begin{array}{c}\text{by the lemma, or by Laplace} \\ \text{expansion along the 1st column}\end{array}\right) \\ &= (t-\lambda_1) \cdot p_B, \end{aligned}$$

so that

$$(t-\lambda_1) \cdot p_B = p_A = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n).$$

Now, we can cancel $t - \lambda_1$ (since this is a nonzero polynomial), and this becomes

$$p_B = (t-\lambda_2)(t-\lambda_3)\cdots(t-\lambda_n),$$

exactly as we wanted to show. □

1.3. Step 2: Separating distinct eigenvalues

Recall the corollary from last time:

> **Corollary 1.3.1.** Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{m\times m}$ and $C \in \mathbb{C}^{n\times m}$ be three matrices such that $\sigma(A) \cap \sigma(B) = \varnothing$. Then, the two $(n+m) \times (n+m)$-matrices
> $$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$
> (written in block-matrix notation) are similar.

Let us use this corollary to show that for any numbers $a, b, c, \ldots, p \in \mathbb{C}$, we have

$$\begin{pmatrix} 1 & a & b & c & d & e \\ & 1 & f & g & h & i \\ & & 2 & j & k & \ell \\ & & & 2 & m & n \\ & & & & 2 & p \\ & & & & & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & a & & & & \\ & 1 & & & & \\ & & 2 & j & k & \\ & & & 2 & m & \\ & & & & 2 & \\ & & & & & 3 \end{pmatrix}$$

(where invisible cells contain 0s). Indeed, the triangular matrices $\begin{pmatrix} 1 & a \\ & 1 \end{pmatrix}$ and $\begin{pmatrix} 2 & j & k & \ell \\ & 2 & m & n \\ & & 2 & p \\ & & & 3 \end{pmatrix}$ have disjoint spectra (i.e., they have no eigenvalues in common), because their diagonals have no entries in common. So, by the corollary,

$$\begin{pmatrix} 1 & a & b & c & d & e \\ & 1 & f & g & h & i \\ & & 2 & j & k & \ell \\ & & & 2 & m & n \\ & & & & 2 & p \\ & & & & & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & a & & & & \\ & 1 & & & & \\ & & 2 & j & k & \ell \\ & & & 2 & m & n \\ & & & & 2 & p \\ & & & & & 3 \end{pmatrix}.$$

Now, the triangular matrices $\begin{pmatrix} 1 & a & & & \\ & 1 & & & \\ & & 2 & j & k \\ & & & 2 & m \\ & & & & 2 \end{pmatrix}$ and $\begin{pmatrix} 3 \end{pmatrix}$ have disjoint spectra, so the corollary yields

$$\begin{pmatrix} 1 & a & & & & \\ & 1 & & & & \\ & & 2 & j & k & \ell \\ & & & 2 & m & n \\ & & & & 2 & p \\ & & & & & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & a & & & & \\ & 1 & & & & \\ & & 2 & j & k & \\ & & & 2 & m & \\ & & & & 2 & \\ & & & & & 3 \end{pmatrix}.$$

Since $\sim$ is an equivalence relation, we can combine the two similarities, and my claim follows.

So when we have a triangular matrix where the diagonal has no interleaving values (i.e., there is never a μ between two λ's on the diagonal when $\mu \neq \lambda$), we can "clean out" all the above-diagonal entries that correspond to different diagonal entries (i.e., that lie above a different diagonal entry than they stand to the right of) by conjugating with an appropriate matrix.

Now, combining this with the Schur triangularization theorem (in its above-stated stronger version), we obtain the following:

Proposition 1.3.2. Let $A \in \mathbb{C}^{n \times n}$ be an $n \times n$-matrix. Then, A is similar to a block-diagonal matrix of the form

$$\begin{pmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_k \end{pmatrix},$$

where each B_i is an upper-triangular matrix with all entries on its diagonal being equal.

This is not yet a Jordan canonical form, but it is already somewhat close. At least, we have separated out all the distinct eigenvalues of A. We now only need to care about the matrices $B_1, B_2, \ldots, B_k$, each of which has only one distinct eigenvalue. Our next goal is to break up these matrices $B_1, B_2, \ldots, B_k$ into Jordan cells (using conjugation).