JORDAN CANONICAL FORM

We will show that every complex $n \times n$ matrix A is linearly conjugate to a matrix $J = T^{-1}AT$ which is in **Jordan canonical form:**

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}$$

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

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

with an eigenvalue λ of A along the diagonal.

Example 1. If a 3×3 matrix A has repeated eigenvalue $\lambda = 5$ with multiplicity 3, there are three possibilities for the Jordan canonical form of A:

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

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 conjugate to the second matrix above.

Let us 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 an eigenvector of J_k with eigenvalue λ, and we call $\{E_2, \ldots, E_k\}$ **generalized eigenvectors**, since they are not true eigenvectors but have a similar property. The whole set $\{E_1, \ldots, E_k\}$ forms a **generalized eigenvector chain** of length k, starting with a true eigenvector E_1 and ending with E_k.

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 kernel of A), and the rest are generalized eigenvectors. The chains consist of

$$\{E_1\}\,,\ \{E_2, E_3\}\,,\ \{E_4\}\,,\ \text{and}\ \{E_5, E_6\}\,.$$

We recall a few important facts:

(1) The kernel of A is the subspace $\operatorname{Ker}(A) = \{V \in \mathbb{C}^n : AV = 0\}$ and $V \in \operatorname{Ker}(A)$ is equivalent to saying that V is an eigenvector of A with eigenvalue 0

(2) For an $n \times n$ matrix, the dimension $r = \dim \operatorname{Ran}(A)$ of the range space and the dimension $k = \dim \operatorname{Ker}(A)$ of the kernel satisfy

$$k + r = n.$$

(3) A is invertible if and only if $\operatorname{Ker}(A) = \{0\}$, for then $r = n$ and $k = 0$.

Theorem. *Let A be an $n \times n$ complex matrix. Then there exists an invertible matrix T such that*

$$T^{-1}AT = J \tag{1}$$

where J is a Jordan form matrix having the eigenvalues of A. Equivalently, the columns of T consist of a set of independent vectors $V_1, \ldots, V_n$ such that

$$AV_j = \lambda_j V_j, \quad \textit{or} \quad AV_j = \lambda_j V_j + V_{j-1}. \tag{2}$$

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

Thus suppose that the theorem has been proved for $r \times r$ matrices for all $r < n$, and consider an $n \times n$ matrix A. We first suppose that A is not invertible, so that in particular $\dim \operatorname{Ran}(A) = r < n$.

Step 1. Consider the restriction of A to the space $\operatorname{Ran}(A)$. This is given by an $r \times r$ matrix (A must send $\operatorname{Ran}(A)$ into itself), so by the inductive hypothesis, there exists a set of linearly independent vectors $W_1, \ldots, W_r$ for $\operatorname{Ran}(A)$ such that

$$AW_j = \lambda_j W_j, \quad \text{or} \quad AW_j = \lambda_j W_j + W_{j-1}.$$

Step 2. Let p be the dimension of the subspace consisting of the intersection $\operatorname{Ker}(A) \cap \operatorname{Ran}(A)$. This means that there are p linearly independent vectors in $\operatorname{Ran}(A)$ which are also in $\operatorname{Ker}(A)$, and so have eigenvalue 0. In particular, among the generalized eigenvector chains of the W_i in the previous step, p of these must have $\lambda = 0$ and *start* with some true eigenvector. Now consider the *end* of such a chain, call it W. Since $W \in \operatorname{Ran}(A)$, there is some vector Y such that $AY = W$.

We do this for each of the p chains and obtain vectors $Y_1, \dots, Y_p$. Note that each of these vectors is the new end of the chain of W_is since the corresponding λ is 0.

Step 3. Now consider the subspace of $\mathrm{Ker}(A)$ spanned by nonzero vectors which are *not* also in $\mathrm{Ran}(A)$. This space has dimension $n-r-p$, and we can find independent vectors $Z_1, \dots, Z_{n-r-p}$ spanning this space, which must satisfy $AZ_j = 0$ since they are in the kernel of A.

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 A 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 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 $\mathrm{Ran}(A)$ and the Z_k are explicitly *not* in the space $\mathrm{Ran}(A)$, 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 so that the vectors Y_j come 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 (2), and that (1) holds where T is the matrix whose columns are the V_i.

To recap what we did: we started with the generalized eigenvector chains (the vectors W_i) lying in the space $\mathrm{Ran}(A)$ 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.

If A is invertible, we consider instead $A' = (A - \lambda_0 I)$, where λ_0 is any eigenvalue of A. This must have nontrivial kernel (since there is at least one eigenvector for λ_0), so the previous algorithm applies to give a matrix T such that $T^{-1}A'T = J'$ is in canonical form. We claim that T also conjugates A to Jordan canonical form:

$$T^{-1}AT = T^{-1}\big(A' + \lambda_0 I\big)T = T^{-1}A'T + \lambda_0 I = J' + \lambda_0 I = J.$$

Notice that J is also a Jordan matrix, having the same eigenvector chains as J' but with shifted eigenvalues $\lambda = \lambda' + \lambda_0$. (In general A and $A + cI$ have the same eigenvectors and generalized eigenvector chains, but their eigenvalues differ by c.)

The clever trick here is that the algorithm requires us to be able to identify a particular eigenspace of A, namely the 0 eigenspace or kernel. If this space is trivial, we shift some other eigenspace (for λ_0 in this case) into this role by adding a constant multiple of I, and the algorithm above works as before, obtaining eigenvector chains and 'growing' those with eigenvalue λ_0. $\square$

Corollary. *Let A be a real $n \times n$ matrix. Then there exists an invertible matrix T such that $T^{-1}AT = J$ has the form*

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}$$

where each block J_k is has one of two forms:

$$J_k = \begin{pmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{pmatrix}, \quad \textit{or} \quad \begin{pmatrix} B_j & I & & & \\ & B_j & I & & \\ & & \ddots & \ddots & \\ & & & B_j & I \\ & & & & B_j \end{pmatrix},$$

where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $B_j = \begin{pmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{pmatrix}$ for real eigenvalues λ_j and complex eigenvalues $\alpha_j \pm i\beta_j$ of A.

Proof. Considering A as a complex matrix, we obtain complex generalized eigenvectors $V_1, \dots, V_n$ from the previous theorem. If an eigenvector λ is real, then it follows by considering the complex conjugate of the equations (2) that the corresponding generalized eigenvectors can be taken to be real, replacing the V_j by $\frac{1}{2}\left(V_j + \overline{V}_j\right)$ if necessary. This results in Jordan blocks of the first type.

If $\lambda = \alpha + i\beta$ is complex, then $\overline{\lambda} = \alpha - i\beta$ must also be an eigenvector, and we may assume that the chains for λ and $\overline{\lambda}$ consist of complex conjugate vectors V_j and $\overline{V}_j$:

$$AV_j = \lambda V_j \ \left[+ V_{j-1}\right], \quad A\overline{V}_j = \overline{\lambda V}_j \ \left[+ \overline{V}_{j-1}\right].$$

Then, letting $W_{2j-1} = \mathrm{Re}\,(V_j) = \frac{1}{2}\left(V_j + \overline{V}_j\right)$ and $W_{2j} = \mathrm{Im}\,(V_j) = \frac{-i}{2}\left(V_j - \overline{V}_j\right)$, it follows that the W_j are independent and that

$$\begin{aligned} AW_{2j-1} &= (\alpha_j W_{2j-1} - \beta_j W_{2j}) \ \left[+ W_{2j-3}\right] \\ AW_{2j} &= (\beta_j W_{2j-1} + \alpha_j W_{2j}) \ \left[+ W_{2j-2}\right]. \end{aligned}$$

Letting T be the matrix whose columns are the real-valued V_i of the first paragraph and the W_j just constructed, we obtain the desired result. □

The Jordan canonical form of A is unique up to permutation of the Jordan blocks. Indeed, the λ_j are the eigenvalues of A, counted with multiplicity, so it suffices to show that two Jordan matrices with the same eigenvalues but different size Jordan blocks (such as the 3×3 matrices of Example 1) cannot be conjugate. This is left as an exercise.