Jordan Form

Since we compute with coordinates and coordinate matrices, when dealing with a linear transformation, it is natural to ask, "What is the nicest/simplest possible coordinate matrix representing our transformation?" If we are allowed to specify both the domain and codomain basis, the answer is quite simple.

Proposition: Let $T : V \to W$ be a linear transformation between two finite dimensional vector spaces over some field. Then there is a basis α for V and basis β for W such that $[T]_\alpha^\beta = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix}$ where the zeros are appropriately sized zero matrices and I is an $r \times r$ identity matrix where $r = \text{rank}(T)$.

Proof: Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for $\ker(T)$. This is a linearly independent subset of V so we can extend it to a basis for V: $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1, \dots, \mathbf{u}_r\}$. We have that $T(V) = T(\text{span}\,\alpha) = \text{span}\,T(\alpha) = \text{span}\,\{T(\mathbf{v}_1)\dots, T(\mathbf{v}_n), T(\mathbf{u}_1), \dots, T(\mathbf{u}_r)\} = \text{span}\,\{T(\mathbf{u}_1), \dots, T(\mathbf{u}_r)\}$ since $T(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \dots, n$. Suppose $c_1T(\mathbf{u}_1) + \cdots + c_rT(\mathbf{u}_r) = \mathbf{0}$ so that $T(c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r) = \mathbf{0}$. Thus $c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r \in \ker(T)$. This means that $c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ for some scalars $a_1, \dots, a_n$. But then $c_1\mathbf{u}_1 + \cdots + c_r\mathbf{u}_r - a_1\mathbf{v}_1 - \cdots - a_n\mathbf{v}_n = \mathbf{0}$ and so $c_1 = \cdots = c_r = -a_1 = \cdots = -a_n = 0$ since $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1, \dots, \mathbf{u}_r\}$ is linearly independent. Therefore, $\{T(\mathbf{u}_1), \dots, T(\mathbf{u}_r)\}$ is linearly indpendent. This is a basis for the range of T so that $r = \text{rank}(T)$. We extend this linearly independent subset of the codomain W to a basis for W, say $\beta = \{T(\mathbf{u}_1), \dots, T(\mathbf{u}_r), \mathbf{w}_1, \dots, \mathbf{w}_m\}$. Then $T(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \dots, n$ implies that the first n columns of $[T]_\alpha^\beta$ are filled with zeros. Evaluating T on the rest of α (i.e., plugging in $\mathbf{u}_j$ for $j = 1, \dots, r$) we have $T(\mathbf{u}_j)$. Since this is the j-th element in of β, the $(n + j)$-th column of $[T]_\alpha^\beta$ has a 1 in the j-th row and zeros elsewhere. Thus $[T]_\alpha^\beta$ has the form we promised. ■

This above result is nice enough, but in the case that T is a linear operator (i.e., $T : V \to V$) it seems strange to have different bases for the domain ($= V$) and codomain (also $= V$). Thus we ask, "What is a nicest/simplest possible coordinate matrix $[T]_\beta^\beta$ for a linear operator $T : V \to V$?" This is a much more difficult problem. In fact, our field starts playing a role in this problem. If the characteristic polynomial of T *splits* (i.e., factors into linear factors), we can demand $[T]_\beta^\beta$ be in *Jordan Form*. This is *almost* diagonal. However, if the characteristic polynomial does not split and we don't want to enlarge our field of scalars, we have to resort to *Rational Canonical Form* (based on so-called invariant factors) or *Primary Rational Canonical Form* (based on elementary divisors). We shall not pursue these forms here.

Assumption: $T : V \to V$ is a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$, say $\dim(V) = n$. Moreover, assume that the characteristic polynomial splits (over $\mathbb{F}$): $\boxed{\det(tI - T) = (t - \lambda_1)^{a_1} \cdots (t - \lambda_\ell)^{a_\ell}}$ where we assume that $\lambda_i \neq \lambda_j$ for $i \neq j$ and $a_i > 0$.

Recall that a_j is the *algebraic multiplicity* of the *eigenvalue* λ_j. Let $E_\lambda = \ker(T - \lambda I) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$ be the *eigenspace* associated with λ. We call $g_\lambda = \dim(E_\lambda)$ the *geometric multiplicity* of λ. Notice that $g_\lambda > 0$ if and only if λ is an eigenvalue. If (and it turns out – only if) $g_j = a_j$ for all $j = 1, \dots, \ell$, we have a basis of eigenvectors, β, and $[T]_\beta^\beta$ is a diagonal matrix with the eigenvalues appearing on the diagonal (each repeated according to its algebraic multiplicity). We now seek to have a basis of vectors that are close to being eigenvectors so our coordinate matrix is close to being diagonal.

Definition: Let $K_\lambda = \{\mathbf{v} \in V \mid (T - \lambda I)^k(\mathbf{v}) = \mathbf{0} \text{ for some } k > 0\}$. We call K_λ a *generalized eigenspace* associated with λ. Its nonzero members (if it has any) are called *generalized eigenvectors.*

Note: Let $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{v} \in K_\lambda$. Let k be the smallest positive integer such that $(T - \lambda I)^k(\mathbf{v}) = \mathbf{0}$. Then $\mathbf{w} = (T - \lambda I)^{k-1}(\mathbf{v}) \neq \mathbf{0}$ but $(T - \lambda I)(\mathbf{w}) = (T - \lambda I)^k(\mathbf{v}) = \mathbf{0}$. This means $\mathbf{w}$ is an eigenvector with eigenvalue λ. In other words, $K_\lambda \neq \{\mathbf{0}\}$ if and only if λ is an eigenvalue.

In fact, $\{\mathbf{0}\} = \ker(I) = \ker(T - \lambda I)^0 \subseteq E_\lambda = \ker(T - \lambda I) \subseteq \ker(T - \lambda I)^2 \subseteq \ker(T - \lambda I)^3 \subseteq \cdots \subseteq \bigcup_{k=1}^\infty \ker(T - \lambda I)^k = K_\lambda$ since a lower power of $T - \lambda I$ killing a vector implies that any higher power does as well. Recall that V is *finite dimensional.* This implies that our *chain* of subspaces $\cdots \subseteq \ker(T - \lambda I)^k \subseteq \ker(T - \lambda I)^{k+1} \subseteq \cdots$ cannot grow forever (each proper containment implies a growth in dimension which is capped at $\dim(V) = n$). In particular, $K_\lambda = \ker(T - \lambda I)^n$. In fact, we can do much better. Once we know about Jordan form, it is obvious that $K_{\lambda_j} = \ker(T - \lambda_j I)^m$ where m is the size of the largest Jordan block associated with λ_j.

Theorem: Our vector space is the direct sum of generalized eigenspaces: $V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \cdots \oplus K_{\lambda_\ell}$. In particular, V has a basis of generalized eigenvectors. Moreover, $K_{\lambda_j} = \ker(T - \lambda_j I)^{a_j}$ and $\dim(K_{\lambda_j}) = a_j$ (the algebraic multiplicity of λ_j) for each $j = 1, \dots, \ell$.

Proof: First, we show that the generalized eigenspaces form a direct sum. For each $i = 1, \dots, \ell$, let k_i be a positive integer such that $K_{\lambda_i} = \ker(T - \lambda_i)^{k_i}$. Let $\mathbf{v} \in K_{\lambda_j} \cap \left(\sum_{i \neq j} K_{\lambda_i}\right)$. Now $(t - \lambda_j)^{k_j}$ and $\prod_{i \neq j}(t - \lambda_i)^{k_i}$ are relatively prime polynomials. Thus, by the extended Euclidean algorithm, there exists polynomials $f(t)$ and $g(t)$ such that $f(t)(t - \lambda_j)^{k_j} + g(t)\prod_{i \neq j}(t - \lambda_i)^{k_i} = 1$. Therefore, $\mathbf{v} = I(\mathbf{v}) = f(T)(T - \lambda_j I)^{k_j}(\mathbf{v}) + g(T)\prod_{i \neq j}(T - \lambda_i I)^{k_i}(\mathbf{v}) = f(T)(\mathbf{0}) + g(T)(\mathbf{0}) = \mathbf{0}$ since vectors in K_{λ_j} are killed by $(T-\lambda_j I)^{k_j}$ and vectors in $\sum_{i\neq j} K_{\lambda_i}$ are killed by $\prod_{i\neq j}(T-\lambda_i I)^{k_i}$. Therefore, $K_{\lambda_j}\cap\left(\sum_{i\neq j} K_{\lambda_i}\right)=\{\mathbf{0}\}$ and thus our sum is direct.

Next, for some $1\leq j\leq \ell$, let $W=(T-\lambda_j I)^{k_j}(V)$ (i.e., K_{λ_j} is the kernel and W is the image of $(T-\lambda_j I)^{k_j}$). Let $\mathbf{v}\in K_{\lambda_j}\cap W$. Since $\mathbf{v}\in W$, there exists some $\mathbf{w}\in V$ such that $(T-\lambda_j I)^{k_j}(\mathbf{w})=\mathbf{v}$. Therefore, by the definition of $\mathbf{w}$ and since $\mathbf{v}$ lies in the kernel of $(T-\lambda_j I)^{k_j}$, we have $(T-\lambda_j I)^{2k_j}(\mathbf{w})=(T-\lambda_j I)^{k_j}[(T-\lambda_j I)^{k_j}(\mathbf{w})]=(T-\lambda_j I)^{k_j}(\mathbf{v})=\mathbf{0}$. Thus $\mathbf{w}\in\ker(T-\lambda_j I)^{2k_j}=\ker(T-\lambda_j I)^{k_j}=K_{\lambda_j}$. In other words, $\mathbf{v}=(T-\lambda_j I)^{k_j}(\mathbf{w})=\mathbf{0}$. Thus $K_{\lambda_j}\cap W=\{\mathbf{0}\}$. Notice that $\dim(K_{\lambda_j})+\dim(W)=\text{nullity}(T-\lambda_j I)^{k_j}+\text{rank}(T-\lambda_j I)^{k_j}=\dim(V)$. Therefore, $V=K_{\lambda_j}\oplus W$. Also, notice that since T commutes with $(T-\lambda_j I)^{k_j}$, we have both $T(K_{\lambda_j})\subseteq K_{\lambda_j}$ and $T(W)\subseteq W$ (i.e., our decomposition is T-invariant).

Let T_1 be the restriction of T to K_{λ_j} and T_2 be the restriction of T to W. Since $K_{\lambda_j}\cap W=\{\mathbf{0}\}$ and K_{λ_j} contains all of the eigenvectors associated with eigenvalue λ_j, we have that λ_j is not an eigenvalue of T_2. Also, since $K_{\lambda_j}\cap K_{\lambda_i}=\{\mathbf{0}\}$ for $i\neq j$ (because the generalized eigenspaces form a direct sum), we have that λ_j is the only eigenvalue of T_1. Next, because $V=K_{\lambda_j}\oplus W$ (and these subspaces are T-invariant), the characteristic polynomial of T is the product of the characteristic polynomials of T_1 and T_2. But λ_j is the only eigenvalue of T_1 and it is not an eigenvalue of T_2. Therefore, the characteristic polynomial of T_1 only has $t-\lambda_j$ factors and T_2's characteristic polynomial cannot have any $t-\lambda_j$ factors. Thus the characteristic polynomial of T_1 must be $(t-\lambda_j)^{a_j}$ and the characteristic polynomial of T_2 must be $\prod_{i\neq j}(t-\lambda_i)^{a_i}$. Consequently, $\dim(K_{\lambda_j})$ (i.e., the dimension of the domain of T_1) must be a_j (i.e., the degree of T_1's characteristic polynomial).

Finally, we have $\dim(K_{\lambda_1}\oplus K_{\lambda_2}\oplus\cdots\oplus K_{\lambda_\ell})=\dim(K_{\lambda_1})+\cdots+\dim(K_{\lambda_\ell})=a_1+\cdots+a_\ell$ (i.e., the degree of the characteristic polynomial of T). Thus $\dim(K_{\lambda_1}\oplus K_{\lambda_2}\oplus\cdots\oplus K_{\lambda_\ell})=n=\dim(V)$. Therefore, $K_{\lambda_1}\oplus K_{\lambda_2}\oplus\cdots\oplus K_{\lambda_\ell}=V$ and so the theorem follows. ■

Corollary: We have $1\leq g_j\leq a_j$ (i.e., the geometric multiplicity never exceeds the algebraic mulitiplicity) since $E_{\lambda_j}\subseteq K_{\lambda_j}$. Also, $E_{\lambda_1}\oplus\cdots\oplus E_{\lambda_\ell}$ (i.e., eigenspaces form a direct sum) since $E_{\lambda_j}\cap\left(\sum_{i\neq j}E_{\lambda_i}\right)\subseteq K_{\lambda_j}\cap\left(\sum_{i\neq j}K_{\lambda_i}\right)=\{\mathbf{0}\}$.

We note that since T commutes with $(T-\lambda I)^k$ for any positive integer k, we have $\ker(T-\lambda I)^k$ is T-invariant. In particular, eigenspaces and generalized eigenspaces are T-invariant. This means that if we create a basis for V by (disjoint) unioning bases for our generalized eigenspaces, our coordinate matrix will be block diagonal with $a_j\times a_j$ blocks ($j=1,\ldots,\ell$).

To finish developing Jordan form we need to seek nicely structured bases for our generalized eigenspaces. In particular, we wish to form bases consisting of *chains* of generalized eigenvectors: $\mathbf{v},(T-\lambda I)(\mathbf{v}),(T-\lambda I)^2(\mathbf{v}),\ldots,(T-\lambda I)^{k-1}(\mathbf{v})$ where $\mathbf{v}\in K_\lambda$ and $(T-\lambda I)^k(\mathbf{v})=\mathbf{0}$. First, we recall a short proof by Mark Wildon establishing the existence a basis for K_λ consisting of such chains. Then we will put this together with our previous theorem to establish the existence of our Jordan canonical form.

Theorem: The generalized eigenspace K_λ has a basis consisting of chains of generalized eigenvectors.

Proof: First, restrict $T-\lambda I$ to the domain K_λ (this is possible since K_λ is $(T-\lambda I)$-invariant). We proceed by induction on the dimension on the domain of our linear operator. The zero dimensional base case is trivial: If $K_\lambda=\{\mathbf{0}\}$, then λ is not an eigenvalue. Its basis is the empty set and thus the theorem is vacuously satisfied. Suppose $K_\lambda\neq\{\mathbf{0}\}$ and let a be the algebraic multiplicity of λ (i.e., $\dim(K_\lambda)=a$) so that $K_\lambda=\ker(T-\lambda I)^a$. Assume the theorem holds for all spaces of dimension less than a.

Notice that $(T-\lambda I)(K_\lambda)$ is properly contained in K_λ since otherwise $K_\lambda=(T-\lambda I)(K_\lambda)=(T-\lambda I)^2(K_\lambda)=\cdots=(T-\lambda I)^a(K_\lambda)=\{\mathbf{0}\}$ (contradiction). Also, if $T-\lambda I=0$ on K_λ, then $K_\lambda=E_\lambda$ and we have a basis of eigenvectors (and thus are done since eigenvectors are chains of length 1).

Therefore, we may assume that $(T-\lambda I)(K_\lambda)$ is non-zero and properly contained in K_λ. We apply an inductive hypothesis to $(T-\lambda I)(K_\lambda)$ and find $\mathbf{v}_1,\ldots,\mathbf{v}_m\in(T-\lambda I)(K_\lambda)$ so that $\mathbf{v}_1,(T-\lambda I)(\mathbf{v}_1),\ldots,(T-\lambda I)^{b_1-1}(\mathbf{v}_1),\ldots,\mathbf{v}_m,(T-\lambda I)(\mathbf{v}_m),\ldots,(T-\lambda I)^{b_m-1}(\mathbf{v}_m)$ is a basis of chains of generalized eigenvectors for $(T-\lambda I)(K_\lambda)$ and $(T-\lambda I)^{b_j}(\mathbf{v}_j)=\mathbf{0}$ for $j=1,\ldots,m$. Consequently, $\dim((T-\lambda I)(K_\lambda))=b_1+\cdots+b_m$.

Notice $\mathbf{v}_j\in(T-\lambda I)(K_\lambda)$ implies we may choose some $\mathbf{u}_j\in K_\lambda$ such that $(T-\lambda I)(\mathbf{u}_j)=\mathbf{v}_j$ (so now $\mathbf{v}_j,\ldots,(T-\lambda I)^{b_j-1}(\mathbf{v}_j)$ becomes $(T-\lambda I)(\mathbf{u}_j),\ldots,(T-\lambda I)^{b_j})(\mathbf{u}_j)$). Clearly $\ker(T-\lambda I)$ contains the linearly indpendent vectors $(T-\lambda I)^{b_1}(\mathbf{u}_1),\ldots,(T-\lambda I)^{b_m}(\mathbf{u}_m)$ (formerly called $(T-\lambda I)^{b_1-1}(\mathbf{v}_1),\ldots,(T-\lambda I)^{b_m-1}(\mathbf{v}_m)$). We extend this set to a basis for $\ker(T-\lambda I)$ by adjoining $\mathbf{w}_1,\ldots,\mathbf{w}_p$.

Claim: $\mathbf{u}_1,(T-\lambda I)(\mathbf{u}_1),\ldots,(T-\lambda I)^{b_1}(\mathbf{u}_1),\ldots,\mathbf{u}_m,(T-\lambda I)(\mathbf{u}_m),\ldots,(T-\lambda I)^{b_m}(\mathbf{u}_m),\mathbf{w}_1,\ldots,\mathbf{w}_p$ is a basis for K_λ consisting of chains of generalized eigenvectors. Notice that $(T-\lambda)^{b_i+1}(\mathbf{u}_i)=\mathbf{0}$ for $i=1,\ldots,m$ and $(T-\lambda I)(\mathbf{w}_q)=\mathbf{0}$ for $q=1,\ldots,p$. Suppose $\sum_{i=1}^m\sum_{j=0}^{b_i}c_{ij}(T-\lambda I)^j(\mathbf{u}_i)+\sum_q d_q\mathbf{w}_q=\mathbf{0}$. Applying $T-\lambda I$ to this equation, we kill off the terms coming from elements of $\ker(T-\lambda I)$. We are left with $\sum_{i=1}^m\sum_{j=0}^{b_i-1}c_{ij}(T-\lambda I)^{j+1}(\mathbf{u}_i)=\mathbf{0}$. But this is just a linear combination of vectors coming from our basis for $(T-\lambda I)(K_\lambda)$. Therefore, $c_{ij}=0$ for all $i=1,\ldots,m$ and $j=0,\ldots,b_i-1$. Thus our equation is becomes: $c_{1b_1}(T-\lambda I)^{b_1}(\mathbf{u}_1)+\cdots+c_{mb_m}(T-\lambda I)^{b_m}(\mathbf{u}_m)+d_1\mathbf{w}_1+\cdots+d_p\mathbf{w}_p=\mathbf{0}$. But these form a basis for $\ker(T-\lambda I)$ and thus the rest of our scalar coefficients are zero. Therefore, our set is linearly independent. Finally, notice that our proposed basis has $(b_1+1)+\cdots+(b_m+1)+p=b_1+\cdots+b_m+m+p$ vectors in it. We already noted that

$\dim((T-\lambda I)(K_\lambda)) = b_1 + \cdots + b_m$ and from our construction of $\mathbf{w}_q$'s we have $\dim(\ker(T-\lambda I)) = m+p$. Therefore, by the rank nullity theorem $\dim(K_\lambda) = b_1 + \cdots + b_m + m + p$. Thus our set also spans K_λ. This establishes our claim. ■

These theorems put together tells us that V has a basis consisting of chains of generalized eigenvectors. Once again, since T commutes with $T - \lambda I$, it is not hard to see that the span of a chain of generalized eigenvectors is a T-invariant subspace of V. Thus if we use a basis consisting of chains of generlized eigenvectors our coordinate matrix will be block diagonal with each block corresponding to some chain. Let us see what such a block looks like. To that end suppose $\mathbf{v}_1 = (T-\lambda I)^{k-1}(\mathbf{v}), \ldots, \mathbf{v}_{k-1} = (T-\lambda I)(\mathbf{v}), \mathbf{v}_1 = \mathbf{v}$ is a chain of generalized eigenvectors with $(T-\lambda I)^k(\mathbf{v}) = \mathbf{0}$. You may notice that we are writing our chains backwards from our previous convention. Also, for convenience let $\mathbf{v}_0 = \mathbf{0}$. Notice that $(T-\lambda I)(\mathbf{v}_j) = \mathbf{v}_{j-1}$. Therefore, $T(\mathbf{v}_j) = \mathbf{v}_{j-1} + \lambda \mathbf{v}_j$. Thus the j-th column of our block will have a 1 in the $(j-1)$-st row and a λ in its j-th row – unless we are looking at the very first column where: $T(\mathbf{v}_1) = \lambda\mathbf{v}_1 + \mathbf{v}_0 = \lambda\mathbf{v}_1$ so the first column is just a λ followed by zeros. We get the following matrix:

$$J_\lambda = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix} = \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} = \lambda I_k + N_k$$

It is interesting to note that $(N_k)^k = 0$ so that N_k is *nilpotent*. More precisely $(N_k)^{k-1} \neq 0$ but $(N_k)^k = 0$ so it is nilpotent of degree k. Obviously the diagonal part (i.e., λI) and the nilpontent part of our Jordan block commute with each other. In general, if β is a basis consisting of chains of generalized eigenvectors, we get

$$[T]_\beta^\beta = J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_s \end{bmatrix}$$

where $J_1, J_2, \ldots, J_k$ are the Jordan blocks corresponding to our various chains of generalized eigenvectors. This is called a *Jordan form* for T. Notice that since each block can be written as a sum of a diagonal and nilpotent part, we can write $J = D + N$ where D is diagonal and N is nilpotent (its degree of nilpotency will be k if the size of the largest Jordan block is $k \times k$). It isn't hard to show that the diagonal and nilpotent parts of J commute: $DN = ND$. So while it is not always possible to diagonalize our linear operator, under the condition our characteristic polynomial splits, we can get close to diagonalizing. We are off by a nilpotent part and better yet, that part commutes with our diagonal part.

The question of whether Jordan form is unique is wrapped up with computing the Jordan form of a matrix. To determine the Jordan form we need to determine the number and length of our chains associated with each eigenvalue. Notice that each chain starts with an eigenvector. Thus $\dim(E_\lambda) = \dim(\ker(T-\lambda I))$ is equal to the number of chains associated with λ. More generally, let $n_k = \dim(\ker(T-\lambda I)^k)$. Let us call $\mathbf{v}$ such that $(T-\lambda I)^k(\mathbf{v}) = \mathbf{0}$ but $(T-\lambda I)^{k-1}(\mathbf{v}) \neq \mathbf{0}$ a generalized eigenvector of degree k (so eigenvectors have degree 1 and the zero vector has degree 0). Then $n_0 = 0$, n_1 is the number of independent eigenvectors, n_2 is the number of independent eigenvectors plus generalized eigenvectors of degree 2. In general, n_k is the number of linearly independent eigenvectors of degree at most k. Notice that a k-chain is made up from one generalized eigenvector of each degree from 1 up to k. Thus the numbers $n_1, n_2, \ldots$ will let us determine how many chains we have (and how long they are). In particular, these numbers completely determine our Jordan block structure associated with λ. Putting this together, we have that our Jordan form is uniquely determined up to rearraging the order of our Jordan blocks.

Example: Consider $B = \begin{bmatrix} 2 & -1 & -1 & 0 & -1 \\ 7 & 7 & 2 & 1 & 3 \\ -4 & -2 & 2 & 0 & -2 \\ 2 & 1 & 1 & 4 & 1 \\ 1 & 1 & 2 & -1 & 5 \end{bmatrix}$. We can use software to find that $\det(tI - B) = (t-4)^5$. Therefore, the only eigenvalue of B is $\lambda = 4$. Next, we compute the nullity of powers of $B - 4I$: $n_0 = 0$, $n_1 = \text{nullity}(B-4I) = 3$, $n_2 = \text{nullity}(B-4I)^2 = 4$, $n_3 = \text{nullity}(B-4I)^3 = 5$, $n_4 = \text{nullity}(B-4I)^2 = 5$ (in fact, $n_k = 5$ for all $k \geq 3$). This means we have 3 chains of generalized eigenvectors. But only one of them is longer than 1 vector long. This chain then also extends to length 3. So we have two 1-chains and one 3-chain. Therefore, the Jordan form of B is $J = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$.

The task of finding a matrix P such that $P^{-1}BP = J$ is considerbly more difficult. This amounts to actually finding our chains of generalized eigenvectors. We simply find one chain at a time and then when finding a new chain, make sure it is independent from the previously found ones. We start with our longest chain. We need a vector $\mathbf{v}_1$ such that $(B-4I)^3\mathbf{v}_1 = \mathbf{0}$ but $(B-4I)^2\mathbf{v}_1 \neq \mathbf{0}$. This is easy since $(B-4I)^3$ is the zero matrix. Examining $(B-4I)^2$, we see that $\mathbf{v}_3 = [1\ 0\ 0\ 0\ 0]^T$ will do the job. Then let $\mathbf{v}_2 = (B-4I)\mathbf{v}_3 = [-2\ 7\ \ -4\ 2\ 1]^T$ and $\mathbf{v}_1 = (B-4I)\mathbf{v}_2 = [0\ 4\ 0\ 0\ \ -4]^T$. Our next longest chains are the only remaining chains (of length 1). To find these we need to find vectors such that $(B-4I)\mathbf{x} = \mathbf{0}$ but $(B-4I)^0\mathbf{x} = \mathbf{x} \neq \mathbf{0}$ *and* they must be independent from our other vector at this level (i.e., independent from $\mathbf{v}_1$). Thus we find a basis for $\ker(B-4I)$ and then extend $\{\mathbf{v}_1\}$ (a linearly indpendent set) to a basis. The extension part is gives us our missing vectors. We find that $\mathbf{u}_1 = [1\ \ -3\ 1\ 0\ 0]^T$, $\mathbf{w}_1 = [-1\ 2\ 0\ 1\ 0]^T$, and a multiple of $\mathbf{v}_1$ form a basis for $\ker(B-4I)$. Thus the first two vectors are our desired extension. Therefore: $\{\mathbf{u}_1, \mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is our basis of chains of generalized eigenvectors. Thus letting $P = \begin{bmatrix} 1 & -1 & 0 & -2 & 1 \\ -3 & 2 & 4 & 7 & 0 \\ 1 & 0 & 0 & -4 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & -4 & 1 & 0 \end{bmatrix}$, we find that $P^{-1}BP = J$ (since we organized our chains so they had length 1, length 1, then length 3).

Since finding the Jordan form of a matrix is relatively easy, but finding the change of basis matrix is hard, we will pass this task off to software. In Maple, after loading the linear algebra package (i.e., `with(LinearAlgebra):`) and defining your matrix (e.g. `A`), you can find a transition matrix with the command `P := JordanForm(A,output='Q');`. Then `J := P^(-1).A.P;` will be your Jordan form.

[Sketch] Finding P – An Algorithm: (I) Compute the characteristic polynomial and determine all eigenvalues, λ, and their algebraic multiplicities, a. (II) For each eigenvalue λ: (1) Compute the nullity of $\ker(T-\lambda I)^k$ (call it n_k) for $k = 1, 2, \ldots$ until $n_k = n_{k+1}$ (this must happen by n_a). The last number in this list is $\dim(K_\lambda)$. Use these numbers to determine the number of and length of each chain of generalized eigenvectors. (2) Let $W = \text{span}$ (previously found basis vectors) (initially $W = \{\mathbf{0}\}$). (3) Determine the largest positive integer k such that $\ker(T-\lambda I)^{k-1} + W$ is properly contained in $\ker(T-\lambda I)^k$. (4) Pick some $\mathbf{v} \in \ker(T-\lambda I)^k$ such that $\mathbf{v} \notin \ker(T-\lambda I)^{k-1} + W$. (5) Add $(T-\lambda I)^{k-1}(\mathbf{v}), (T-\lambda I)^{k-2}(\mathbf{v}), \ldots, (T-\lambda I)(\mathbf{v}), \mathbf{v}$ to your basis. (6) If you have not found $\dim(K_\lambda)$ vectors, go back to (2). (III) Now let assemble the (coordinate representations) of each basis vector found for each generalized eigenspace in some order (making sure to keep chains together and in the order listed in (5)) into your matrix P.

This leaves many questions. One big one might be, "Why do we care?" In the end, having a standard form is incredibly useful. If you want to test a theorem about linear operators, you most likely just need to think about how it works in the case your matrix is in Jordan form. For example, we already know that $\det(A)$ is the product of the eigenvalues of A (counting multiplicity) and $\text{trace}(A)$ is the sum. This is obvious from the Jordan form. Notice that the number of Jordan blocks counts the number of linearly independent eigenvectors. Our matrix is diagonalizable if and only if all of our Jordan blocks are 1×1!

If we let $f(t) = \det(tI - A)$ (i.e., the characteristic polynomial of A), then the Cayley-Hamilton Theorem states that $f(A) = 0$ (you cannot just plug $t = A$ into the definition to prove this – why?). Thus A is a root of a polynomial of degree n (given A is $n \times n$). One might ask what the minimum degree polynomial might be. The *minimal polynomial* for A is the monic (=leading coefficient is 1) polynomial $m(t)$ such that $m(A) = 0$ and A is not the root of any lower degree polynomial. It isn't hard to see from the Jordan form of A, if the size of the largest Jordan block associated with λ_i is $m_i \times m_i$ (for each eigenvalue λ_i), then $m(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_\ell)^{m_\ell}$. Not only does this imply that the minimal polynomial divides the characteristic polynomial, but it also reveal the theorem: A is diagonalizable if and only if its minimal polynomial has no repeated roots!

Finally, Jordan form is useful if one wants to define functions of matrices. Suppose you have a function $f(x)$ defined via a power series: $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots = \sum_{k=0}^{\infty} c_k x^k$. Notice that if $A = PJP^{-1}$ then $A^k = (PJP^{-1}) \cdots (PJP^{-1}) = PJP^{-1}PJP^{-1} \cdots PJP^{-1} = PJ \cdots JP^{-1} = PJ^kP^{-1}$. Thus (excusing sum analytic/convergence type issues), $f(A) = \sum_{k=0}^{\infty} c_k A^k = \sum_{k=0}^{\infty} c_k PJ^kP^{-1} = P\left(\sum_{k=0}^{\infty} c_k J^k\right)P^{-1} = Pf(J)P^{-1}$. Therefore, if you can sort out how to evaluate your function on a Jordan block, you can figure out how to evaluate it on any matrix!

In particular, $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. One can show that if x and y are *commuting* formal variables, then $e^{x+y} = e^x e^y$. Notice that $J = D + N$ where D and N are the diagonal and nilpotent parts of the Jordan form J. But D and N commute! Thus $e^J = e^{D+N} = e^D e^N$. Since computing powers of a diagonal matrix just amounts to computing powers of its diagonals, applying a function to a diagonal matrix just amounts to applying that function to its diagonals. Thus we can easily compute e^D. On the other hand, a high enough power of a nilpotent matrix is zero, so a power series $f(N)$ is actually a finite sum (a polynomial in N). Thus we can compute e^N. Therefore, we can effectively compute $e^A = Pe^JP^{-1} = Pe^De^NP^{-1}$. [Although there are easier ways to compute e^A.] Why it this useful? The answer is differential equations and much more!