JORDAN CANONICAL FORM

We will show that, if V is a finite dimensional complex vector space, then every operator $T \in \mathcal{L}(V)$ has a basis in which its matrix is in **Jordan canonical form:**

$$M(T) = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}$$

where each **Jordan block** J_k is a matrix of the form

$$J_k = \begin{pmatrix} \lambda_k & 1 & & & \\ & \lambda_k & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_k & 1 \\ & & & & \lambda_k \end{pmatrix}$$

with an eigenvalue λ_k of T along the diagonal. Up to changing the order of the J_ks in $M(T)$, this matrix is uniquely determined by T.

Example 1. If $\dim(V) = 3$ and $T \in \mathcal{L}(V)$ has repeated eigenvalue $\lambda = 5$ with multiplicity 3, there are three possibilities for the Jordan canonical form of T:

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \qquad \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

The first consists of three 1×1 Jordan blocks, the second consists of a 2×2 Jordan block and a 1×1 block, and the third consists of a single 3×3 Jordan block. You might expect $\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}$ to be a fourth possibility, but this is equivalent to the second matrix above, as it is just a rearrangement of the blocks.

Consider for a moment how a $k \times k$ Jordan block J acts with respect to the standard basis vectors $e_i \in \mathbb{C}^k$:

$$\begin{aligned} Je_1 &= \lambda e_1, \\ Je_2 &= \lambda e_2 + e_1, \\ &\vdots \; \vdots \\ Je_k &= \lambda e_k + e_{k-1}. \end{aligned}$$

Thus e_1 is a true eigenvector of J_k with eigenvalue λ, and we call $\{e_2, \ldots, e_k\}$ **generalized eigenvectors**, since they are not true eigenvectors but satisfy a similar equation, namely (1) below . The whole set $\{e_1, \ldots, e_k\}$ forms a **generalized eigenvector chain** of length k. Such a chain by definition starts with a true eigenvector (e_1 in this case) and then consists of generalized eigenvectors which satisfy the defining equation

$$Je_i = \lambda e_i + e_{i-1}. \tag{1}$$

Example 2. The matrix

$$A = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in Jordan canonical form and consists of a 1×1 block with eigenvalue 3, one 2×2 block and one 1×1 block both with eigenvalue 2, and a 2×2 block with eigenvalue 0. The true eigenvectors are e_1, e_2, e_4 and e_5 (the latter spans the 1 dimensional nullspace of A), and the rest are generalized eigenvectors. The eigenvector chains are

$$\{e_1\}, \ \{e_2, e_3\}, \ \{e_4\}, \text{ and } \{e_5, e_6\}.$$

We recall a few important facts:

(1) The nullspace of T is the subspace $\mathrm{Null}(T) = \{v \in V : Tv = 0\}$ and $v \in \mathrm{Null}(T)$ is equivalent to saying that v is an eigenvector of T with eigenvalue 0.
(2) For $T \in \mathcal{L}(V)$, the dimension $r = \dim \mathrm{Ran}(T)$ of the range and the dimension $k = \dim \mathrm{Null}(T)$ of the nullspace satisfy
$$k + r = n = \dim(V).$$
(3) T is invertible if and only if $\mathrm{Null}(T) = \{0\}$, for then $r = n$ and $k = 0$.

Theorem. *Let $T \in \mathcal{L}(V)$ where V is a complex finite dimensional vector space. Then there exists a basis $\{v_1, \ldots, v_n\}$ for V such that*

$$M(T, \{v_1, \ldots, v_n\}) = J \tag{2}$$

where J is a Jordan form matrix having the eigenvalues of T. Equivalently, there is a basis $\{v_1, \ldots, v_n\}$ such that

$$Tv_j = \lambda_j v_j, \quad \text{or} \quad Tv_j = \lambda_j v_j + v_{j-1}. \tag{3}$$

The matrix J is uniquely determined by T up to changing the order of the Jordan blocks J_i.

Proof. This proof is due to Fillipov, and proceeds by induction on $n = \dim(V)$. The case $n = 1$ is trivial since any 1×1 matrix is already in canonical form.

Thus suppose that the theorem has been proved for all operators on vector spaces of dimension strictly less than $n = \dim(V)$, and consider $T \in \mathcal{L}(V)$. We first suppose that T is not invertible, so that in particular $\dim \mathrm{Ran}(T) = r < n$.

Step 1. Consider the restriction of T to the space $\mathrm{Ran}(T)$. Since $\mathrm{Ran}(T)$ is an invariant subspace of dimension less than that of V, by the inductive hypothesis, there exists a basis $\{w_1, \ldots, w_r\}$ for $\mathrm{Ran}(T)$ such that

$$Tw_j = \lambda_j w_j, \quad \text{or} \quad Tw_j = \lambda_j w_j + w_{j-1}.$$

We arrange these into generalized eigenvector chains as described above.

Step 2. Let p be the dimension of the subspace $\mathrm{Null}(T) \cap \mathrm{Ran}(T)$. This means that there are p linearly independent vectors in $\mathrm{Ran}(T)$ which are also in $\mathrm{Null}(T)$, and are therefore true eigenvectors with eigenvalue 0. In particular, among the generalized eigenvector chains in the previous step, there are p chains which have $\lambda = 0$ and *start* with some true eigenvector (note that these p eigenvectors give a basis of $\mathrm{Null}(T) \cap \mathrm{Ran}(T)$). Now consider the *end* of such a chain, call it w. Since $w \in \mathrm{Ran}(T)$, there is some vector $y \in V$ such that

$$Ty = w = 0\,y + w. \tag{4}$$

We do this for each of the p chains and obtain vectors $y_1, \dots, y_p$. Note that each of these vectors forms the *new end* for a chain obtained in the previous step (since $\lambda = 0$, (4) is equivalent to the defining equation (1).)

Step 3. Now consider a complementary subspace U of $\mathrm{Null}(T) \cap \mathrm{Ran}(T)$ in $\mathrm{Null}(T)$, for instance by completing a basis for $\mathrm{Null}(T) \cap \mathrm{Ran}(T)$ to a basis for $\mathrm{Null}(T)$, and letting U be the span of those basis vectors are *not* also in $\mathrm{Ran}(T)$. This space has dimension $n - r - p$, and we denote the basis for U by $\{z_1, \dots, z_{n-r-p}\}$, whose elements must satisfy $Tz_j = 0$ since they are in the nullspace of T.

Now we claim that the set $w_1, \dots, w_r, y_1, \dots, y_p, z_1, \dots, z_{n-r-p}$ is independent. Indeed, suppose that

$$\sum_i a_i w_i + \sum_j b_j y_j + \sum_k c_k z_k = 0.$$

Applying T to both sides, we find that

$$\sum_i a_i \begin{bmatrix} \lambda_i w_i \\ \text{or} \\ \lambda_i w_i + w_{i-1} \end{bmatrix} + \sum_j b_j w_{i_j} = 0.$$

None of the w_{i_j}s appearing in the second sum can appear in the first sum, since they are the end of a Step 1 chain for which $\lambda_{i_j} = 0$. Thus we conclude that all the b_j must be 0. So we now have

$$\sum_i a_i w_i + \sum_k c_k z_k = 0.$$

But here the w_i are in the subspace $\mathrm{Ran}(T)$ and the z_k are explicitly *not* in the space $\mathrm{Ran}(T)$, and since they are separately independent it follows that $a_i = c_k = 0$ for all i, k, so that the whole set is independent.

Now we rename the vectors $w_1, \dots, w_r, y_1, \dots, y_p, z_1, \dots, z_{n-r-p}$ to $v_1, \dots, v_n$, reordering everything so that the vectors y_j appear at the end of the corresponding chain of w_i's where they belong. It follows that the set $v_1, \dots, v_n$ satisfies (3); equivalently, (2) holds where J is the corresponding Jordan form matrix.

To recap what we did: we started with the generalized eigenvector chains (the vectors w_i) lying in the space $\mathrm{Ran}(T)$ which were afforded to us by induction. We then appended a y_j to the end of each of those chains with eigenvalue 0, and then added additional length 1 chains of the z_k with eigenvalue 0. In particular, note that all the chains with nonzero eigenvalue are already obtained in Step 1, and that we are always 'growing' or adding chains with eigenvalue 0.

In the case that T is invertible, we consider instead $T' = (T - \lambda_0 I)$, where λ_0 is any eigenvalue of T. This must have nontrivial nullspace (since there is at least one eigenvector for λ_0), so the previous algorithm applies to give a basis $\{v_1, \dots, v_n\}$

such that $M(T', \{v_1, \dots, v_n\})$ is in canonical form. But then $M(T, \{v_1, \dots, v_n\})$ is likewise in canonical form since

$$M(T) = M(T') + \lambda_0 I.$$

In other words, the basis of generalized eigenvectors for T' is also a basis of generalized eigenvectors for T, and the Jordan form matrix for T consists of the Jordan form matrix for T', but with new eigenvalues $\lambda_i = \lambda_i' + \lambda_0$.

The clever trick here is that the algorithm requires us to be able to identify a particular eigenspace of T, namely the 0 eigenspace or nullspace. If this space is trivial, we shift some other eigenspace (for λ_0 in this case) into this role by subtracting a constant multiple of I, and the algorithm above works as before, obtaining eigenvector chains and 'growing' those with eigenvalue λ_0.

To prove uniqueness, we need to show that the data making up J, namely the eigenvalues and the number and size of the Jordan blocks, are completely determined by T, independent of any choices we have made for bases and so on. Obviously the eigenvalues are just the eigenvalues of T, so these are fixed. Also the *number* of Jordan blocks with a given eigenvalue λ must be given by $\dim \mathrm{Null}(T - \lambda I)$, since each block corresponds with a single independent eigenvector (the start of the corresponding chain).

That the sizes of the Jordan blocks (lengths of the generalized eigenvector chains) are invariantly determined is slightly more subtle. To see this, fix an eigenvalue λ and consider the operators $(T - \lambda I)^2$ and $T - \lambda I$. Observe that the nullspace of $(T - \lambda I)^2$ consists not only of the true eigenvectors with eigenvalue λ (i.e. the first generalized eigenvectors in any chains with eigenvalue λ), but also the *second* generalized eigenvectors in any chains with eigenvalue λ as well. Likewise $(T - \lambda I)^3$ kills not only the first and second but also the third generalized eigenvectors in any λ-chains, and so on. In particular, the dimensions of $\mathrm{Null}((T - \lambda I)^k)$ are (weakly) increasing as k increases.

We claim that the sizes of the Jordan blocks can be read off from the sequence of $\dim \mathrm{Null}((T - \lambda I)^k)$ as k increases. Indeed, the difference

$$\dim \mathrm{Null}((T - \lambda I)^2) - \dim \mathrm{Null}(T - \lambda I)$$

tells you how many chains with eigenvalue λ have length *at least* two (i.e. how many chains have second generalized eigenvectors), and in general the difference

$$\dim \mathrm{Null}((T - \lambda I)^k) - \dim \mathrm{Null}((T - \lambda I)^{k-1})$$

is the number of λ-chains having length at least k. By seeing how the number of chains of length at least k changes as k increases, you can determine the number of chains of length *equal to* k. Note in particular that the dimension stabilizes at some point ($k = n$ at most, since there can be no chains of length longer than n), so for some k_0,

$$\dim \mathrm{Null}((T - \lambda I)^{k_0+1}) - \dim \mathrm{Null}((T - \lambda I)^{k_0}) = 0$$

and therefore k_0 is the length of the longest λ-chains. $\square$

Let $T \in \mathcal{L}(V)$ with characteristic polynomial

$$p_T(z) = \det(zI - T).$$

The **algebraic multiplicity** of an eigenvalue λ_0 is the multiplicity of λ_0 as a root of $p_T(z)$. In other words, the algebraic multiplicity is the largest k such that $(z-\lambda_0)^k$ divides $p_T(z)$.

The **geometric multiplicity** of λ_0 is the maximal number of linearly independent eigenvectors with eigenvalue λ_0. Equivalently, it is the dimension of $\text{Null}(T-\lambda_0 I)$.

Corollary. *For any eigenvalue λ_0 of $T \in \mathcal{L}(V)$, the geometric multiplicity of λ_0 is less than or equal to the algebraic multiplicity.*

Proof. This follows immediately from the Jordan Canonical Form theorem. Indeed, the algebraic multiplicity is the number of generalized eigenvectors with eigenvalue λ_0 (since there is precisely one of these for each diagonal entry of J which is equal to λ_0 and the sum of these is the algebraic multiplicity), whereas the geometric multiplicity is the number of true eigenvectors. $\square$

Corollary. *$T \in \mathcal{L}(V)$ has a diagonal matrix with respect to some basis if and only if the geometric and algebraic multiplicities are equal for each of its eigenvalues.*