JORDAN CANONICAL FORM

Steven Miller sjmiller@math.ohio-state.edu

I. Introduction:

We've seen that not every matrix is diagonalizable. For example, consider

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Then direct calculation shows that it is not diagonalizable. Why do we care about diagonalizing matrices? The main reason is ease of computation. If we can write $\mathbf{A = S \Lambda S^{-1}}$, then $\mathbf{A^{1000} = S \Lambda^{1000} S^{-1}}$, and the calculations can be performed very quickly. If we had to multiply 1,000 powers of A, this would be very time consuming. Theoretically, we may not need such a time-saving method, but if we're trying to model any physical system or economic model, we're going to want to run calculations on a computer. And if the matrix is decently sized, very quickly these calculations will cause noticeable time-lags.

Jordan Canonical Form is the answer. The Question? What is the 'nicest' form we can get an arbitrary matrix into. We already know that, to every eigenvalue, there is a corresponding eigenvector. If an nxn matrix has n linearly independent eigenvectors, then it is diagonalizable. Hence,

Theorem 1: ***If an nxn matrix A has n distinct eigenvalues, then A is diagonalizable, and for the diagonalizing matrix S we can take the columns to be the n eigenvectors ($S^{-1} A S = \Lambda$).***

In the proof of the above, we see all we needed was n linearly independent vectors. So we obtain

Theorem 2: ***If an nxn matrix A has n linearly independent eigenvectors, then A is diagonalizable, and for the diagonalizing matrix S we can take the columns to be the n eigenvectors ($S^{-1} A S = \Lambda$).***

Now consider the case of an nxn matrix A that does not have n linearly independent eigenvectors. Then we have

Theorem 3: ***If an nxn matrix does not have n linearly independent eigenvectors, then A is not diagonalizable.***

Proof: Assume A is diagonalizable by the matrix S.
Then $S^{-1} A S = \Lambda$, or $A = S \Lambda S^{-1}$.
The standard basis vectors $e_1, ..., e_n$ are eigenvectors of Λ, and as S is invertible, we get $Se_1, ..., Se_n$ are eigenvectors of A, and these n vectors are linearly independent. (Why?) But this contradicts the fact that A does not have n linearly independent eigenvectors. Contradiction, hence A is not diagonalizable.

So, in studying what can be done to an arbitrary nxn matrix, we need only study matrices that do not have n linearly independent eigenvectors.

Jordan Canonical Form Theorem (JCF):
Let A be an nxn matrix. Then there exists an invertible matrix M such that $M^{-1} A M = J$, where J is a block diagonal matrix, and each block is of the form

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

Note J^{1000} is much easier to computer than A^{1000}. In fact, there is an explicit formula for J^{1000} if you know the eigenvalues and the sizes of each block.

II. Notation:

Recall that λ is an eigenvalue of A if $\text{Det}(A - \lambda I) = 0$, and v is an eigenvector of A with eigenvalue λ if $Av = \lambda v$. We say v is a generalized eigenvector of A with eigenvalue λ if there is some number N such that $(A-\lambda I)^N v = 0$. Note all eigenvectors are generalized eigenvectors.

For notational convenience, we write **gev** for generalized eigenvector, or **λ-gev** for generalized eigenvector corresponding to λ.

We say the **λ-Eigenspace of A** is the subspace spanned by the eigenvectors of A that have eigenvalue λ. Note that this is a subspace, for if v and w are eigenvectors with eigenvalue λ, then av + bw is an eigenvector with eigenvalue λ.

We define the **λ-Generalized Eigenspace of A** to be the subspace of vectors killed by some power of (A-λI). Again, note that this is a subspace.

III. Needed Theorems:

Fundamental Theorem of Algebra: Any polynomial with complex coefficients of degree n has n complex roots (not necessarily distinct).

Cayley-Hamilton Theorem: Let $p(\lambda) = \text{Det}(A-\lambda I)$ be the characteristic polynomial of A. Let $\lambda_1, ..., \lambda_k$ be the distinct roots of this polynomial, with multiplicities $n_1, ..., n_k$ (so $n_1 + ... + n_k = n$). Then we can factor $p(\lambda)$ as

$$p(\lambda) = (\lambda - \lambda_1)^{n1} (\lambda - \lambda_2)^{n2} * ... * (\lambda - \lambda_k)^{nk},$$

and the matrix A satisfies

$$p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk} = 0,$$

Schur's Lemma (Triangularization Lemma): Let A be an nxn matrix. Then there exists a unitary U such that $U^{-1} A U = T$, where T is an upper triangular matrix.

Proof: construct U by fixing one column at a time.

IV. Reduction to Simpler Cases:

In the rest of this handout, we will always assume A has eigenvalues $\lambda_1, ..., \lambda_k$, with multiplicities $n_1, ..., n_k$ (so $n_1 + ... + n_k = n$). We will show that we

can find n_1 λ-gev, n_2 λ-gev, ..., n_k λ-gev, such that these n vectors are linearly independent (LI). These will then form our matrix M.

So, if we can show that the n generalized eigenvectors are linearly independent, and that each one 'block diagonalizes' where it should, it is enough to study each λ separately.

For example, we'll show it's sufficient to consider $\lambda = 0$. Let λ be an eigenvalue of A. Then if v_j is a generalized eigenvector of A with eigenvalue λ, then v_j is a generalized eigenvector with eigenvalue 0 of $B = A - \lambda I$:

$$A\, v_j = \lambda\, v_j + v_{j-1} \quad \rightarrow \quad B\, v_j = 0\, v_j + v_{j-1}.$$

So, if we can find n_j LinIndep gev for B corresponding to 0, we've found n_j LinIindep gev for A corresponding to λ.

The next simplification is that if we can find n_j LinIndep gev for $U^{-1} B U$, then we've found n_j LinIndep gev for B. The proof is a straightforward calculation: let v_1, ..., v_m be the m LinIndp gev for $U^{-1} B U$; then $U^{-1} v_1$, ..., $U^{-1} v_m$ will be m LinIndep gev for B.

Lemma 4: Let $p(\lambda) = (\lambda - \lambda_1)^{n1} (\lambda - \lambda_2)^{n2} * ... * (\lambda - \lambda_k)^{nk}$ be the char poly of A, so $p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk}$. For $1 \leq i \leq k$, consider $(A - \lambda_i I)$. This matrix has exactly n_i LinIndep generalized eigenvectors with eigenvalue 0, hence A has n_i LinIndep generalized eigenvectors with eigenvalue λ_i.

Proof: For notational simplicity, we'll prove this for $\lambda = \lambda_1$, and let's write m for the multiplicity of λ (so $m = n_1$). Further, by the above arguments we see it is sufficient to consider the case $\lambda = 0$. By the Triangularization Lemma, we can put $B = A - \lambda I$ (which has first eigenvalue = 0) into upper triangular form. What we need from the proof is that if we take the first column of U to be v, where v is an eigenvector of B corresponding to eigenvalue 0, then the first column of $T = U_1^{-1} B U_1$ would be $(0,0, ..., 0)^T$.

The lower (n-1)x(n-1) block of T_n, call it C_{n-1}, is upper triangular, hence the eigenvalues of B appear as the entries on the main diagonal. Hence we can

again apply the triangularization argument to C_{n-1}, and get an (n-1)x(n-1) unitary matrix U_{2b}, such that $U_{2b}^{-1} C_{n-1} U_{2b} = T_{n-1}$ has first column (0,0,....0), and the rest is upper triangular. Hence we can form an nxn unitary matrix U_2

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & & & & \\ 0 & & U_{2b} & & \\ \dots & & & & \\ 0 & & & & \end{pmatrix}$$

Then $U_2^{-1} U_1^{-1} B U_1 U_2 =$

$$\begin{pmatrix} 0 & * & * & \dots & * \\ 0 & 0 & * & \dots & * \\ 0 & 0 & * & \dots & * \\ \dots & \dots & \dots & & \dots \\ 0 & 0 & 0 & \dots & * \end{pmatrix}$$

The net result is that we've now rearranged our matrix so that the first two entries on the main diagonal are zero. By 'triangularizing' like this m times, we can continue so that the upper mxm block is upper triangular, with zeros along the main diagonal, and the remaining entries on the main diagonal are non-zero (as we are assuming the multiplicity was m). Call this matrix T_m. Note there is a unitary U such that $T_m = U^{-1}BU$. *Remember, T_m and B are nxn matrices, not mxm matrices.*

Sublemma 1: At most m vectors can be killed by powers of T_m.
Proof: direct calculation: When we multiply powers of T_m, we still have an upper triangular matrix. The entries on the main diagonal are zero for the first m terms, and then non-zero for the remaining terms (because the multiplicity of the eigenvalue $\lambda = 0$ is exactly m). Hence the vectors e_{m+1}, e_{m+2}, ..., e_n are not killed by powers of T_m, and so powers of T_m can have a nullspace of dimension at most m.

We now show that exactly m vectors are killed by T_m. This follows immediately from

Sublemma 2: Let C be an mxm upper triangular matrix with zeros along the main diagonal. Then C^m is the zero matrix.
Proof: straightforward calculation, left to the reader.

Hence the nullspace of $(T_m)^m$ (and all higher powers) is exactly m, which proves there are m generalized eigenvectors of B with eigenvalue $\lambda = 0$. These vectors are LinIndep: As T_m is upper triagonal with zeros on the main diagonal for the first m entries, T_m has m LinIndep gev $e_1, \ldots, e_m$ with eigenvalue 0. As $B = UT_mU^{-1}$, B has m LinIndep gev $Ue_1, \ldots, Ue_m$ with eigenvalue 0 (show that B cannot have any more LinIndep gev with $\lambda = 0$).

Returning to the proof of Lemma 4, we see that there are exactly n_1 vectors killed by $(A-\lambda_1 I)^{n1}$, ..., n_k vectors killed by $(A-\lambda_k I)^{nk}$.

The only reason we go thru this triagonalizing is to conclude that there are exactly n_i vectors killed by $(A-\lambda_i I)^{ni}$. Try to prove this fact directly!

IV. Appendix: Representation of λ-Generalized Eigenvectors.

We know that if λ is an eigenvalue with multiplicity m, there are m generalized eigenvectors, satisfying $(A - \lambda I)^m v = 0$. We describe a very useful way to write these eigenvectors. Let us assume there are t eigenvectors, say $v_1, \ldots, v_t$. We know there are m λ-gev. Note if v is a λ-gev, so is $(A-\lambda I)v$, $(A-\lambda I)^2 v$, …., $(A-\lambda I)^m v$. Of course, some of these may be the zero vector.

We claim that each eigenvector is the termination of some chain of λ-gev. In particular, we have

$$(A-\lambda I)\, v_{1,a} = v_{1,a-1}$$
$$(A-\lambda I)\, v_{1,a-1} = v_{1,a-2}$$
$$\ldots$$
$$(A-\lambda I)\, v_1 = 0 \qquad \text{where } v_1 = v_{1,1}.$$

and

$$(A-\lambda I)\, v_{2,b} = v_{2,b-1}$$
$$(A-\lambda I)\, v_{2,b-1} = v_{2,b-2}$$
$$\vdots$$
$$(A-\lambda I)\, v_2 = 0 \qquad \text{where } v_2 = v_{2,1}.$$

all the way down to

$$(A-\lambda I)\, v_{t,r} = v_{t,r-1}$$
$$(A-\lambda I)\, v_{t,r-1} = v_{t,r-2}$$
$$\vdots$$
$$(A-\lambda I)\, v_t = 0 \qquad \text{where } v_t = v_{t,1},$$

and $a + b + \ldots + r = m$.

We emphasize that we have not shown that such a sequence of λ-gev exists. Later we shall show how to construct these vectors, and then in Lemma 8 we will prove they are Linearly Independent. For now, we assume their existence (and linear independence), and complete the proof of Jordan Canonical Form.

Let us say a λ-gev is a ***pure-gev*** if it is not an eigenvector. Thus, in the above we have t eigenvectors, and m-t pure-generalized eigenvectors. For notational convenience, we often label the λ-generalized eigenvectors by $v_1, \ldots, v_m$. Thus, for a given j, we have $(A-\lambda I)v_j = 0$ if v_j is an eigenvector, and $(A-\lambda I)v_j = v_{j-1}$ if v_j is a pure-gev.

V. Linear Independence of the λ-Generalized Eigenspaces.

Assume the n_1 gev corresponding to λ_1 are linearly independent amongst themselves, and the same for the n_2 gev corresponding to λ_2, ... We now show that the n gev are linearly independent. This fact complete the proof of Jordan Canonical Form (of course, we still must prove the n_i λ_i-gev are linearly independent).

Assume we have some linear combination of the n gev equaling zero. By LC λ_i-gev we mean a linear combination of the n_i λ_i-gev. (This is just to simplify notation).

Then (LC λ_1-gev) + (LC λ_2-gev) + ... + (LC λ_k-gev) = 0.

We'll show first that the coefficients in the first linear combination are all zero. Recall the characteristic polynomial is

$$p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk}.$$

Define

$$g_1(A) = (A - \lambda_2 I)^{n2} (A - \lambda_3 I)^{n3} * ... * (A - \lambda_k I)^{nk},$$

$g_1(A)$ kills (LC λ_2-gev), $g_1(A)$ kills (LC λ_3-gev),...., $g_1(A)$ kills (LC λ_k-gev). Why? For example, for the λ_2-gev, they are all killed by $(A-\lambda_2 I)^{n2}$, and hence as $g_1(A)$ contains this factor, they are all killed.

What does $g_1(A)$ do to (LC λ_1-gev)? Again, for notational simplicity we'll write m for n_1, and v_1, ..., v_m for the corresponding m λ_1-gev.

We can look at it factor by factor, as all the different terms $(A - \lambda_i I)$ commute.

Lemma 5: For i > 1, let the v_j's be the gev corresponding to λ_1.
If v_j is a pure-gev, then $(A - \lambda_i I)\, v_j = (\lambda_1 - \lambda_i)\, v_j + v_{j-1}$
If v_j is an eigenvector, then $(A - \lambda_i I)\, v_j = (\lambda_1 - \lambda_i)\, v_j$.

Again, the proof is a calculation: if v_j is a pure-gev,

$$\begin{aligned}(A - \lambda_i I)\, v_j &= (A - \lambda_1 I + \lambda_1 I - \lambda_i I)\, v_j \\ &= (A - \lambda_1 I)\, v_j + (\lambda_1 I - \lambda_i I)\, v_j \\ &= v_{j-1} + (\lambda_1 - \lambda_i)\, v_j\end{aligned}$$

The proof when v_j is an eigenvector is similar.

Now we examine $g_1(A)$ ((LC λ_1-gev) + (LC λ_2-gev) + ... + (LC λ_k-gev)) = 0. Clearly $g_1(A)$ kills the last k-1 linear combinations, and we are left with

$$g_1(A)\ (\text{LC } \lambda_1\text{-gev}) = 0$$

Let's say the LC λ_1-gev = $a_1 v_1 + ... + a_m v_m$. We need to show that all the a_j's are zero. (*Remember we are assuming the v_j's are linear independent – we will prove this fact when we construct the v_j's*). Assume $a_m \neq 0$. From our labeling, v_m is either an eigenvector, or a pure-gev that starts a chain leading to an eigenvector: v_m, $(A - \lambda_i I)\, v_m$, $(A - \lambda_i I)2\, v_m$, …. Note no other chain will contain v_m.

We claim that $g_1(A)$ (LC λ_1-gev) will contain a non-zero multiple of v_m. Why? When each factor $(A - \lambda_i I)$ hits a v_j, one gets back $(\lambda_1 - \lambda_i)\, v_j + v_{j-1}$ if v_j is not an eigenvector, and $(\lambda_1 - \lambda_i)\, v_j$ if v_j is an eigenvector. Regardless, we always get back a non-zero multiple of v_j, as $\lambda_1 ? \lambda_i$.

Hence direct calculation shows the coefficient of v_m in $g_1(A)$ (LC λ_1-gev) is

$$a_m (\lambda_1 - \lambda_2)^{n2} (\lambda_1 - \lambda_3)^{n3} * ... * (\lambda_1 - \lambda_k)^{nk}$$

As we are assuming the different λ's are distinct (remember we grouped the eigenvalues together to have multiplicity), this coefficient is non-zero. As we are assuming v_1 thru v_m are linearly independent, the only way the coefficient of v_m can be zero is if $a_m = 0$. Similar reasoning implies $a_{m-1} = 0$, and so on. Hence we have proved:

Theorem 5: Assuming that the n_i generalized eigenvectors associated to the eigenvalue λ_i are linearly independent (for $1 \leq i \leq n$), then the n generalized eigenvectors are linearly independent. Furthermore, there is an invertible M such that $M^{-1} A M = J$.

The only item not immediately clear is what M is. As an exercise, show that one may take M to be the generalized eigenvectors of A. They must be put in a special order. For example, one may group all the λ_1-gev together, the λ_2-gev together, and so on. For each i, order the λ_i-gev as follows: say there are t eigenvectors which give sequences v_1, …, $v_{1,a}$, v_2, …, $v_{2,b}$,…, v_t,…,$v_{t,r}$. Then this ordering works (exercise).

VI. Finding the λ-gev:

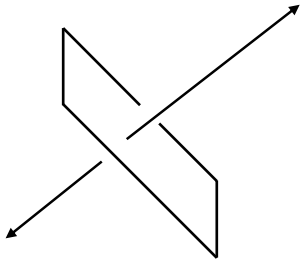
The above arguments show we need only find the n_i generalized eigenvectors corresponding to the eigenvalue λ_i; these will be of the form $(A-\lambda_i I)\, v_j = 0$ or $(A-\lambda_i I)\, v_j = v_{j-1}$. Moreover, we've also seen we may take $\lambda_i = 0$ without loss of generality. For notational convenience, we write λ for λ_i and m for n_i.

So we ***assume the multiplicity of $\lambda = 0$ to be m.*** Hence in the sequel we show how to find m generalized eigenvectors of an mxm matrix whose m^{th} power vanishes. (By the triangularizing we're done, finding m such generalized eigenvectors for this is equivalent to finding m generalized eigenvectors for the original nxn matrix A).

We define the following spaces, where A is our mxm matrix:

1. N(A) = Nullspace(A). The dimension of this is the number of eigenvectors, as we are assuming $\lambda = 0$.
2. $V_1 = W_1 = N(A)$
3. $V_i = N(A^i)$, all vectors killed by A^i. Note that V_m is the entire space.
4. $W_i = \{w \in N(A^i)$ such that $w \perp N(A^{i-1})\}$, for $2 \le i \le m$.

For example, assume we are in R^3, and A^2 is the zero matrix. Let's consider V_2. For definiteness, assume V_1 is 1-dimensional, and V_2 is 3-dimensional. W_1 is just V_1. The problem is, if y_1 and y_2 are two vectors killed by A^2 and not by A, then it is possible that $y_1 - y_2$ (or some linear combination) is killed by A.



In the picture above, the line represents W_1 and the plane represents W_2. Anything in the 3-space above is killed by A^2, and *only* those vectors along the line are killed by just A.. It is possible to take two vectors in R^3 that are linearly independent, neither of which lie on the line, but their difference *does* lie on the line.

Why are we constructing such spaces as W_2? Why isn't V_2 good enough? The reason is we want a very nice basis. The first basis vector will just be a vector in $V_1 = W_1$. For the other two directions, we can take two vectors perpendicular to W_1. (How? This is a 3-dimensional space – simply apply Gram-Schmidt).

The advantage of such a basis is that if z_1 and z_2 are linearly independent vectors in W_2, then the only way a $z_1 + b\ z_2$ can be in W_1 is for $a = b = 0$. Why? W_2 is a subspace, and as z_1 and z_2 are perpendicular to W_1, so is their linear combination. So their linear combination is still in the plane perpendicular to W_1, and as long as a and b are not both zero, it will not be the zero vector in the plane, hence it will be killed by A^2 and not A.

What we are really doing is *Partial Orthogonal Complementation.* Instead of finding the orthogonal complement of V_1 in R^m, we are finding the orthogonal complement in V_2.

Let L be the smallest integer such that A^L is the zero matrix. Then we only need to look up to W_L and V_L. V_L will be an m-dimensional space (as every vector is killed by A^L). We'll have a nice basis for V_L, consisting of the bases of W_1, ..., W_L. The advantage of this decomposition is that the spaces are mutually perpendicular, and if you have a linear combination of vectors in W_j, then the only way it can be in a W_b with $b < j$ is if the combination is the zero vector.

Lemma 6: $\dim(W_{i-1}) \geq \dim(W_i)$, for i = 2, 3, ..., L.
Proof: Assume not: let $N = \dim(W_i)$. So consider a basis of W_i: $z_1, z_2, ..., z_N$, and the vectors $Az_1, ..., Az_N$. Clearly each Az_j is in V_{i-1}. We claim each Az_j must have some component in W_{i-1}. Why? The smallest power that kills each z_j is A^i. If there was no component in W_{i-1}, then A^{i-2} would kill z_j, contradiction.

Let $P = P_{i-1}$ be the projection operator from V_{i-1} to W_{i-1}. Note $P^2 = P$, and by the above arguments each vector $Az_1, ..., Az_N$ has a non-zero component in W_{i-1}. Therefore the N vectors $PAz_1, ..., PAz_N$ are N non-zero vectors in W_{i-1}.

As we are assuming that $\dim(W_{i-1}) < \dim(W_i)$, the N vectors PAz_1 thru PAz_N cannot be linearly independent, for the dimension of a subspace is the

maximal number of linear independent vectors you can have in that space. Hence there exist constants, not all zero, such that

$$a_1 PAz_1 + ... + a_N PAz_N = 0$$

Hence $PA(a_1z_1 + ... + a_Nz_N) = 0$. By the definition of W_i, the linear combination $a_1z_1 + ... + a_Nz_N$ is in W_i. Therefore, the smallest power of A that kills it is I *unless* it is the zero vector. As we are assuming the vectors z_1 through z_N are linearly independent, it is only the zero vector if $a_1 = ... = a_N = 0$.

As $i > 1$, A cannot kill $a_1z_1 + ... + a_Nz_N$ unless this is the zero vector. Could PA kill a non-zero vector? No: by definition, if $a_1z_1 + ... + a_Nz_N$ is not the zero vector, then it is in W_i. Therefore $A(a_1z_1 + ... + a_Nz_N)$ has a non-zero component in W_{i-1} (if not, that contradicts $a_1z_1 + ... + a_Nz_N$ is in W_i). Therefore, $PA(a_1z_1 + ... + a_Nz_N)$ cannot be zero, as $A(a_1z_1 + ... + a_Nz_N)$ has a component in W_{i-1}. Therefore, the only way $PA(a_1z_1 + ... + a_Nz_N)$ can be the zero vector is if $a_1z_1 + ... + a_Nz_N$ is the zero vector, which forces $a_1 = ... = a_N = 0$. Contradiction. QED.

REMARK: by Lemma 6, we see our previous example is impossible. An example that is consistent with Lemma 6 is to consider R^5, let $V_1 = W_1$ be three-dimensional, and W_1 a plane perpendicular to W_1.

Lemma 7: $\dim(W_i) \geq 1$ for i = 1, 2, ..., L.

Proof: As L is the smallest integer such that A^L is the zero matrix, then there must be a vector killed by A^L but not by A^{L-1}. Whence $\dim(W_L)$ is at least 1, so by Lemma 6 we obtain $\dim(W_i)$ is at least 1 for i = 1, 2, ..., L.

We now show how to construct the m generalized eigenvectors. We find bases for the spaces W_1, W_2, ..., W_m. We then use A to 'pullback'.

It's easier to explain by an example: assume the dimensions are as follows. Let's take m = 12, and L = 5. For definiteness sake, consider the following:

	$\mathbf{V_1}$	$\mathbf{V_2}$	$\mathbf{V_3}$	$\mathbf{V_4}$	$\mathbf{V_5}$	$\mathbf{V_6}$
	$\mathbf{W_1}$	$\mathbf{W_2}$	$\mathbf{W_3}$	$\mathbf{W_4}$	$\mathbf{W_5}$	$\mathbf{W_6}$
dimW:	4	3	2	2	1	0
basis:	$u_1,...,u_4$	$v_{1,...,}v_3$	$w_{1,}w_2$	x_1,x_2	y	

pullback: $A^4 y \leftarrow A^3 y \leftarrow A^2 y \leftarrow A y \leftarrow y$

Now, W_4 is 2-dimensional.

WE DO NOT KNOW THAT Ay IS IN W_4 !!! IT IS QUITE POSSIBLE THAT Ay IS KILLED BY A^4 AND NO SMALLER POWER OF A WITHOUT BEING IN W_4 !!!

We know y is killed by A^5 and no smaller power of A; hence Ay is killed by A^4 and no smaller power of A. But this does not mean that Ay is in W_4. Fortunately, there is a huge degree of non-uniqueness in the Jordan Canonical Form. We did not need Ay to be in W_4 – all we needed was Ay (and A^2y, A^3y, ...) to be killed by A^4 and nothing lower (A^3 and nothing lower, ...). We'll see below how to handle this.

So for now, all we know is that Ay is in V_4, with a non-zero projection in W_4; that A^2 y is in V_3 with a non-zero projection in W_3, and so on.

W_4 is 2-dimensional. Choose a vector x in W_4 such that x is linearly independent with the projection of Ay onto W_4. Then this x gives us another Jordan Block:

	$\mathbf{V_1}$	$\mathbf{V_2}$	$\mathbf{V_3}$	$\mathbf{V_4}$	$\mathbf{V_5}$	$\mathbf{V_6}$
	$\mathbf{W_1}$	$\mathbf{W_2}$	$\mathbf{W_3}$	$\mathbf{W_4}$	$\mathbf{W_5}$	$\mathbf{W_6}$
dimW:	4	3	2	2	1	0
basis:	$u_1,...,u_4$	$v_{1,...,}v_3$	$w_{1,}w_2$	x_1,x_2	y	

pullback $A^4 y \leftarrow A^3 y \leftarrow A^2 y \leftarrow A y \leftarrow y$
pullback $A^3 x \leftarrow A^2 x \leftarrow A x \leftarrow x$

We continue the game (noting that Ax is in V_3, but not necessarily in W_3). We already have two candidates for directions in V_3, namely A^2y and Ax. We'll show later that though they are not necessarily in W_3, they are killed by A^3 and not A^2, and that their projections onto W_3 are linearly independent.

We need to find 3 directions in W_2. The projections of A^3y and A^2x give us at most two (these two directions could be the same – again, we will show later that this cannot happen). As W_2 is a 3 dimensional space, we can find a vector v in W_3 that is linearly independent with the projections of A^3y and A^2x:

$\mathbf{V_1}$		$\mathbf{V_2}$		$\mathbf{V_3}$		$\mathbf{V_4}$		$\mathbf{V_5}$	$\mathbf{V_6}$
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							

We then have to find four directions in W_1, and have three candidates. We'll see later these three candidates are linearly independent, hence we can find a fourth vector u linearly independent with the rest:

$\mathbf{V_1}$		$\mathbf{V_2}$		$\mathbf{V_3}$		$\mathbf{V_4}$		$\mathbf{V_5}$	$\mathbf{V_6}$
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							
u									

We now have enough (m) candidates. We will show that they are linearly independent.

First we prove that $A^2 y$ and Ax are linearly independent; then we will prove $A^3 y$, $A^2 x$ are linearly independent, and so on. (Proof suggested by L. Fefferman and O. Pascu). Assume $a A^2 y + b Ax = 0$. Then $A(a Ay + b x) = 0$. But Ay has a non-zero projection in W_4, and we've chosen x to be linearly independent in W_4 with the projection of Ay. Therefore the smallest power that can kill this combination is A^4, *unless* it is the zero combination. Hence the only way this can be killed by A is if $a = b = 0$.

Similarly, assume a $A^3 y + b A^2 x = 0$. Then $A^2(a Ay + b x) = 0$, and by the same argument as above, a = b = 0. Note that we did not need Ay to be in W_4, only that it had a non-zero projection there.

By construction, v is linearly independent with $A^3 y$ and $A^2 x$. What about A^4y, A^3x, and Av? Assume $a A^4y + b A^3x + cAv = 0$. Then again we obtain $A(aA^3y + bA^2x + cv) = 0$, and the construction of v forces a = b = c = 0.

Lemma 8: Assume now some linear combination of the m generalized eigenvectors constructed above is zero. Then all the coefficients are zero.

Proof:

V_1		V_2		V_3		V_4		V_5	V_6
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							
u									

Assume the coefficient of y, a, is non-zero. Then we have

a y = - (rest of terms), where the rest is killed by A^4, but y isn't.

Hence a must be zero.

Now assume the coefficients of Ay and x are a and b, respectively. Then

a Ay + b x = - (rest of terms), rest killed by A^3.

As x is linearly independent with the projection of Ay onto W_4, a Ay + bx is killed by A^4 and not A^3 *unless* this combination is the zero vectory. As the rest of the terms *are* killed by A^3 , this implies a = b = 0.

Continuing to argue in this way, we obtain all the coefficients are zero, and hence the generalized eigenvectors are linearly independent.

We can now build up our matrix M and J! At last!

For each λ, we have associated generalized eigenvectors. For definiteness sake, let's consider the above case, and I'll leave to you the generalization. We have 4 blocks corresponding to $\lambda = 0$: one block with 5 generalized eigenvectors (starting with the eigenvector A^4y, and ending with y), another block with 4 gev (starting with the eigenvector A^3x, and ending with x), another block of length 2, and one of length 1.

We can order the blocks anyway we want – that will just change the order of the block in J; however, in each block we must write the vectors starting with the eigenvector on the far left, and going to the highest generalized eigenvector on the far right:

$$\begin{pmatrix} A^4y & A^3y & A^2y & Ay & y & A^3x & A^2x & Ax & x & Av & v & u \end{pmatrix}$$

Another possible arrangements would be

$$\begin{pmatrix} A^3x & A^2x & Ax & x & A^4y & A^3y & A^2y & Ay & y & Av & v & u \end{pmatrix}$$

and so on. I'll leave it to you to verify that $M^{-1} A M = J$.

VII. Calculation Shortcut:

When trying to find bases for the spaces W_i, there is a nice shortcut. First, we find a basis for V_i, or start to. How do we find vectors killed by A^i? We just have to find the nullspace of A^i. We do this by Gaussian Elimination, reducing A^i to an upper triangular matrix U. We can assume we've already found bases for W_1 thru W_{i-1}, or equivalently, a basis for V_{i-1}. Let's say the basis for V_{i-1} is $b_1, b_2, ..., b_q$. Then if we add these q rows to U, forming a new matrix U', we observe the following:

(1) If U' v = 0, then v is killed by A^i (from the first m rows of U' are the same as those of U).
(2) If U' v = 0, then v is perpendicular to W_1 thru W_{i-1}: this follows immediately from the fact that we put the basis for W_1 thru W_{i-1} as the last q rows of U', and so this forces v to be perpendicular to these spaces.

Also, if an eigenvalue has multiplicity 3 or less, counting the number of linearly independent eigenvectors gives us the Jordan Form. Why? If there are 3 LI eigenvectors, it's diagonalizable. If there is only 1, it must be a 3x3 block. If there are 2, we must have a 2x2 and a 1x1 block. Note we have no idea what M looks like.

Also note that this argument fails for multiplicity 4 and greater. If we have multiplicity 4 and 2 eigenvectors, it could be 2x2, 2x2, or it could be 3x3, 1x1.

Note the difference between theory and practice: theoretically, we know that bases for the different W_i exist, so with a wave of the hand we have them to work with. But if we were actually going to Jordanize large matrices, finding bases for all these spaces takes time, and we don't always need all those basis elements. Often it's enough to just find vectors in V_i; for example, show that if instead of taking y in W_L we took y in V_L the pullback process would work. Then there could be many i where we've pulled-back all the vectors we need, and hence there would be no need there to find a basis. If this doesn't make too much sense, don't worry: it's late at night here for me, and at this stage in your life, you won't be dealing with terrible Jordanizations where this would really make a difference. I just want to emphasize that often you can come up with a theoretical line of argument that, in practice, will yield the correct answer, but be so computationally inefficient that a better way is greatly desired.

SUMMARY: HOW TO JORDANIZE:

STEP 1: Find the eigenvalues, their multiplicities, and all the eigenvectors.
STEP 2: For each eigenvalue λ and it's multiplicity m, calculate $(A-\lambda I)$, $(A-\lambda I)^2$, ..., $(A-\lambda I)^m$.
STEP 3: Find bases for the spaces W_i described above. This will yield bases for V_i. Use the calculation shortcut to find the bases.
STEP 4: 'Pullback' vectors as described, add in vectors linearly independent with projections as needed.