Math 108B — Professor: Padraic Bartlett

Lecture 7: Similar/Elementary Matrices; Jordan Canonical Form

Week 8 — *UCSB 2014*

Over the last two weeks of lecture, we proved the **spectral theorem**, which stated that any real-valued symmetric matrix A can be written as the product UDU^T, for some real-valued diagonal matrix D and orthogonal real-valued matrix U. This was a remarkably strong result as contrasted with the Schur decomposition, which stated that any arbitrary matrix A can be written in the form URU^*, for some upper-triangular R and unitary U.

Over the next two weeks, we are going to attempt to study another decomposition result, namely the Jordan canonical form! In next week's talks we will explicitly state what the Jordan canonical form is. For now, however, consider the following question: suppose we take an arbitrary matrix A. What kinds of matrices B can we find such that $A = UBU^{-1}$, for some U? We know that if we restrict U to be unitary, the best we can get is upper-triangular (Schur!), but what if we weaken our restrictions on U? Can we make B into something nicer? How much nicer?

With this desire in mind, we introduce our first topic of the week: similar matrices!

1 Similar Matrices

We have been working with the concept of **similarity** for essentially this entire course, so this definition should look fairly familiar:

Definition. Two matrices A, B are called **similar** if there is some matrix U such that $A = UBU^{-1}$. If we want to specify what U is, we can specifically state that A and B are similar **via** U.

To illustrate this definition, we rephrase some of the results we've recently proven using this language of similarity:

Theorem. (Spectral theorem.) Any real-valued symmetric matrix is similar to a real-valued diagonal matrix via a real-valued orthogonal matrix.

Theorem. (Schur decomposition.) Any $n \times n$ matrix A is similar to an upper-triangular matrix via a unitary matrix.

Theorem. The matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not similar to any diagonal matrix D.

Notice that similarity is an **equivalence relation**: i.e.

1. **Reflexivity**: any matrix A is similar to itself, because $A = IAI^{-1}$.
2. **Symmetry**: If A is similar to B via a matrix U, then B is similar to A via the matrix U^{-1}; this is because $A = UBU^{-1} \Rightarrow B = U^{-1}AU$.

3. **Transitivity**: If A is similar to B via a matrix U, and B is similar to C via a matrix V, then A is similar to C; this is because we can write $A = UBU^{-1} = UVCV^{-1}U^{-1} = (UV)C(UV)^{-1}$.

This is a nice property! In particular, to show that two matrices A, B are similar we will often just show that they are both similar to some matrix C, and use symmetry and transitivity to conclude that A and B are similar to each other.

In this language, the question we mentioned at the start of class is the following: given an arbitrary $n \times n$ matrix A, what "nice" kinds of matrices can we find that are similar to A? The Schur decomposition tells us that A is similar to an upper-triangular matrix: can we improve this to something nicer than upper-triangular?

Working with general matrices is a pain; they can do lots of awful things. Instead, let's work with upper-triangular matrices! Because any matrix is similar to an upper-triangular matrix, if we can find some "nicer" form that upper-triangular matrices are similar to, then (by transitivity) every matrix will be similar to that "nicer" form!

This simplifies things somewhat. However, we can simplify things more: instead of looking at similarities via arbitrary matrices, why not look at similarity via particularly nice and easy-to-understand matrices? In particular, recall the concept of **elementary matrices**, from last quarter:

2 Review: Elementary Matrices

Definition. The first matrix, $E_{\text{multiply entry k by }\lambda}$, is the matrix corresponding to the linear map that multiplies its k-th coördinate by λ and does not change any of the others:

$$E_{\text{multiply entry } k \text{ by } \lambda} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \lambda & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

This matrix has 1's down its diagonal and 0's elsewhere, with an exception for the value at (k, k), which is λ.

The second matrix, $E_{\text{switch entry } k \text{ and entry } l}$, corresponds to the linear map that swaps its k-th coördinate with its l-th coördinate, and does not change any of the others:

$$E_{\text{switch entry } k \text{ and entry } l} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$

You can create this matrix by starting with a matrix with 1's down its diagonal and 0's elsewhere, and switching the k-th and l-th columns.

Finally, the third matrix, $E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l}$, for $k \neq l$, corresponds to the linear map that adds λ copies of its k-th coördinate to its l-th coördinate and does not change any of the others:

$$E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l} = \begin{bmatrix} 1 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & \lambda & 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix}$$

This matrix has 1's down its diagonal and 0's elsewhere, with an exception for the value in row l, column k, which is λ.

These three matrices are called the **elementary** matrices, if you don't remember them from the previous quarter of 108.

As a reminder, this is how elementary matrices interact with matrix multiplication:

Theorem 1. *Take any $n \times n$ matrix A. Suppose that we are looking at the products $E \cdot A$, $A \cdot E$, where E is one of our elementary matrices. Then, we have the following three possible situations:*

- *If $E = E_{\text{multiply entry } k \text{ by } \lambda}$, then $E \cdot A$ would be the matrix A with its k-th row multiplied by λ, and $A \cdot E$ is the matrix A with its k-th columns multiplied by λ.*
- *If $E = E_{\text{switch entry } k \text{ and entry } l}$, then $E \cdot A$ would be the matrix A with its k-th and l-th rows swapped, and $A \cdot E$ is the matrix A with its k-th and l-th columns swapped.*
- *If $E = E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l}$, then $E \cdot A$ would be the matrix A with λ copies of its k-th row added to its l-th row, and $A \cdot E$ is a matrix A with λ copies of its l-th column added to its k-th column.*

Finally, we also calculated the inverses of these matrices:

Theorem 2. *Take any $n \times n$ elementary matrix E. Then, we have the following:*

- *If $E = E_{\text{multiply entry } k \text{ by } \lambda}$, then $E^{-1} = E_{\text{multiply entry } k \text{ by } \frac{1}{\lambda}}$, whenever $\lambda \neq 0$.*
- *If $E = E_{\text{switch entry } k \text{ and entry } l}$, then $E^{-1} = E$; i.e. E is its own inverse.*
- *If $E = E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l}$, then $E^{-1} = E_{\text{add } -\lambda \text{ copies of entry } k \text{ to entry } l}$.*

We calculate a few quick examples to illustrate what's going on here:

Example. Suppose that A is some arbitrary 5×5 matrix. Calculate EAE^{-1}, where E is one of the following matrices:

- $E_{\text{multiply entry 2 by }\lambda}$, for $\lambda \neq 0$.
- $E_{\text{switch entry 1 and entry 4}}$.
- $E_{\text{add }\lambda\text{ copies of entry 2 to entry 5}}$

Proof. We simply calculate, starting first with $E_{\text{multiply entry 2 by }\lambda}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/\lambda & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \lambda a_{21} & \lambda a_{22} & \lambda a_{23} & \lambda a_{24} & \lambda a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/\lambda & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12}/\lambda & a_{13} & a_{14} & a_{15} \\ \lambda a_{21} & a_{22} & \lambda a_{23} & \lambda a_{24} & \lambda a_{25} \\ a_{31} & a_{32}/\lambda & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42}/\lambda & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52}/\lambda & a_{53} & a_{54} & a_{55} \end{bmatrix}.$$

Note that this is just the original matrix A, with its second row scaled by λ and its second column scaled by $1/\lambda$; furthermore, note that the entry $(2, 2)$ that is in both this row and this column is unchanged after this similarity transformation.

We now consider $E_{\text{switch entry 1 and entry 4}}$:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{44} & a_{42} & a_{43} & a_{41} & a_{45} \\ a_{24} & a_{22} & a_{23} & a_{21} & a_{25} \\ a_{34} & a_{32} & a_{33} & a_{31} & a_{35} \\ a_{14} & a_{12} & a_{13} & a_{11} & a_{15} \\ a_{54} & a_{52} & a_{53} & a_{51} & a_{55} \end{bmatrix}.$$

Note that this is the original matrix A with its 1st and 4th rows swapped, as well as its 1st and 4th columns.

We finally consider $E_{\text{add } \lambda \text{ copies of entry 2 to entry 5}}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \lambda & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ \lambda a_{21} + a_{51} & \lambda a_{22} + a_{52} & \lambda a_{21} + a_{53} & \lambda a_{21} + a_{54} & \lambda a_{21} + a_{55} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} - \lambda a_{15} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} - \lambda a_{25} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} - \lambda a_{35} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} - \lambda a_{45} & a_{43} & a_{44} & a_{45} \\ \lambda a_{21} + a_{51} & \lambda a_{22} + a_{52} - \lambda a_{55} & \lambda a_{21} + a_{53} & \lambda a_{21} + a_{54} & \lambda a_{21} + a_{55} \end{bmatrix}.$$

Note that this is just the original matrix A with λ copies of its 2nd row added to its 5th row, and $-\lambda$ copies of its 5th column added to its 2nd column. □

3 The Interactions Between Elementary Matrices and Similarity

The results above from our example hold in general:

Proposition. Suppose that E is an elementary matrix of the form $E_{\text{multiply entry k by } \lambda}$, for $\lambda \neq 0$, and A is an $n \times n$ matrix. Then EAE^{-1} is the matrix A with its k-th row scaled by λ and k-th column scaled by $\frac{1}{\lambda}$. In particular, note that the entry $A(k,k)$ is unchanged, as it is in both the k-th row and k-th column.

Proposition. Suppose that E is an elementary matrix of the form $E_{\text{switch entry } k \text{ and entry } l}$, and A is an $n \times n$ matrix. Then EAE^{-1} is the matrix A with its k-th and l-th rows swapped, as well as its k-th and l-th columns.

Proposition. Suppose that E is an elementary matrix of the form $E_{\text{add } \lambda \text{ copies of entry } k \text{ to entry } l}$, and A is an $n \times n$ matrix. Then EAE^{-1} is the matrix A, with λ copies of its k-th row added to its l-th row, and $-\lambda$ copies of its l-th column added to its k-th column.

Note that on the HW, you studied this product! In particular, you proved that whenever A is upper-triangular and $l < k$, then EAE^{-1} is still an upper-triangular matrix.

The second of these properties can be generalized in a fairly beautiful way:

4 Permutation Matrices and Similarity

Definition. A **permutation** σ of the list $(1, \ldots n)$ is simply some way to reorder this list into some other $(\sigma(1), \ldots \sigma(n))$. (If you prefer to think about functions, σ is simply a bijection from $\{1, \ldots n\}$ to $\{1, \ldots n\}$.)

Given a permutation σ, there is clearly some way to "undo" σ and return our list to its original state. This "unshuffling" is itself a permutation, as it gives us a way to reorder the elements $(1, \ldots n)$; we denote it by σ^{-1}. (Again, if you prefer functions, σ^{-1} is jus the inverse of the map σ.)

For example, the map

$$\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$$

is a permutation of the list $(1, 2, 3)$. The inverse of this map is easy to construct; just "undo" the above, i.e. set

$$\sigma^{-1}(2) = 1, \sigma^{-1}(3) = 2, \sigma^{-1}(1) = 3.$$

Definition. Given any permutation σ of $(1, \ldots n)$, the **permutation matrix** P_σ is simply the $n \times n$ matrix whose i-th column is given by $e_{\sigma(i)}$. In other words,

$$P_\sigma = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{e_{\sigma(1)}} & \vec{e_{\sigma(2)}} & \ldots & \vec{e_{\sigma(n)}} \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

For example, if we use the permutation σ on the list $(1, 2, 3)$ that we defined above, we have

$$P_\sigma = \begin{bmatrix} \vdots & \vdots & \vdots \\ \vec{e_{\sigma(1)}} & \vec{e_{\sigma(2)}} & \vec{e_{\sigma(3)}} \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \vec{e_2} & \vec{e_3} & \vec{e_1} \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Proposition. If σ is a permutation of $(1, \ldots n)$ with inverse σ^{-1}, then P_σ is an invertible matrix with inverse P_{σ^T}.

Proof. The rows of P_σ form an orthonormal basis for $\mathbb{R}^n$, because they're all just $\vec{e_i}$'s but reordered. Therefore P_σ is an orthonormal matrix, and we know that orthonormal matrices have inverses given by their transposes! $\square$

These matrices interact with arbitrary matrices in the following way:

Proposition. Take an arbitrary $n \times n$ matrix A, and an arbitrary permutation σ of $(1, \ldots n)$. Then AP_σ is just the matrix A, but with its columns permuted by σ.

Proof. Just look at the product:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{e_{\sigma(1)}} & \vec{e_{\sigma(2)}} & \dots & \vec{e_{\sigma(n)}} \\ \vdots & \vdots & & \vdots \end{bmatrix}.$$

An entry (i, j) in this product is simply the dot product of the row $(a_{i1}, \dots a_{in})$ of A with the j-th column $\vec{e_{\sigma(j)}}$ of P_σ. This is precisely $a_{i,\sigma(j)}$! In other words, the entry in (i, j) of our product is just the entry in $(i, \sigma(j))$ of A itself. This just means that that AP_σ is the matrix A with its columns permuted by σ, as claimed. □

Proposition. Take an arbitrary $n \times n$ matrix A, and an arbitrary permutation σ of $(1, \dots n)$. Then $P_\sigma^T A$ is just the matrix A, but with its rows permuted by σ.

Proof. Again, just look at the product:

$$\begin{bmatrix} \dots & \vec{e_{\sigma(1)}} & \dots \\ \dots & \vec{e_{\sigma(2)}} & \dots \\ & \vdots & \\ \dots & \vec{e_{\sigma(n)}} & \dots \end{bmatrix} \cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}.$$

An entry (i, j) in this product is simply the dot product of the row $e_{\sigma(i)}$ of P_σ^T with the j-th column $(a_{1j}, \dots a_{nj})$ of A. This is precisely $a_{\sigma(i),j}$! In other words, the entry in (i, j) of our product is just the entry in $(\sigma(i), j)$ of A itself. This just means that that $P_\sigma^T A$ is the matrix A with its rows permuted by σ, as claimed. □

The reason we care about this is the following proposition. It's kinda complicated, but the result we get from it is really useful:

Proposition. Suppose that A is a $n \times n$ matrix with the following form:

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

In the above matrix, the matrix B_1 is a $l_1 \times l_1$ matrix, B_2 is a $l_2 \times l_2$ matrix, and in general B_i is a $l_i \times l_i$ matrix, with the constants $l_1, \dots l_k$ chosen so that $l_1 + \dots + l_k = n$. The rest of the $n \times n$ matrix A is filled in with 0's, as indicated.

Consider the permutation $\sigma_{a,b}$ on $(1, \dots n)$ defined as follows:

$$\sigma_{a,b}(x) = x, \text{ if } x \leq a.$$
$$\sigma_{a,b}(x) = x + b, \text{ if } a < x \leq n - b.$$
$$\sigma_{a,b}(x) = x + 2b - n - 1, \text{ if } n - b < x \leq n.$$

In other words, $\sigma_{a,b}$ fixes the first a elements of $(1, \dots n)$, and then cycles the remaining elements of $(1, \dots n)$ by b. For example, $\sigma_{2,3}$ can be thought of as a permutation of $(1, 2, 3, 4, 5, 6, 7)$ such that

$$
\begin{aligned}
&\sigma(1) = 1, \sigma(2) = 2, \sigma(3) = 3 + 3 = 6, \sigma(4) = 4 + 3 = 7, \sigma(5) = 5 + 2 \cdot 3 - 7 - 1 = 3,\\
&\sigma(6) = 6 + 2 \cdot 3 - 7 - 1 = 4, \sigma(7) = 7 + 2 \cdot 3 - 7 - 1 = 5: \text{ in other words,}\\
\Rightarrow &(1, 2, 3, 4, 5, 6, 7) \xrightarrow{\sigma} (1, 2, 6, 7, 3, 4, 5).
\end{aligned}
$$

Pick any integer i, and let $a = l_1 + \dots l_{i-1}, b = l_k$. Then

$$
P^T_{\sigma_{a,b}} A P_{\sigma_{a,b}} = \left[\begin{array}{c|c|c||c|c|c|c|c}
B_1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \hline
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline
0 & \dots & B_{i-1} & 0 & 0 & 0 & \dots & 0 \\ \hline\hline
0 & \dots & 0 & B_k & 0 & 0 & \dots & 0 \\ \hline
0 & \dots & 0 & 0 & B_i & 0 & \dots & 0 \\ \hline
0 & \dots & 0 & 0 & 0 & B_{i+1} & \dots & 0 \\ \hline
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline
0 & \dots & 0 & 0 & 0 & 0 & \dots & B_{k-1}
\end{array}\right].
$$

In other words, $P^T_{\sigma_{a,b}} A P_{\sigma_{a,b}}$ is the same original matrix as A, except the "blocks" $B_i, \dots B_k$ have been "cycled through:" i.e. block B_k is where B_i used to be, B_i is where B_{i+1} was, B_{i+1} is where B_{i+2} was, and so on/so forth.

Proof. Again, just look at the product! In particular, start with $P^T_\sigma A$:

$$
P^T_{\sigma_{a,b}} A = \begin{bmatrix}
\dots & \vec{e_{\sigma(1)}} & \dots \\
\dots & \vec{e_{\sigma(2)}} & \dots \\
 & \vdots & \\
\dots & \vec{e_{\sigma(n)}} & \dots
\end{bmatrix} \cdot \left[\begin{array}{c|c|c||c|c|c|c|c}
B_1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \hline
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline
0 & \dots & B_{i-1} & 0 & 0 & 0 & \dots & 0 \\ \hline\hline
0 & \dots & 0 & B_i & 0 & 0 & \dots & 0 \\ \hline
0 & \dots & 0 & 0 & B_{i+1} & 0 & \dots & 0 \\ \hline
0 & \dots & 0 & 0 & 0 & B_{i+2} & \dots & 0 \\ \hline
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline
0 & \dots & 0 & 0 & 0 & 0 & \dots & B_k
\end{array}\right]
$$

As noted before, this is the matrix A but with its rows permuted by σ. But what does this mean? Well, σ does nothing to the first $l_1 + \dots l_{i-1}$ rows, so the blocks $B_1, \dots B_{i-1}$ are untouched. After that, σ shifts the remaining rows by l_k forward: i.e. it sends A to the following matrix:

$$\left[\begin{array}{c|c|c||c|c|c|c|c} B_1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ \hline \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline 0 & \ldots & B_{i-1} & 0 & 0 & 0 & \ldots & 0 \\ \hline\hline 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & B_k \\ \hline 0 & \ldots & 0 & B_i & 0 & \ldots & 0 & 0 \\ \hline 0 & \ldots & 0 & 0 & B_{i+1} & \ldots & 0 & 0 \\ \hline \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & \ldots & 0 & 0 & 0 & \ldots & B_{k-1} & 0 \end{array}\right]$$

Now, what happens if we multiply this on the right by P_σ? Well, we know that multiplying by P_σ on the right just permutes columns by σ. Again, what does this mean? Well, σ does nothing to the first $l_1 + \ldots l_{i-1}$ columns, so the blocks $B_1, \ldots B_{i-1}$ are again untouched. After that, σ shifts the remaining columns by l_k forward: i.e. we get

$$P^T_{\sigma_{a,b}} A P_{\sigma_{a,b}} = \left[\begin{array}{c|c|c||c|c|c|c|c} B_1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ \hline \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline 0 & \ldots & B_{i-1} & 0 & 0 & 0 & \ldots & 0 \\ \hline\hline 0 & \ldots & 0 & B_k & 0 & 0 & \ldots & 0 \\ \hline 0 & \ldots & 0 & 0 & B_i & 0 & \ldots & 0 \\ \hline 0 & \ldots & 0 & 0 & 0 & B_{i+1} & \ldots & 0 \\ \hline \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & B_{k-1} \end{array}\right].$$

But this is what we claimed! So we're done with our proof. $\square$

The main reason we care about this proof is that it gives us the following remarkably useful corollary:

Corollary 3. *Suppose that A is a $n \times n$ matrix with the following form:*

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

Take any other matrix C that consists of the same blocks on the diagonal, except in some other order: i.e.

$$C = \left[\begin{array}{c|c|c|c} B_{\sigma(1)} & 0 & \ldots & 0 \\ \hline 0 & B_{\sigma(2)} & \ldots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \ldots & B_{\sigma(k)} \end{array}\right],$$

for some permutation σ of $(1, \ldots k)$. Then A and C are similar.

Proof. Repeatedly apply our earlier proposition to move the blocks of A around using the $P_{\sigma_{a,b}}$ matrices. Since we performed these switches by multiplying on the left by P_σ^T and on the right by this matrix's inverse, P_σ, these switches preserve similarity. So we can just perform whatever sequence of switches is necessary to turn A into C, and we're done! $\square$

For notational convenience, make the following definition:

Definition. If A is a matrix in the form

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

call A a **block-diagonal** matrix.

In next week's classes, we will prove that all matrices are similar to something called their Jordan Canonical Form, using only three tools:

- The Schur decomposition.
- The lemmas about how conjugating by an elementary matrix changes a matrix.
- The result above about how we can "rearrange" the blocks of a block-diagonal matrix while preserving similarity.