Math 108B Professor: Padraic Bartlett

Lecture 8: The Jordan Canonical Form

Weeks 9-10 *UCSB 2014*

In these last two weeks, we will prove our last major theorem, which is the claim that all matrices admit something called a Jordan Canonical Form. First, recall the following definition from last week's classes:

Definition. If A is a matrix in the form

$$A = \left[\begin{array}{c|c|c|c} B_1 & 0 & \ldots & 0 \\ \hline 0 & B_2 & \ldots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \ldots & B_k \end{array}\right],$$

where each B_i is some $l_i \times l_i$ matrix and the rest of the entries of A are zeroes, we say that A is a **block-diagonal** matrix. We call the matrices $B_1, \ldots B_k$ the **blocks** of A.

For example, the matrix

$$A = \left[\begin{array}{ccc||ccc||cc} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 0 & \pi & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \pi \end{array}\right].$$

is written in block-diagonal form, with blocks $B_1 = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, $B_2 = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$, and $B_3 = \begin{bmatrix} \pi & 1 \\ 0 & \pi \end{bmatrix}$.

The blocks of A in our example above are particularly nice; each of them consists entirely of some repeated value on their diagonals, 1's on the entries directly above the main diagonal, and 0's everywhere else. These blocks are sufficiently nice that we give them a name here:

Definition. A block B_i of some block-diagonal matrix is called a **Jordan block** if it is in the form

$$\begin{bmatrix} \lambda & 1 & 0 & 0 & \dots & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}.$$

In other words, there is some value λ such that B_i is a matrix with λ on its main diagonal, 1's in the cells directly above this diagonal, and 0's elsewhere.

Using this definition, we can create the notion of a **Jordan normal form**:

Theorem. Suppose that A is similar to an $n \times n$ block-diagonal matrix B in which all of its blocks are Jordan blocks; in other words, that $A = UBU^{-1}$, for some invertible U. We say that any such matrix A has been written in **Jordan canonical form.** (Some authors will say "Jordan normal form" instead of "Jordan canonical form:" these expressions define the same object.)

The theorem we are going to try to prove this week is the following:

Theorem. Any $n \times n$ matrix A can be written in Jordan canonical form.

This result is (along with the Schur decomposition and the spectral theorem) one of the crown jewels of linear algebra; it is a remarkably strong result about what (up to similarity) **any** linear map must look like, and one of the more powerful tools a linear algebraicist has at their disposal.

Perhaps surprisingly, though, its proof is not very involved! In this talk, we present a somewhat more esoteric and interesting proof than most textbooks traditionally develop, as described by Brualdi in his paper "The Jordan Canonical Form: an Old Proof." Most modern textbooks use the concept of generalized eigenvectors and null spaces to show that the Jordan Canonical Form must exist; while these proofs are certainly enlightening and worth reading (if we have time in week 10, I'll say a word about how they go) I find them to be less interesting from the perspective of explaining how one actually **constructs** a Jordan canonical form (as they tend to mostly be proofs that assert that such things exist!)

Our proof here, however, is quite explicitly constructive, and to boot fairly elementary! All we will need to perform this proof are the following results:

- The Schur decomposition, which will tell us that every matrix is similar to some upper-triangular matrix.
- Our results last week on how conjugating by elementary matrices changes a matrix.
- Our results last week about how conjugating a matrix by a permutation matrix shuffles its rows and columns.

We construct our proof via a series of lemmas, based on these three results above:

Lemma 1. *Suppose that R is some upper-triangular matrix, and that r_{ii}, r_{jj} are a pair of distinct diagonal entries of R. Without loss of generality, assume that $i < j$. Consider the matrix ERE^{-1}, where E is the elementary matrix $E_{\text{add } h \text{ copies of entry } j \text{ to entry } i}$.*

Suppose that h is chosen such that $r_{ij} + h(r_{jj} - r_{ii}) = 0$. Then ERE^{-1} is a matrix in which its (i, j)-th entry is 0, and moreover agrees with R everywhere except for at the entries of row i with coordinates $j, \ldots n$ and the entries of column j with coordinates $1, \ldots i$.

Before we give our proof, we first illustrate what we're talking about with an example:

Example. Consider the matrix

$$R = \begin{bmatrix} 2 & 1 & 0 & 1 & 1 & 1 & 7 & 0 \\ 0 & 3 & 1 & 4 & 5 & 2 & 3 & 2 \\ 0 & 0 & 5 & 0 & 2 & 2 & 3 & 6 \\ 0 & 0 & 0 & 5 & 1 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 & 5 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notice that the entries $r_{33} = 5, r_{77} = 2$ are distinct. So: let's examine what happens if we multiply R on the left by

$$E = E_{\text{add } h \text{ copies of entry 7 to entry 3}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & h & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and on the right by E^{-1}, which we know from last week is

$$E^{-1} = E_{\text{add } -h \text{ copies of entry 3 to entry 1}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -h & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Well: on one hand we know that ER is the matrix R with h copies of its seventh row added to its third row, because this is what elementary matrices do when you multiply them. On the other hand, we know that $(ER)E^{-1}$ is just the matrix ER with -h copies of its third column added to its seventh column, again as we've proven in week 8!

Therefore, ERE^{-1} is just the matrix R with h copies of its seventh row added to its third row, and $-h$ copies of its third column added to its seventh column. In other words,

$$ERE^{-1} = \begin{bmatrix} 2 & 1 & 0 & 1 & 1 & 1 & 7-0h & 0 \\ 0 & 3 & 1 & 4 & 5 & 2 & 3-1h & 2 \\ 0 & 0 & 5 & 0 & 2 & 2 & 3+2h-5h & 6+2h \\ 0 & 0 & 0 & 5 & 1 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 & 5 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Suppose we pick h such that $r_{37} + h(r_{33} - r_{77}) = 5 + 2h - 5h = 0$; i.e. $h = 1$. In this situation, we have that

$$ERE^{-1} = \begin{bmatrix} 2 & 1 & 0 & 1 & 1 & 1 & 7 & 0 \\ 0 & 3 & 1 & 4 & 5 & 2 & 2 & 2 \\ 0 & 0 & 5 & 0 & 2 & 2 & 0 & 8 \\ 0 & 0 & 0 & 5 & 1 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 & 5 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In other words, R is similar to a matrix in which entry $(3,7)$ is now zero, and that differs from A only in the entries in the same column as $(3,7)$ with smaller x-coordinate, or the same row as $(3,7)$ with larger y-coordinate.

Proof. Literally do what we did in the example, but with a general matrix. Specifically: take any matrix

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \ddots & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix},$$

and suppose that $a_{ii} \neq a_{jj}$. Consider the matrix $E = E_{\text{add } h \text{ copies of entry } j \text{ to entry } i}$. We know that E^{-1} is the matrix $E_{\text{add } -h \text{ copies of entry } j \text{ to entry } i}$, and thus that ERE^{-1} will be the matrix that has h copies of its j-th row added to its i-th row, and $-h$ copies of its i-th column added to its j-th column.

In particular, this means that the only entries that will be changed are those in the i-th row and j-th column; furthermore, the only entries that will be changed are those of the form (i,k), $j \leq k$ or (k,j), $k \leq i$; this is because the only entries that get changed when we add the i-th row or j-th column to something are those where that row/column are nonzero! Because our matrix is upper-triangular, this means that the only changed entries are precisely those in column i or row j and either above or to the right of (i,j).

Also, in particular, the entry in (i,j) is now $a_{ij} + ha_{jj} - ha_{ii}$. If we pick h such that $a_{ij} + h(a_{jj} - a_{ii}) = 0$ (which we can do whenever $a_{jj} - a_{ii} \neq 0$, then the cell $(i,j) = 0$! So we've proven our claim. $\square$

Lemma 2. *Suppose that R is an upper-triangular matrix. Then R is similar to another upper-triangular matrix R', such that*

- *R and R' share the same entries on their diagonal: in other words, $r_{ii} = r'_{ii}$.*
- *If r'_{ii} and r'_{jj} are distinct, with $i < j$, then the entry r'_{ij} of R is zero.*

This is essentially just repeated application of Lemma 1. Again, we illustrate our proof with an example before starting the proof itself. It's pretty long, so feel free to read through it until you feel like you get the idea!

Example. Consider the matrix

$$R = \begin{bmatrix} 1 & 4 & 4 & 0 & 14 & 1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We're going to repeatedly use Lemma 1, starting on the second column and working bottom-to-top through each column in order, until we get a matrix R' where whenever r'_{ii} and r'_{jj} are distinct, $i < j$, we have $r'_{ij} = 0$.

We start with $(1, 2)$. Because $r_{11} = 1 \neq r_{22} = 2$, Lemma 1 tells us to try multiplying on the left by $E = E_{\text{add } h \text{ copies of entry 2 to entry 1}}$ and on the right by $E^{-1} = E_{\text{add } -h \text{ copies of entry 2 to entry 1}}$. (In general, this process of multiplying on the left by E and on the right by E^{-1} is called **conjugating by** E.)

This yields (using the observations we've been making about how elementary matrix multiplication works) the matrix

$$\begin{bmatrix} 1 & 4 + 2h - 1h & 4 + 1h & 0 & 14 + 3h & 1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

If we set $h = -4$, this zeroes out the entry $(1, 2)$, and changes the entries $(1, 3)$ and $(1, 5)$ to 0, 2 respectively. In other words, we get the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We now move to the third column. The entry $(2,3)$ is nonzero, but $r_{33} = r_{22} = 1$, so we can't do anything about it, and the entry $(1,3)$ is already 0, so we don't do anything here as well.

In the four column, we can do a bit more. We start by looking at $(3,4)$, which is nonzero and features $r_{33} = 2 \neq r_{44} = 3$, our lemma suggests that we conjugate by $E = E_{\text{add } h \text{ copies of entry 4 to entry 3}}$. This gives us

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 2 & 1 & 0-1h & 3 & 0 \\ 0 & 0 & 2 & 1+3h-2h & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

To zero out entry $(3,4)$, we set $h = -1$. This gives us the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 2 & 1 & 1 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We proceed in this manner throughout the entire matrix. Our next entry in column 4that is nonzero and changeable is $(2,4)$, which is currently 1 and opposite $r_{22} = 2, r_{44} = 3$. Conjugating by $E = E_{\text{add } -1 \text{ copies of entry 4 to entry 2}}$ zeroes out this entry and doesn't change anything else in our matrix.

In column 5, the entries $(4,5), (3,5)$ are already 0, and the entry $(2,5)$ has $r_{22} = r_{55}$, so we can't do anything there. We can do something about $(1,5)$; once again, Lemma 1 says that if we conjugate by $E_{\text{add } -2 \text{ copies of entry 5 to entry 1}}$, we will zero out that entry. This results in the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2+2h-h & 1+h \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Finally, in column 6 we note that we can zero out entry $(5,6)$ by conjugating by $E_{\text{add 1 copy of entry 6 to entry 5}}$. This yields

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & 0-3h \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1+h-2h \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & -3 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We can also zero out entry $(2,6)$ by conjugating by $E_{\text{add } -3 \text{ copies of entry 6 to entry 2}}$. This yields

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & -3+h-2h \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The last nonzero entry in this column, $(1,6)$, is across from $a_{11} = a_{66} = 1$, so we can do nothing more.

So what have we done? We took R, and repeatedly conjugated it by elementary matrices until we got a matrix with our desired property (i.e. that whenever r'_{ii} and r'_{jj} are distinct, with $i < j$, we have $r'_{ij} = 0$.)

Proof. Our proof is precisely the same algorithm as described above, but in abstract. Take any upper-triangular matrix R. Starting with the second column of R and working left to right, run the following algorithm:

1. Suppose that we are currently on column j.

2. Starting from $(j-1, j)$ and working our way up the column, run Lemma 1 on each cell (i,j) that it can be ran on. This zeroes out (i,j) whenever $r_{ii} \neq r_{jj}$.

3. Notice that Lemma 1 in particular guarantees that the only cells changed by conjugating by E are those in the same column as (i,j) but in rows $k \leq i$, or in the same row as (i,j) but in columns $j \leq k$, this does not change any cells that we've already checked.

4. Once you are done with column j, go to the next column and repeat this process.

By Lemma 1, this creates a matrix that is similar to our original matrix R by some chain of elementary matrices, and has our desired property (i.e. whenever r'_{ii} and r'_{jj} are distinct, with $i < j$, we have $r'_{ij} = 0$.) $\square$

Lemma 3. *Suppose that R is an $n \times n$ upper-triangular matrix. Then R is similar to a block-diagonal matrix*

$$\left[\begin{array}{c|c|c|c} B_1 & 0 & \dots & 0 \\ \hline 0 & B_2 & \dots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \dots & B_k \end{array}\right],$$

where each block B_i is of the following form:

- *B_i is upper-triangular.*

- *The entries on the diagonal of B_i are all equal.*

Again, we illustrate our proof methods (basically just using Lemma 2 along with our results on how conjugating by a permutation matrix works) with an example before our proof:

Example. Consider the matrix R from before:

$$R = \begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

As we did in our earlier example, we can find a matrix R' that is similar to R, such that whenever $r'_{ii} \neq r'_{jj}$, $i < j$ we have $r'_{ij} = 0$:

$$R' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Now, what we want to do is somehow shuffle the diagonal of R' around, so that equal terms are collected! We can do this with a permutation matrix, P_σ. First, notice that because $P_\sigma^{-1} = P_\sigma^T$ as shown last week, we know that $P_\sigma^T R' P_\sigma$ is a similar matrix to R'. Furthermore, as we studied last week, multiplying a matrix on the left by P_σ^T just shuffles its rows by σ, and multiplying it on the right by P_σ just shuffles its columns by σ. Doing both, then, shuffles R's rows and columns both by σ! In other words, if we denote this matrix as C, it is a matrix such that $C(i,j) = R'(\sigma(i), \sigma(j))$.

So: let σ be some permutation of $(1, 2, 3, 4, 5, 6)$ that corresponds to ordering the diagonal of R' from smallest to greatest: i.e. pick some way σ to reorder the diagonal entries of R' such that $r'_{\sigma(1),\sigma(1)} \leq r'_{\sigma(2),\sigma(2)} \leq \ldots \leq r'_{\sigma(6),\sigma(6)}$. Furthermore, we have some diagonal elements that are identical; therefore, the above criteria doesn't fully specify what permutation we should pick. In this situation, pick our permutation such that it preserves the original ordering on those elements: in other words, if $r_{ii} = r_{jj}$ and $i < j$, we should have $\sigma(i) < \sigma(j)$.

If we want this property to hold for our R' in this problem, we should pick the permutation

$$\sigma : \begin{array}{|c|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & 5 & 6 \\ \hline \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \hline 1 & 6 & 2 & 3 & 5 & 4 \\ \hline \end{array}.$$

Then, if we look at $P_\sigma^T R' P_\sigma$, i.e. the matrix whose rows and columns are shuffled by σ as defined above, we get

$$C = P_\sigma^T R' P_\sigma = \left[\begin{array}{cc||ccc||c} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right].$$

This is in the form that we were looking for. Therefore, we have shown that our original matrix R is similar to a block diagonal matrix C with the properties requested!

Proof. Once again, our proof is just a generalization of our example above. Take any $n \times n$ upper-triangular matrix R. First, apply Lemma 2 to find some R' that is similar to R, such that whenever r'_{ii} and r'_{jj} are distinct, with $i < j$, we have $r_{ij} = 0$.

Now, let σ be some permutation of $(1, \dots n)$ that corresponds to ordering the diagonal of R' from smallest to greatest: i.e. pick some way σ to reorder the diagonal entries of R' such that $r'_{\sigma(1),\sigma(1)} \leq r'_{\sigma(2),\sigma(2)} \leq \ldots \leq r'_{\sigma(n),\sigma(n)}$. Furthermore, in the event that we have diagonal elements of R that are the same, we will have multiple different permutations that will order our diagonal elements. In this situation, pick our permutation such that it preserves the original ordering on those elements: in other words, if $r_{ii} = r_{jj}$ and $i < j$, we should have $\sigma(i) < \sigma(j)$.

Let P_σ be the permutation matrix corresponding to σ, and look at $(P_\sigma)^T R' P_\sigma$. This is the matrix that we get by permuting R''s rows and columns by σ! Notice that this matrix has the following properties:

1. The diagonal of this new matrix is the old diagonal of R', but permuted by σ: i.e. it's now sorted from least to greatest. In particular, this means that every repeated value on the diagonal of our matrix is now clumped together.

2. If we picked two diagonal elements r_{ii}, r_{jj} that were distinct, we knew that the cell across from them in R' was zero. This property does not go away when we permute rows or columns if we simply track where r_{ii}, r_{jj} move to! Therefore, our new matrix $(P_\sigma)^T R' P_\sigma$ still has this property.

3. Moreover, our new matrix is still upper-triangular. To see why, simply take any cell in the new matrix, with coordinates $(\sigma(i), \sigma(j))$. We want this cell to contain a zero if $\sigma(j) < \sigma(i)$, as these are precisely the cells with coordinates below the main diagonal.

   There are two possibilities: either

   (a) the diagonal entries $r_{\sigma(i),\sigma(i)}, r_{\sigma(j),\sigma(j)}$ are distinct, and therefore this cell contains a zero by property 2, or

   (b) these diagonal entries are the same. In this case, we know that if $\sigma(i) > \sigma(j)$, we constructed our permutation such that $i > j$ as well. But this means that our new matrix's cell $(\sigma(i), \sigma(j))$ came from a cell (i, j) in our original matrix with $i > j$. But these cells are precisely those that are below the main diagonal, and thus contain a 0!

So our matrix is upper-triangular.

4. If we combine properties 1 and 2 above, this tells us that our matrix is in block diagonal form! In particular, for each distinct value $\lambda_1 \leq \ldots \leq \lambda_k$ that shows up on the diagonal of $(P_\sigma)^T R' P_\sigma$, associate a block B_{λ_i}. Because the entries (i,j) across from distinct λ_i, λ_j are all zero from (2), these blocks are each in fact blocks, as nothing outside of them is zero.

5. Now look within any given $k_i \times k_i$ block B_i. These blocks, by construction, are upper-triangular with a repeated value λ_i on their diagonal. This is precisely what we wanted!

So we have proven our claim: any upper-triangular matrix is similar to a block-diagonal matrix where each block is upper-triangular, and each block's diagonal consists of a repeated value. $\square$

At this stage, we have something that is nearly in Jordan canonical form! We just have to slightly modify these blocks. We do this with the following lemma:

Lemma 4. *Suppose that A is a block-diagonal matrix with a block B of the following form:*

- *B_i is upper-triangular.*
- *The entries on the diagonal of B_i are all equal.*

Then we can conjugate A by appropriate elementary matrices so that B is replaced with a block-diagonal matrix made out of Jordan blocks, and no other blocks of A are changed. In other words, A is similar to a matrix in which block B is replaced with a block-diagonal matrix made out of Jordan blocks.

Again, we illustrate our proof methods with an example before our proof:

Example. Consider the following matrix C:

$$C = \left[\begin{array}{cc||ccc||c} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right].$$

We'll perform the process that Lemma 4 wants us to calculate an example of, on each of its blocks. The only tools we will use here are the following:

1. Conjugating a matrix by an elementary matrix of the form $E_{\text{multiply entry } k \text{ by } h}$, for nonzero h. Notice that doing this will scale all of the entries in row k by h and all of the entries in column k by $\frac{1}{h}$. In particular, note that this will not change the entry (k,k) on the diagonal that is in both that row and that column. Also note that this will only change entries in whichever block contains the cell (k,k), as no other blocks contain elements from the k-th row or column.

2. Conjugating a matrix by an elementary matrix of the form $E_{\text{add } h \text{ copies of entry } k \text{ to entry } l}$. As used many times in these notes, this will add h copies of row k to row l, and also add $-h$ copies of column l to column k. In particular, note that if we ask for $l < k$, this property preserves that our matrix is upper-triangular. Also note that if l, k come from within the same block, then the only entries changed are in that block, as no other blocks contain those rows/columns.

3. Conjugating a matrix by a permutation matrix P_σ, where the elements permuted by σ all correspond to the rows/columns of one specific block. This, by construction, only messes with one of our blocks, and if we're careful (as shown earlier) will preserve our upper-triangular structure as well.

We start with the top block $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$. We want this to become a Jordan block; i.e. we want that -1 to become a 1. This is not hard to do: just conjugate by the matrix $E = E_{\text{multiply entry 2 by } -1}$. This scales row 2 by -1 and column 2 by $1/-1 = -1$; i.e. we get

$$ECE^{-1} = \left[\begin{array}{cc||ccc||c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right].$$

So our first block is OK!

We move to our second block. This is tricker: we need to somehow make $(3,5)$ zero, even though the entries $(3,3)$ and $(5,5)$ are identical. How can we do this?

Here's one solution: instead of using the cells $(3,3),(5,5)$, which are both identical, use the cells $(3,4)$ and $(5,5)$ which are distinct! In particular, consider conjugation by the matrix $E_{\text{add } h \text{ copies of entry 5 to entry 4}}$. This will add h copies of row 5 to row 4 and then subtract h copies of column 4 from column 5; i.e. it will produce the matrix

$$\left[\begin{array}{cc||ccc||c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 3-1h & 0 \\ 0 & 0 & 0 & 2 & 0+2h-2h & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right] = \left[\begin{array}{cc||ccc||c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 3-1h & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right]$$

Notice that the entry in $(4,5)$ just gets canceled out, and all we have done to our matrix is give us a way to subtract h from $(3,5)$! In particular, if we set $h = 3$, we have made $(3,5)$ zero.

Consequently, we have now written our second block as a union of Jordan blocks. In other words, we have shown that our matrix is similar to the matrix

$$\left[\begin{array}{cc||cc||c||c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 2 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right],$$

This is a block-diagonal matrix made out of Jordan blocks!

Proof. Our proof here, like every proof in this lecture, is just a generalization of the methods used in our example.

First, notice that for the reasons explained in our example, if we limit ourselves to just using the tools

1. conjugating a matrix by an elementary matrix of the form $E_{\text{multiply entry } k \text{ by } h}$, for nonzero λ, and

2. conjugating a matrix by an elementary matrix of the form $E_{\text{add } h \text{ copies of entry } k \text{ to entry } l}$,

our actions will only change entries inside of one fixed block. So, for the remainder of this proof, we can simply pretend that we're working within a single block

$$B = \begin{bmatrix} \lambda & * & \dots & * \\ 0 & \lambda & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \end{bmatrix}.$$

We will turn B into a block-diagonal matrix consisting of Jordan blocks column-by-column, working left-to-right.

We start with the second column: in this case, we are looking at the 2×2 matrix given by the cells in B's first two rows and columns,

$$B_2 = \begin{bmatrix} \lambda & * \\ 0 & \lambda \end{bmatrix}$$

There are two possibilities here:

1. The entry $*$ in cell $(1, 2)$ is 0. In this case, B_2 is already a block-diagonal matrix consisting of the two 1×1 Jordan blocks $\boxed{\lambda}$.

2. The entry $*$ in cell $(1, 2)$ is not 0. In this case, conjugate our matrix by the elementary matrix $E = E_{\text{multiply entry 1 by } 1/*}$. This scales our first row by $1/*$ and our first column by $*$. In other words, the only cell that gets changed in our 2×2 matrix is $(1, 2)$, which gets scaled by $1/*$: i.e. it is replaced by 1! So our matrix is now in the form $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, and thus in particular is a Jordan block.

So we've succeeded in our "base case;" i.e. where we're just dealing with our second column. We now describe a process that, assuming we've succeeded in making the cells in the first k rows and columns into a block-diagonal matrix made out of Jordan blocks, will let us extend our result to the next row and column. (In a sense, you can regard this as a proof by induction, though I want to point out that we are describing an explicit algorithm that we can concretely use to make Jordan blocks.)

If we're assuming that we've succeeded on our first $k-1$ rows and columns and are now looking at the next column k, we're looking at the following $k \times k$ matrix:

$$B_k = \left[\begin{array}{cccccc|c} \lambda & \delta_1 & 0 & 0 & \dots & 0 & * \\ 0 & \lambda & \delta_2 & 0 & \dots & 0 & * \\ 0 & 0 & \lambda & \delta_3 & \dots & 0 & * \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & \delta_{k-2} & * \\ 0 & 0 & 0 & \dots & 0 & \lambda & * \\ 0 & 0 & 0 & \dots & 0 & 0 & \lambda \end{array}\right].$$

The δ_i's are all either 0's or 1's, because we've insured that the cells in the first $k-1$ rows/columns form a block diagonal matrix made out of Jordan blocks. The $*$'s are all values in the k-th column, that are arbitrary and that we don't know yet.

We deal with the entries in this k-th column with a few different techniques:

1. Suppose that there is some nonzero entry (i,k) in this last column, $(i,k) = t$, such that the entry in $(i, i+1)$ (i.e. δ_i) is 1. In other words, suppose that our square looks like the following:

$$B_k = \left[\begin{array}{ccccc|c} \lambda & \dots & 0 & 0 & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \lambda & 1 & \dots & t \\ 0 & \dots & 0 & \lambda & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \lambda \end{array}\right].$$

In this situation, our life is pretty easy. Just conjugate by the elementary matrix $E_{\text{add } t \text{ copies of entry } k \text{ to entry } i+1}$. This adds t copies of the k-th row of our matrix to the $i+1$-th row, and subtracts t copies of the $i+1$-th column of our matrix from the k-th column. In other words, it gives us the matrix

$$\left[\begin{array}{ccccc|c} \lambda & \dots & 0 & 0 & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \lambda & 1 & \dots & t-1t \\ 0 & \dots & 0 & \lambda & \dots & *-\lambda t+\lambda t \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \lambda \end{array}\right] = \left[\begin{array}{ccccc|c} \lambda & \dots & 0 & 0 & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \lambda & 1 & \dots & 0 \\ 0 & \dots & 0 & \lambda & \dots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \lambda \end{array}\right].$$

This is the same matrix as we had before, except (i,k) is now zero! Therefore, by conjugating our block by elementary matrices, we can zero out these kinds of cells.

2. This means that if we have any nonzero cells in our last column, there isn't a 1 in the cell $(i, i+1)$; i.e. we have $\delta_i = 0$, in our earlier notation for B_k.

   Consider first the case where there are two such values in our last column: i.e. we have cells $(i,k), (j,k)$ such that $(i,k) = s \neq 0$, $(j,k) = t \neq 0$, and $\delta_i = \delta_j = 0$:

$$B_k = \left[\begin{array}{cccccccc|c}
\lambda & \dots & 0 & 0 & \dots & 0 & 0 & \dots & * \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & \lambda & 0 & \dots & 0 & 0 & \dots & s \\
0 & \dots & 0 & \lambda & \dots & 0 & 0 & \dots & * \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & \lambda & 0 & \dots & t \\
0 & \dots & 0 & 0 & \dots & 0 & \lambda & \dots & * \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & \lambda
\end{array}\right].$$

   Take the first $k-1$ rows/columns of B_k. These form a block-diagonal matrix whose blocks are Jordan blocks! In particular, notice that because the entries $\delta_i, \delta_j = 0$, these Jordan blocks do not contain the cells $(i, i+1), (j, j+1)$. Therefore, if we look at B_k and draw in the Jordan blocks that contain the cells $(i,i), (j,j)$, our matrix is actually in the following form:

$$B_k = \left[\begin{array}{cccccccc|c}
\lambda & & & & & & & & * \\
 & \ddots & & & & & & & \\
 & & \boxed{J_s} & & & & & & \vdots \\
 & & & 0 & & & & & s \\
 & & & \lambda & & & & & \vdots \\
 & & & & \ddots & & & & \\
 & & & & & \boxed{J_t} & & & \vdots \\
 & & & & & & 0 & & t \\
 & & & & & & \lambda & & \vdots \\
 & & & & & & & \ddots & \\
 & & & & & & & & \lambda
\end{array}\right].$$

   Assume that J_s is a $k_s \times k_s$ matrix and J_t is a $k_t \times k_t$ matrix. Assume without loss of generality that $k_s \geq k_t$ (the opposite case will have an identical proof to what we do here, with s and t's roles exchanged.) In this situation, perform the following sequence of conjugations:

   - First, conjugate by $E_{\text{add } -s/t \text{ copies of entry } j \text{ to entry } i}$.
   - Then, conjugate by $E_{\text{add } -s/t \text{ copies of entry } j-1 \text{ to entry } i-1}$.

- Then, conjugate by $E_{\text{add } -s/t \text{ copies of entry } j-2 \text{ to entry } i-2}$.

$\vdots$

- Then, conjugate by $E_{\text{add } -s/t \text{ copies of entry } j-k_s \text{ to entry } i-k_s}$.

Claim: this makes entry $(j,k) = 0$ without changing the values of any other cells in our matrix! To see this, simply draw a few stages in our process. After the first conjugation by $E_{\text{add } -s/t \text{ copies of entry } j \text{ to entry } i}$, our matrix (when zoomed in on the appropriate rows/columns) looks like

$$
\begin{bmatrix}
\ddots & & & & & & & & & & \\
& \overbrace{\begin{array}{|ccccc|} \hline \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & & \\ & & & \lambda & 1 \\ & & & & \lambda \\ \hline \end{array}}^{J_s} & & & & s/t & & \\
& & & & & (s/t)\lambda-(s/t)\lambda & \dots & s-(s/t)t \\
& & \ddots & & & & & \\
& & & \overbrace{\begin{array}{|ccccc|} \hline \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & & \\ & & & \lambda & 1 \\ & & & & \lambda \\ \hline \end{array}}^{J_t} & & & \dots & t \\
& & & & & & \ddots &
\end{bmatrix}
$$

Note that the contents of cells $(i,j),(i,k)$ are both 0 above, and the only nonzero cell outside of the last column and our blocks is the s/t in cell $(i-1,j)$.

After performing the second conjugation by $E_{\text{add } -s/t \text{ copies of entry } j-1 \text{ to entry } i-1}$, we have

$$
\begin{bmatrix}
\ddots & & & & & & & & & \\
& \overbrace{\begin{array}{|ccccc|} \hline \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & & \\ & & & \lambda & 1 \\ & & & & \lambda \\ \hline \end{array}}^{J_s} & & s/t & & & \\
& & & (s/t)\lambda-(s/t)\lambda & (s/t)-(s/t)1 & & \\
& & & & 0 & \dots & 0 \\
& & \ddots & & & & \\
& & & \overbrace{\begin{array}{|cccccc|} \hline \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \ddots & & & \\ & & & \lambda & & 1 \\ & & & & & \lambda \\ \hline \end{array}}^{J_t} & & \dots & t \\
& & & & & \ddots &
\end{bmatrix}
$$

Note that the contents of cells $(i-1, j), (i-1, j-1)$ are both 0 above, and furthermore that we've moved our s/t exactly one cell up and to the left to $(i-2, j-1)$.

Just to hammer the point home, look at the third conjugation by $E_{\text{add } -s/t \text{ copies of entry } j-2 \text{ to entry } i-2}$:

$$\begin{bmatrix}
\ddots & & & & & & & & & & & & & \\
& \lambda & 1 & & & & & & & & & & & \\
& & \lambda & 1 & & & & & & s/t & & & & \\
& & & \ddots & & & & & & (s/t)\lambda-(s/t)\lambda & (s/t)-(s/t) & & & \\
& & & & \lambda & 1 & & & & & 0 & 0 & & \\
& & & & & \lambda & & & & & & 0 & \dots & 0 \\
& & & & & & \ddots & & & & & & & \\
& & & & & & & \lambda & 1 & & & & & \\
& & & & & & & & \lambda & 1 & & & & \\
& & & & & & & & & \ddots & & & & \\
& & & & & & & & & & \lambda & 1 & & \\
& & & & & & & & & & & \lambda & \dots & t \\
& & & & & & & & & & & & \ddots &
\end{bmatrix}$$

Again, the cells $(i-2, j-1), (i-2, j-2)$ above are both 0, and we've again moved our s/t one cell up and to the left to the cell $(i-3, j-2)$.

So: we keep doing this! After each conjugation, we are effectively just moving the s/t cell up one and to the left one; we can do this as long as we have rows of J_t and columns of J_s to combine! In particular, because there are more rows of J_t than columns of J_s by assumption, we can do this up until we get to our last conjugation $E_{\text{add } -s/t \text{ copies of entry } j-k_s \text{ to entry } i-k_s}$. At this stage, we start with a s/t in cell $(i-k_s, j+1-k_s)$, and our conjugation gives us the following:

$$\begin{bmatrix}
\ddots & & & & & & & & & & & & & \\
& \lambda & 1 & & & & & (s/t)\lambda-(s/t)\lambda & (s/t)-(s/t) & & & & & \\
& & \lambda & 1 & & & & & & & & & & \\
& & & \ddots & & & & & & \ddots & \ddots & & & \\
& & & & \lambda & 1 & & & & & 0 & 0 & & \\
& & & & & \lambda & & & & & & 0 & \dots & 0 \\
& & & & & & \ddots & & & & & & & \\
& & & & & & & \lambda & 1 & & & & & \\
& & & & & & & & \lambda & 1 & & & & \\
& & & & & & & & & \ddots & & & & \\
& & & & & & & & & & \lambda & 1 & & \\
& & & & & & & & & & & \lambda & \dots & t \\
& & & & & & & & & & & & \ddots &
\end{bmatrix}$$

In particular, notice that when we add s/t copies of the $i-k_s$-th column, we're only adding a λ, because we're at the end of the block J_s (and therefore there's a 0 above the λ at $(i-k_s, i-k_s)$.)

So, what's the net result of all of these conjugations? We've zeroed out the entry (i,k), and not changed any of the other entries in our matrix!

3. Therefore, by 2, if there are two nonzero entries still left in the last column, we can repeatedly get rid of one of them until we have at most one nonzero cell (i,k) in the last column. If there are no nonzero entries, we are done! Otherwise there is exactly one left. We know that by 1 this nonzero entry is not across from a 1 in $(i,i+1)$: in other words, our square looks like the following:

$$B_k = \left[\begin{array}{cccccc|c} \lambda & \dots & 0 & 0 & \dots & & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & & \vdots \\ 0 & \dots & \lambda & 0 & \dots & & t \\ 0 & \dots & 0 & \lambda & \dots & & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & & \lambda \end{array}\right].$$

Conjugate by the following pair of matrices:

- First, conjugate by $E_{\text{multiply entry } k \text{ by } t}$. This scales row k by t and column k by $1/t$: i.e. it scales (i,k) to 1, and doesn't change anything else.
- Finally, pick out some permutation σ that cycles the values $i+1, \dots k$ forward by 1 (i.e. it sends k to $i+1$, $i+1$ to $i+2$, …and $k-1$ to k.) Then conjugating by P_σ just permutes the rows and columns of B_k in the same way: i.e. it uses the k-th row and column as its $i+1$-th row and column, and shuffles the other rows and columns forward by one. But what is this matrix? It's mostly just the same matrix as B_k was before, except now the 1-cell (i,k) is in $(i,i+1)$. In other words, our matrix is now in Jordan normal form!

So we're done! $\square$

Finally, if we combine all of our results together, we get our claimed proof:

Theorem. Any $n \times n$ matrix A can be written in Jordan canonical form.

Proof. Simply use all of our results at once! I.e.

1. The Schur decomposition tells us that any $n \times n$ matrix A is similar to an upper-triangular matrix.
2. Lemma 2 lets us say that any upper-triangular matrix is similar to a block-diagonal matrix in which each block is upper-triangular and has the same values on its diagonal.
3. Repeatedly applying Lemma 4 to any matrix of the above form tells us that any matrix of the above form is similar to some matrix in Jordan canonical form.

Therefore, by simply chaining our similarity results together, we have that any $n \times n$ matrix A is similar to some other matrix in Jordan canonical form. $\square$

So we're done! We've proven one of the strongest classification theorems in linear algebra, and furthermore done so in a beautifully concrete and algorithmic fashion that we could easily implement on a computer. A nice place to end our course, right?