12. Jordan Canonical Form

12.1 Definition: Let $\mathbb{F}$ be a field. For $m \in \mathbb{Z}^+$ and $\lambda \in \mathbb{F}$, we define the $m \times m$ **Jordan block** for the eigenvalue λ to be the $m \times m$ matrix

$$J_\lambda^m = \begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & 1 & & \\ & & & \ddots & \ddots & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}.$$

For $n \in \mathbb{Z}^+$, a matrix $A \in M_{n \times n}(\mathbb{F})$ is said to be in **Jordan form** when it is in the block-diagonal form

$$A = \begin{pmatrix} J_{\lambda_1}^{m_1} & & & \\ & J_{\lambda_2}^{m_2} & & \\ & & \ddots & \\ & & & J_{\lambda_l}^{m_l} \end{pmatrix}$$

for some $l, m_i \in \mathbb{Z}^+$ and $\lambda_i \in \mathbb{F}$.

12.2 Note: Our goal in this chapter is to prove that for every linear map $L : U \to U$ on a finite dimensional vector space U over a field $\mathbb{F}$, if $f_L(x)$ splits then there exists an ordered basis $\mathcal{A}$ for U such that the matrix $[L]_\mathcal{A}$ is in Jordan form, and that this Jordan form matrix is unique up to the order of the Jordan blocks.

Recall that when $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ is an ordered basis for U, the matrix $[L]_\mathcal{A}$ is given by the formula

$$[L]_\mathcal{A} = \big([Lu_1]_\mathcal{A}, [Lu_2]_\mathcal{A}, \cdots, [Lu_n]_\mathcal{A}\big) \in M_n(\mathbb{F}).$$

If follows immediately from this formula that when

$$\mathcal{A} = \big\{u_{11}, u_{12}, \cdots, u_{1\,m_1}, u_{21}, u_{22}, \cdots u_{2\,m_2}, \cdots, u_{l1}, u_{l2}, \cdots, u_{l\,m_l}\big\}$$

is an ordered basis for U, the matrix $[L]_\mathcal{A}$ is of the required Jordan form with blocks $J_{\lambda_i}^{m_i}$ precisely when for each index i with $1 \leq i \leq l$, we have

$$Lu_{i1} = \lambda_i u_{i1}\ ,\ Lu_{i2} = u_{i1} + \lambda_i u_{i2},\ ,\ Lu_{i3} = u_{i2} + \lambda_i u_{i2}\ ,\ \cdots\ ,\ Lu_{i\,m_i} = u_{i\,m_i-1} + \lambda_i u_{i\,m_i}.$$

We can also write the above equations as

$$(L - \lambda_i I)u_{i1} = 0\ , (L - \lambda_i I)u_{i2} = u_{i1}\ ,\ (L - \lambda I)u_{i3} = u_{i2}\ ,\ \cdots\ ,\ (L - \lambda_i I)u_{i\,m_i} = u_{i\,m_{i-1}}.$$

Notice that when these equations hold, we have

$$0 = (L - \lambda I)u_{i,1} = (L - \lambda I)^2 u_{i,2} = (L - \lambda I)^3 u_{i,3} = \cdots = (L - \lambda I)^{m_i} u_{i,m_i}.$$

These considerations lead us to make the following definitions.

12.3 Definition: Let $L : U \to U$ where U is a finite dimensional vector space over a field $\mathbb{F}$, and let $\lambda \in \text{Spec}\,(L)$. The **generalized eigenspace** of L for λ is the vector space

$$K_\lambda = K_\lambda(L) = \left\{ u \in U \middle| (L - \lambda I)^k u = 0 \text{ for some } k \in \mathbb{Z}^+ \right\}.$$

A **cycle** of generalized eigenvectors for λ is an ordered m-tuple $(u_1, u_2, \cdots, u_m)$ with each $u_i \in U$ such that

$$(L - \lambda I)u_1 = 0 \;,\; (L - \lambda I)u_2 = u_1 \;,\; (L - \lambda I)u_3 = u_2 \;,\; \cdots \;,\; (L - \lambda I)u_m = u_{m-1}.$$

Note that for each index k we have $(L - \lambda I)^k u_k = 0$ so that $u_k \in K_\lambda$.

12.4 Note: The discussion in Note 12.2 shows that for an ordered basis $\mathcal{A}$ for U, the matrix $[L]_\mathcal{A}$ is in Jordan form if and only if $\mathcal{A}$ is an ordered union of cycles of generalized eigenvectors (for various eigenvalues).

12.5 Definition: Let $L : U \to U$ be a linear map on a vector space U over a field $\mathbb{F}$. Let $V \subseteq U$ be a subspace. We say that V is an **invariant subspace** of U under L when $L(V) \subseteq V$. Note that when V is invariant under L, the restriction of L to V gives a linear map $L : V \to V$.

12.6 Theorem: *Let $L : U \to U$ be a linear map on a finite dimensional vector space U over a field $\mathbb{F}$. Suppose that $f_L(x)$ splits. Let $\lambda \in \text{Spec}\,(L)$ and let $\mu \in \mathbb{F}$. Then*
(1) If $M : U \to U$ is linear and commutes with L, then K_λ is invariant under M. In particular, K_λ is invariant under L, and under $L - \mu I$.
(2) When $\mu \neq \lambda$ the map $(L - \lambda I) : K_\mu \to K_\mu$ is an isomorphism.
(3) We have $\dim(K_\lambda) \leq m_\lambda$ where m_λ is the algebraic multiplicity of λ.

Proof: Let $M : U \to U$ be any map which commutes with L and let $u \in K_\lambda$, say $(L - \lambda I)u^k = 0$. Snce M commutes with L, M also commutes with $L = \lambda I$ and so we have

$$(L - \lambda I)^k (Mu) = M(L - \lambda I)^k u = M(0) = 0$$

and so $M(u) \in K_\lambda$. This proves Part (1).

To prove Part (2), suppose that $\mu \neq \lambda$ and suppose, for a contradiction, that the map $(L - \lambda I) : K_\mu \to K_\mu$ is not an isomorphism. Choose $0 \neq u \in K_\mu$ such that $(L - \lambda I)u = 0$. Choose $k \in \mathbf{N}$ so that $(L - \mu I)^k u \neq 0$ but $(L - \mu I)^{k+1} u = 0$ and let $v = (L - \mu I)^k u$. Note that $v \neq 0$ and $v \in E_\mu$. Since $(L - \lambda I)$ commutes with $(L - \mu I)$ and $(L - \lambda I)u = 0$ we have

$$(L - \lambda I)v = (L - \lambda I)(L - \mu I)^k u = (L - \mu I)^k (L - \lambda I)u = (L - \mu I)^k (0) = 0$$

so that $v \in E_\lambda$. Thus $v \neq 0$ and $v \in E_\lambda \cap E_\mu$. But this is not possible since $\mu \neq \lambda$ so that $E_\lambda \cap E_\mu = \{0\}$.

By Part (1), K_λ is an invariant subspace so we can let $M : K_\lambda \to K_\lambda$ be the restriction of L. Note that $\lambda \in \text{Spec}\,(M)$ and, since $f_M(x)$ divides $f_L(x)$ (as we can see by choosing any basis for K_λ, extending it to a basis for U, and considering the associated matrices for L and M), it follows that $f_M(x)$ splits and $\text{Spec}\,(M) \subseteq \text{Spec}\,(L)$. By Part (2), when $\mu \neq \lambda$ the map $(L - \mu I) : K_\lambda \to K_\lambda$ is an isomorphism, so that μ is not an eigenvalue of M, and so we have $\text{Spec}\,(M) = \{\lambda\}$. Thus $f_M(x) = \pm(x - \lambda)^d$ where $d = \dim(K_\lambda)$ and $d \leq m_\lambda$.

12.7 Theorem: *Let $L : U \to U$ be a linear map on a finite dimensional vector space U over a field $\mathbb{F}$. Suppose that $f_L(x)$ splits. Then*

$$U = \bigoplus_{\lambda \in \mathrm{Spec}\,(L)} K_\lambda(L).$$

Proof: We prove by induction on the number of distinct eigenvalues of L. When L has only one eigenvalue λ, by Schur's Theorem (or the Cayley-Hamilton Theorem) we have $(L - \lambda I)^d = 0$ where $d = \dim U$ so $U = \mathrm{Null}(L - \lambda I)^d = K_\lambda(L)$.

Suppose that L has at least 2 distinct eigenvalues and suppose the theorem holds for any linear map M with fewer eigenvalues that L. Let $\lambda \in \mathrm{Spec}\,(L)$. Since U is finite dimensional, we can choose $p \in \mathbb{Z}$ with $1 \le p \le d = \dim U$ such that

$$U = \mathrm{Range}(L - \lambda I)^0 \supsetneqq \mathrm{Range}(L - \lambda I)^1 \supsetneqq \cdots \supsetneqq \mathrm{Range}(L - \lambda I)^p = \mathrm{Range}(L - \lambda I)^{p+1}.$$

Note that since we have $\mathrm{Range}(L - \lambda I)^p = \mathrm{Range}(L - \lambda I)^{p+1}$, it follows that the map $(L - \lambda I) : \mathrm{Range}(L - \lambda I)^p \to \mathrm{Range}(L - \lambda I)^{p+1}$ is surjective hence isomorphic, and so we have $\mathrm{Range}(L - \lambda I)^p = \mathrm{Range}(L - \lambda I)^k$ for all $k \ge p$. It follows that we also have

$$\{0\} = \mathrm{Null}(L - \lambda I)^0 \subsetneqq \mathrm{Null}(L - \lambda I)^1 \subsetneqq \cdots \subsetneqq \mathrm{Null}(L - \lambda I)^p = \mathrm{Null}(L - \lambda I)^k$$

for all $k \ge p$, so $K_\lambda(L) = \left\{ u \in U \,\middle|\, (L - \lambda I)^k u = 0 \text{ for some } k \in \mathbb{Z}^+ \right\} = \mathrm{Null}(L - \lambda I)^p$. Let

$$V = \mathrm{Range}(L - \lambda I)^p.$$

Since $L - \lambda I : V \to V$, we also have $L : V \to V$. Let $M : V \to V$ be the restriction of L to V. Since $f_M(x)$ divides $f_L(x)$, it follows that $f_M(x)$ splits and that $\mathrm{Spec}\,(M) \subseteq \mathrm{Spec}\,(L)$. Since $M - \lambda I = L - \lambda I : V \to V$ is an isomorphism, λ is not an eigenvalue of M so we have $\mathrm{Spec}\,(M) \subseteq \mathrm{Spec}\,(L) \setminus \{\lambda\}$. By the induction hypothesis, we have

$$V = \bigoplus_{\nu \in \mathrm{Spec}\,(M)} K_\nu(M).$$

Let $\mu \in \mathrm{Spec}\,(L)$ with $\mu \ne \lambda$. We have

$$\begin{aligned} u \in K_\mu(M) &\iff u \in V \text{ and } (M - \mu I)^k u = 0 \text{ for some } k \in \mathbb{Z}^+ \\ &\iff u \in V \text{ and } (L - \mu I)^k u = 0 \text{ for some } k \in \mathbb{Z}^+ \\ &\iff u \in V \text{ and } u \in K_\mu(L) \end{aligned}$$

so $K_\mu(M) = K_\mu(L) \cap V$. Since $(L - \lambda I) : K_\mu(L) \to K_\mu(L)$ is an isomorphism, it follows that $K_\mu(L) \subseteq \mathrm{Range}(L - \lambda I)^p = V$ and so $K_\mu(M) = K_\mu(L) \cap V = K_\mu(L)$, as claimed. Thus we have

$$V = \bigoplus_{\mu \in \mathrm{Spec}\,(M)} K_\mu(M) = \bigoplus_{\mu \in \mathrm{Spec}\,(L) \setminus \{\lambda\}} K_\mu(L).$$

To prove $U = \bigoplus\limits_{\lambda \in \mathrm{Spec}\, L} K_\lambda(L)$, we show that $U = \sum\limits_{\lambda \in \mathrm{Spec}\,(L)} K_\lambda(L)$ and $\sum\limits_{\lambda \in \mathrm{Spec}\,(L)} \dim(K_\lambda(L)) = d$ where $d = \dim U$. Given $x \in U$, let $y = (L - \lambda I)^p x \in V$. For each $\mu \in \mathrm{Spec}\,(L) \setminus \{\lambda\}$, choose $v_\mu \in K_\mu(L)$ so that $y = \sum v_\mu$. Since $(L - \lambda I)$, hence also $(L - \lambda I)^p$, is an automorphism from $K_\mu(L)$, we can choose $u_\mu \in K_\mu(L)$ so that $(L - \lambda I)^p u_\mu = v_\mu$. Then we have $(L - \lambda I)^p (x - \sum u_\mu) = y - \sum v_\mu = 0$ so we can choose $u_\lambda = x - \sum\limits_{\mu \ne \lambda} u_\mu \in K_\lambda(L)$ to get $x = u_\lambda + \sum u_\mu$. This proves that $U = \sum\limits_{\lambda \in \mathrm{Spec}\,(L)} K_\lambda(L)$. Finally, we note that since $\dim\big(K_\lambda(L)\big) \le m_\lambda$ we have $\sum\limits_{\lambda \in \mathrm{Spec}\,(L)} \dim\big(K_\lambda(L)\big) \le \sum\limits_{\lambda \in \mathrm{Spec}\,(L)} m_\lambda = \dim U$.

12.8 Theorem: *Let $L : U \to U$ be a linear map on a finite dimensional vector space U over a field $\mathbb{F}$ and let $\lambda \in \mathrm{Spec}\,(L)$. Then K_λ has an ordered basis which is an ordered union of cycles of generalized eigenvectors.*

Proof: Consider the restriction of $L - \lambda I$ to K_λ. Choose $m \in \mathbb{Z}^+$ so that

$K_\lambda = \mathrm{Range}(L - \lambda I)^0 \supsetneqq \mathrm{Range}(L - \lambda I)^1 \supsetneqq \cdots \supsetneqq \mathrm{Range}(L - \lambda I)^m = \mathrm{Range}(L - \lambda I)^{m+1}.$

Note that $\mathrm{Range}(L-\lambda I)^m = \{0\}$ since $K_\lambda = \mathrm{Null}(L-\lambda I)^m$, and $\mathrm{Range}(L-\lambda I)^{m-1} \subseteq E_\lambda$ because if $u \in \mathrm{Range}(L - \lambda I)^{m-1}$ then $(L - \lambda I)u \in \mathrm{Range}(L - \lambda I)^m = \{0\}$.

We describe a recursive procedure for constructing an ordered union of cycles which is an ordered basis for K_λ in which, at the k^{th} step, we obtain a basis for $\mathrm{Range}(L-\lambda I)^{m-k}$. We begin with the empty set, which is a basis for $\{0\} = \mathrm{Range}(L - \lambda I)^m$. At the 1^{st} step we choose a basis $\{u_{1,1}, u_{2,1}, \cdots, u_{n,1}\}$ for $\mathrm{Range}(L - \lambda I)^{m-1} \subseteq E_\lambda$. Suppose, inductively, that after the k^{th} step we have constructed a basis

$$\mathcal{A} = \{u_{1,1}, u_{1,2}, \cdots, u_{1,\ell_1}, u_{2,1}, \cdots, u_{2,\ell_2}, \cdots, u_{r,1}, \cdots, u_{r,\ell_r}\}$$

for $\mathrm{Range}(L - \lambda I)^{m-k} \subseteq K_\lambda$, where $\{u_{1,1}, \cdots, u_{r,1}\}$ is a basis for $\mathrm{Range}(L - \lambda I)^{m-k} \cap E_\lambda$ and $u_{i,j} = (L - \lambda I)u_{i,j+1}$ for all $1 \le i \le r$ and $1 \le j < \ell_r$. At the $(k + 1)^{\text{st}}$ step, for each index i with $1 \le i \le r$ we choose $u_{i,\ell_i+1} \in K_\lambda$ such that $(L - \lambda I)u_{i,\ell_i+1} = u_{i,\ell_i}$ and, in addition, we extend the basis $\{u_{1,1}, \cdots, u_{r,1}\}$ for $\mathrm{Range}(L - \lambda I)^{m-k} \cap E_\lambda$ to obtain a basis $\{u_{1,1,}, \cdots, u_{r,1}, u_{r+1,1} \cdots, u_{s,1}\}$ for $\mathrm{Range}(L - \lambda I)^{m-k-1} \cap E_\lambda$. We then extend the ordered basis $\mathcal{A}$ to the ordered set

$$\mathcal{B} = \{u_{1,1,}, \cdots, u_{1,\ell_1}, u_{1,\ell_1+1}, u_{2,1,}, \cdots, u_{2,\ell_2+1}, \cdots, u_{r,1}, \cdots, u_{r,\ell_r+1}\ ,\ u_{r+1,1}, \cdots, u_{s,1}\}.$$

It remains to prove that $\mathcal{B}$ is a basis for $\mathrm{Range}(L - \lambda I)^{m-k-1}$.

Consider the map $M = (L - \lambda I) : \mathrm{Range}(L - \lambda I)^{m-k-1} \to \mathrm{Range}(L - \lambda I)^{m-k}$. This map is surjective so that $\mathrm{rank}(M) = \dim\big(\mathrm{Range}(L - \lambda I)^{m-k}\big) = |\mathcal{A}|$, and we have $\mathrm{Null}(M) = \mathrm{Range}(L - \lambda I)^{m-k-1} \cap E_\lambda$ so that $\mathrm{nullity}(M) = s$, and it follows that

$$\dim\big(\mathrm{Range}(L - \lambda I)^{m-k-1}\big) = \mathrm{rank}(M) + \mathrm{nullity}(M) = |\mathcal{A}| + s.$$

Now consider the map $N = (L-\lambda I) : \mathrm{Span}(\mathcal{B}) \subseteq \mathrm{Range}(L-\lambda I)^{m-k-1} \to \mathrm{Range}(L-\lambda I)^{m-k}$. Since $(L - \lambda I)(u_{i,j+1}) = u_{i,j}$ it follows that N is surjective so that $\mathrm{rank}(N) = |\mathcal{A}|$, and since $(L - \lambda I)(u_{i,1}) = 0$ it follows that $\mathrm{Range}(L - \lambda I)^{m-k-1} \cap E_\lambda \subseteq \mathrm{Null}(N)$ so that $\mathrm{nullity}(N) \ge s$. Since $\mathcal{B}$ is obtained from $\mathcal{A}$ by adding exactly s elements, it follows that

$$|\mathcal{A}| + s = |\mathcal{B}| \ge \dim\big(\mathrm{Span}\,(\mathcal{B})\big) = \mathrm{rank}(N) + \mathrm{nullity}(N) \ge |\mathcal{A}| + s$$

so that $|\mathcal{B}| = \dim\big(\mathrm{Span}\,(\mathcal{B})\big) = \dim\big(\mathrm{Range}(L - \lambda I)^{m-k-1}\big)$. Thus $\mathcal{B}$ is a basis for $\mathrm{Range}(L - \lambda I)^{m-k-1}\big)$, as required.

12.9 Theorem: (*Jordan Form*) *Let $L : U \to U$ be a linear map on a finite dimensional vector space U over a field $\mathbb{F}$. Suppose that $f_L(x)$ splits over $\mathbb{F}$. Then there exists an ordered basis $\mathcal{A}$ for U such that*

$$[L]_\mathcal{A} = \begin{pmatrix} J_{\lambda_1}^{m_1} & & & \\ & J_{\lambda_2}^{m_2} & & \\ & & \ddots & \\ & & & J_{\lambda_l}^{m_l} \end{pmatrix}$$

for some $l, m_i \in \mathbb{Z}^+$ and some $\lambda_i \in \mathbb{F}$, with the Jordan blocks $J_{\lambda_i}^{m_i}$ uniquely determined (up to order).

Proof: The existence of the ordered basis which puts the map L into Jordan form follows from the previous two theorems. It remains to show that the Jordan blocks are uniquely determined (up to order). Note that for $m \in \mathbb{Z}^+$ and $\lambda \in \mathbb{F}$ we have

$$\left(J_\lambda^m - \lambda I\right) = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix}, \quad \left(J_\lambda^m - \lambda I\right)^2 = \begin{pmatrix} 0 & 0 & 1 & & & \\ & 0 & 0 & 1 & & \\ & & & \ddots & \ddots & \\ & & & \ddots & \ddots & 1 \\ & & & & & 0 \\ & & & & & 0 \end{pmatrix}$$

and so on until

$$\cdots, \quad \left(J_\lambda^m - \lambda I\right)^{m-1} = \begin{pmatrix} & 0 & 1 \\ & & 0 \\ & & \end{pmatrix}, \quad \left(J_\lambda^m - \lambda I\right)^m = 0\,.$$

It follows that for $0 \leq k \leq m$ we have

$$\operatorname{rank}\left(J_\lambda^m - \lambda I\right)^k = m - k.$$

Also notice that for $\mu \in \mathbb{F}$ with $\mu \neq \lambda$ we have

$$\left(J_\lambda^m - \mu I\right) = \begin{pmatrix} \lambda-\mu & 1 & & & \\ & \lambda-\mu & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda-\mu \end{pmatrix}$$

and so

$$\operatorname{rank}\left(J_\lambda^m - \mu I\right) = m\,.$$

Now suppose that there exists a basis $\mathcal{A}$ as stated in the theorem, so that $A = [L]_\mathcal{A}$ is in Jordan form with Jordan blocks $J_{\lambda_i}^{m_i}$. The eigenvalues of L are the same as the eigenvalues of A, namely $\lambda_1, \cdots, \lambda_l$. For indices i such that $\lambda_i \neq \lambda$, we have $\operatorname{rank}(J_{\lambda_i}^{m_i} - \lambda I)^k = m_i$. For indices i with $\lambda_i = \lambda$ and $m_i \geq k$, we have $\operatorname{rank}(J_{\lambda_i}^{m_i} - \lambda I)^k = m_i - k$. For indices i such that $\lambda_i = \lambda$ and $m_i < k$, we have $\operatorname{rank}(J_{\lambda_i}^{m_i} - \lambda I)^k = 0$. Let a_k be the number of indices i such that $\lambda_i = \lambda$ and $m_i = k$, and let b_k be the number of indices i such that $\lambda_i = \lambda$ and $m_i \geq k$. Then since $\left[(L - \lambda_i I)^k\right]_\mathcal{A} = (A - \lambda I)^k$, which is the block diagonal matrix with blocks $(J_{\lambda_i}^{m_i} - \lambda I)^k$, we see that

$$\begin{aligned} \operatorname{rank}(L - \lambda I)^k &= n - \left(1\,a_1 + 2\,a_2 + 3\,a_3 + \cdots + (k-1)a_{k-1} + k\,b_k\right) \\ &= n - \left(b_1 + b_2 + b_3 + \cdots + b_k\right) \end{aligned}$$

and so we have

$$b_k = \operatorname{rank}(L - \lambda I)^{k-1} - \operatorname{rank}(L - \lambda I)^k$$

and hence

$$a_k = b_k - b_{k+1} = \operatorname{rank}(L - \lambda I)^{k-1} - 2\operatorname{rank}(L - \lambda I)^k + \operatorname{rank}(L - \lambda I)^{k+1}\,.$$

This formula shows that the blocks $J_{\lambda_i}^{m_i}$ are uniquely determined (up to order).