Jordan Normal Form

§1. Jordan's Theorem

Definition The n by n matrix $J_{\lambda,n}$ with λ's on the diagonal, 1's on the superdiagonal and 0's elsewhere is called a *Jordan block* matrix. A *Jordan* matrix or matrix in *Jordan normal form* is a block matrix that is has Jordan blocks down its block diagonal and is zero elsewhere.

Theorem Every matrix over **C** is similar to a matrix in Jordan normal form, that is, for every A there is a P with $J = P^{-1}AP$ in Jordan normal form.

§2. Motivation for proof of Jordan's Theorem

Consider Jordan block $A = J_{\lambda,n}$, for example,

$$A = J_{5,3} = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

We see that

$$\begin{aligned} A\mathbf{e}_1 &= 5\mathbf{e}_1 \\ A\mathbf{e}_2 &= \mathbf{e}_1 + 5\mathbf{e}_2 \,. \\ A\mathbf{e}_3 &= \mathbf{e}_2 + 5\mathbf{e}_3 \end{aligned}$$

Writing $A_5 = A - 5I$ this becomes:

$$\begin{aligned} A_5\mathbf{e}_1 &= \mathbf{0} \\ A_5\mathbf{e}_2 &= \mathbf{e}_1 \,. \\ A_5\mathbf{e}_3 &= \mathbf{e}_2 \end{aligned}$$

which can be conveniently rewritten as a *string* of length 3 with value 5:

$$\mathbf{e}_3 \xrightarrow{A_5} \mathbf{e}_2 \xrightarrow{A_5} \mathbf{e}_1 \xrightarrow{A_5} \mathbf{0}$$

Since $A_5\mathbf{e}_1 = \mathbf{0}$, $\mathbf{e}_1$ is an eigenvector with value 5. $(A_5)^2\mathbf{e}_2 = \mathbf{0}$ and $(A_5)^3\mathbf{e}_3 = \mathbf{0}$ and so $\mathbf{e}_2$ and $\mathbf{e}_3$ are called generalized eigenvectors. Although there is no basis of eigenvectors, there is a basis of generalized eigenvectors.

Definition Define $A_\lambda = A - \lambda I$. Call $\mathbf{v} \neq \mathbf{0}$ a *generalized eigenvector* with value λ for A if $(A_\lambda)^p\mathbf{v} = \mathbf{0}$ for some natural p. If $p = 1$, $\mathbf{v}$ is called an *eigenvector*.

§3. Proof of Jordan's Theorem

Introduction to the proof Although there is no basis of eigenvectors, we show there is a basis of generalized eigenvectors. More specifically we find a collection of strings:

$$\begin{array}{ccccccc} \mathbf{w}_{1,n_1} & \xrightarrow{A_{\lambda_1}} & \cdots & \xrightarrow{A_{\lambda_1}} & \mathbf{w}_{1,1} & \xrightarrow{A_{\lambda_1}} & \mathbf{0} \\ & & \vdots & & \vdots & & \vdots \\ \mathbf{w}_{k,n_k} & \xrightarrow{A_{\lambda_k}} & \cdots & \xrightarrow{A_{\lambda_k}} & \mathbf{w}_{k,1} & \xrightarrow{A_{\lambda_k}} & \mathbf{0} \end{array}$$

such that the $\mathbf{w}_{i,j}$'s form a basis. With respect to this basis the matrix of A is in Jordan normal form because the i-th string generates a Jordan block J_{λ_i,n_i}, and conversely a Jordan matrix generates a collection of strings of basis vectors. Accordingly we concern ourselves with generating strings of basis vectors.
The proof we give is due to Filippov (see *Linear Algebra and Its Applications* by G. Strang).

Proof Let A be n by n. The case $n = 1$ is trivial. By "strong" induction, assume every smaller size matrix can be put in Jordan normal form, which by the comments above, amounts to the existence of strings.

A has an eigenvector $\mathbf{v}$ with value λ. Since $A_\lambda\mathbf{v} = \mathbf{0}$, we have $r \overset{\text{def}}{=} \dim \operatorname{Ker} A_\lambda > 0$. By the Rank+Nullity Theorem (or directly, since the row reduced form of A_λ has r free variables there must be $n - r$ pivots) we have $\dim \operatorname{Range} A_\lambda = n - r < n$. Call $W = \operatorname{Range} A_\lambda$.

Step 1 $A_\lambda(W) \subseteq W$ so A_λ induces a transformation $T\colon W \to W$. Since $\dim(W) < n$, the matrix of T is of smaller size than n so by induction there are strings:

$$\begin{array}{ccccccc}
\mathbf{w}_{1,n_1} & \xrightarrow{A_{\lambda_1}} & \dots & \xrightarrow{A_{\lambda_1}} & \mathbf{w}_{1,1} & \xrightarrow{A_{\lambda_1}} & \mathbf{0} \\
 & & \vdots & & \vdots & & \vdots \\
\mathbf{w}_{k,n_k} & \xrightarrow{A_{\lambda_k}} & \dots & \xrightarrow{A_{\lambda_k}} & \mathbf{w}_{k,1} & \xrightarrow{A_{\lambda_k}} & \mathbf{0}
\end{array}$$

where the $\mathbf{w}_{i,j}$'s form a basis for W — here we used the fact that $(A_\lambda)_{\mu_i} = A_{\lambda+\mu_i} \stackrel{\text{def}}{=} A_{\lambda_i}$.
Step 2 Let $q = \dim(W \cap \operatorname{Ker} A_\lambda)$. Since $\mathbf{w}_{j,1} \in \operatorname{Ker} A_{\lambda_j}$, q of the above strings are A_λ strings, say the first q: $\lambda_j = \lambda$ for $1 \leq j \leq q$. At the other end of these strings, $\mathbf{w}_{j,n_j} \in W = \operatorname{Range} A_\lambda$ so there are $\mathbf{y}_j$ with $\mathbf{y}_j \xrightarrow{A_\lambda} \mathbf{w}_{j,n_j}$ for $1 \leq j \leq q$.
Step 3 Since $\operatorname{Ker} A_\lambda$ is r dimensional and meets W on a q dimensional subspace, some $r-q$ dimensional subspace Z of $\operatorname{Ker} A_\lambda$ meets W only at $\mathbf{0}$. Let $\mathbf{z}_1, \dots, \mathbf{z}_{r-q}$ be a basis for Z.
This gives $q + (n-r) + (r-q) = n$ vectors in strings:

$$\begin{array}{ccccccccc}
\mathbf{y}_1 & \xrightarrow{A_\lambda} & \mathbf{w}_{1,n_1} & \xrightarrow{A_\lambda} & \dots & \xrightarrow{A_\lambda} & \mathbf{w}_{1,1} & \xrightarrow{A_\lambda} & \mathbf{0} \\
 & & & & \vdots & & \vdots & & \vdots \\
\mathbf{y}_q & \xrightarrow{A_\lambda} & \mathbf{w}_{q,n_q} & \xrightarrow{A_\lambda} & \dots & \xrightarrow{A_\lambda} & \mathbf{w}_{q,1} & \xrightarrow{A_\lambda} & \mathbf{0} \\
 & & \mathbf{w}_{q+1,n_{q+1}} & \xrightarrow{A_{\lambda_{q+1}}} & \dots & \xrightarrow{A_{\lambda_{q+1}}} & \mathbf{w}_{q+1,1} & \xrightarrow{A_{\lambda_{q+1}}} & \mathbf{0} \\
 & & & & \vdots & & \vdots & & \vdots \\
 & & \mathbf{w}_{k,n_k} & \xrightarrow{A_{\lambda_k}} & \dots & \xrightarrow{A_{\lambda_k}} & \mathbf{w}_{k,1} & \xrightarrow{A_{\lambda_k}} & \mathbf{0} \\
 & & & & & & \mathbf{z}_1 & \xrightarrow{A_\lambda} & \mathbf{0} \\
 & & & & & & \vdots & & \vdots \\
 & & & & & & \mathbf{z}_{r-q} & \xrightarrow{A_\lambda} & \mathbf{0}
\end{array}$$

It suffices to show they are linearly independent, so assume

$$\sum_i a_i \mathbf{y}_i + \sum_{i,j} b_{ij} \mathbf{w}_{i,j} + \sum_i c_i \mathbf{z}_i = \mathbf{0}.$$

Applying A_λ gives a linear combination, L, in $\mathbf{w}_{i,j}$'s as one can see by referring to the strings above. Using $A_{\lambda_r} \mathbf{w}_{s,r} = \mathbf{w}_{s,r-1}$ together with $A_\lambda = A_{\lambda_r} + (\lambda_r - \lambda)I$ shows $A_\lambda \mathbf{w}_{s,r} = \mathbf{w}_{s,r-1} + (\lambda_r - \lambda)\mathbf{w}_{s,r}$, hence the coefficient of the $\mathbf{w}_{j,n_j}$ for $1 \leq j \leq q$ in linear combination L is a_j. By linear independence of the $\mathbf{w}_{i,j}$'s we obtain $a_j = 0$. So

$$\sum_{i,j} b_{ij} \mathbf{w}_{i,j} + \sum_i c_i \mathbf{z}_i = \mathbf{0}.$$

But $\sum_{i,j} b_{ij} \mathbf{w}_{i,j} = \mathbf{0}$ and $\sum_i c_i \mathbf{z}_i = \mathbf{0}$ since W and Z meet only at $\mathbf{0}$. By linear independence in W and Z, $b_{ij} = 0$ and $c_i = 0$. ▮