Math 108B | Professor: Padraic Bartlett

Lecture 5: The Schur Decomposition

Week 5 | *UCSB 2014*

Repeatedly through the past three weeks, we have taken some matrix A and written A in the form

$$A = UBU^{-1},$$

where B was a diagonal matrix, and U was a change-of-basis matrix.

However, on HW #2, we saw that this was not always possible: in particular, you proved in problem 4 that for the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there was no possible basis under which A would become a diagonal matrix: i.e. you proved that there was no diagonal matrix D and basis $B = \{(b_{11}, b_{21}), (b_{12}, b_{22})\}$ such that

$$A = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \cdot D \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}^{-1}.$$

This is a bit of a shame, because diagonal matrices (for reasons discussed earlier) are pretty fantastic: they're easy to raise to large powers and calculate determinants of, and it would have been nice if every linear transformation was diagonal in some basis. So: what now? Do we simply assume that some matrices cannot be written in a "nice" form in any basis, and that we should assume that operations like matrix exponentiation and finding determinants is going to just be awful in many situations?

The answer, as you may have guessed by the fact that these notes have more pages after this one, is no! In particular, while diagonalization[1] might not always be possible, there is something fairly close that is - the **Schur decomposition**.

Our goal for this week is to prove this, and study its applications. To do this, we need one quick deus ex machina:

Theorem. Suppose that V is a n-dimensional vector space over $\mathbb{C}$, and T is a linear transformation from $V \to V$. Then T has a complex-valued eigenvalue with corresponding nontrivial eigenvector: i.e. there is some vector $\vec{v} \neq \vec{0} \in V$ such that $T(\vec{v}) = \lambda\vec{v}$.

If we find a basis B for V, and write T as a matrix over the basis B and $\vec{v}$ as a vector in the base B, this is equivalent to the following theorem:

Theorem. Suppose that T is a complex-valued $n \times n$ matrix. Then T has a complex-valued eigenvalue with corresponding nontrivial eigenvector: i.e. there is some vector $\vec{v} \neq \vec{0} \in \mathbb{C}^n$ such that $T(\vec{v}) = \lambda\vec{v}$.

We aren't going to prove this theorem in this course, because it essentially boils down to a very large theorem from complex analysis that is very far outside of the aim of this course: the **fundamental theorem of algebra**[2] Talk to me if you want to know how this works!

Instead, we're going to study applications of this result:

[1] We say that a linear transformation is **diagonalizable** if it can be written as a diagonal matrix in some basis.

[2]

1 The Schur Decomposition

The Schur decomposition is the following result:

Theorem. (Schur decomposition): For any $n \times n$ matrix A with entries from $\mathbb{C}$, there is some orthonormal basis B of $\mathbb{C}$ and some upper-triangular[3] matrix R with entries in $\mathbb{C}$.

$$A = \begin{bmatrix} \vdots & & \vdots \\ \vec{b_1} & \dots & \vec{b_n} \\ \vdots & & \vdots \end{bmatrix} \cdot R \cdot \begin{bmatrix} \vdots & & \vdots \\ \vec{b_1} & \dots & \vec{b_n} \\ \vdots & & \vdots \end{bmatrix}^{-1}.$$

In other words, for any $n \times n$ complex-valued matrix there is an orthonormal basis in which that matrix is upper-triangular!

We prove this theorem here, provide an example of such a decomposition, and finally use this decomposition to calculate something that would otherwise be fairly difficult!

First, the proof:

Proof. We proceed in four stages.

1. First, find an eigenvalue λ_1 of A. We are guaranteed that some such λ_1 exists, by our earlier result.

2. Now, let E_λ denote the set of all vectors $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$. This is a subspace of $\mathbb{C}^n$, as linear combinations of vectors in this space are still in this space. Therefore, it has an orthonormal basis! Pick some orthonormal basis $\vec{b_{1\lambda_1}}, \dots \vec{b_{k\lambda_1}}$ for this space.

3. Now, extend this basis to an orthonormal basis for all of $\mathbb{C}^n$! This is easy to do: one by one, pick a vector not in the span of our basis, run Gram-Schmidt on that vector to make it orthogonal to everything in our basis, and add in this new orthogonal vector $\vec{c_i}$ to our basis. Do this until we have n vectors in our basis, at which point we have an orthonormal basis for $\mathbb{C}^n$.

4. Now, write our matrix A in the orthonormal basis $\{\vec{b_{1\lambda_1}}, \dots \vec{b_{k_1\lambda_1}}, \vec{c_1}, \dots \vec{c_{n-k_1}}\}$. What does this look like? Well: we know that for each $\vec{b_{i\lambda_1}}$, by definition, we have $A\vec{b_{i\lambda_1}} = \lambda_1\vec{b_{i\lambda_1}}$, which in our orthonormal basis is the vector with i-th entry λ_1 and the rest 0. In other words, we know the first k columns of our matrix! In particular, our matrix

Theorem. (The Fundamental Theorem of Algebra) If $f(x)$ is a nonconstant polynomial with coefficients from some field F, with $F = \mathbb{R}$ or $\mathbb{C}$, then it has a root in $\mathbb{C}$. In other words, every nonconstant polynomial $f(x)$ has some corresponding value $r \in \mathbb{C}$ such that $f(r) = 0$.

[3] A matrix is called **upper-triangular** if all of its entries below the main diagonal are 0. For example, $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ is upper-triangular.

has the following form:

$$A_B = \begin{array}{c} \begin{array}{cc} \overbrace{\hphantom{\lambda_1 \ddots \lambda_1}}^{k_1 \text{ entries}} & \overbrace{\hphantom{A_{\text{rem}}}}^{\text{n-}k_1 \text{ entries}} \end{array} \\ \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & A_{\text{rem}} \\ & & \lambda_1 & \\ \hline & & & \\ & 0 & & A_2 \\ & & & \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} k_1 \text{ entries} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n-k_1 \text{ entries} \end{array}$$

To be specific: the above matrix consists of four distinct "blocks:"

(a) a $k_1 \times k_1$ diagonal matrix in the upper-left, with λ on its diagonal and 0's elsewhere,

(b) a $n - k_1 \times k_1$ matrix in the lower-left made entirely of 0's,

(c) a $k_1 \times n - k_1$ matrix in the uppper-right corner, which we name A_{rem}, and

(d) a $n - k_1 \times n - k_1$ matrix in the lower-right corner, which we name A_2.

5. Consider the matrix A_2. This is, by definition, a linear map that takes in vectors from the span of the set $\{\vec{c_1}, \dots \vec{c_{n-k_1}}\}$ and outputs vectors within that same span! (Think about why that is for a second. Persuade yourself that it is true.) In particular, because A_2 is now a square matrix over some vector space, we can take A_2 and put it back into steps 1-4 above! We can do this whenever this matrix A_2 exists, which is whenever $n - k_1$ is nonzero.

6. What do we have now? Well: we have an orthonormal basis

$$\{\vec{b_{1\lambda_1}}, \dots \vec{b_{k_1\lambda_1}}, \vec{b_{1\lambda_2}}, \dots \vec{b_{k_2\lambda_2}}, \vec{c_1'}, \dots \vec{c_{n-k_1-k_2}}\},$$

and have also shown that the matrix A written under the above basis has the form

$$A_B = \begin{array}{c} \begin{array}{ccc} \overbrace{\hphantom{\lambda_1 \ddots \lambda_1}}^{k_1 \text{ entries}} & \overbrace{\hphantom{\lambda_2 \ddots \lambda_2}}^{k_2 \text{ entries}} & \overbrace{\hphantom{A_{\text{rem}_2}}}^{\text{n-}k_1 - k_2 \text{ entries}} \end{array} \\ \left[\begin{array}{ccc|ccc|c} \lambda_1 & & & & & & \\ & \ddots & & & A_{\text{rem}_1} & & A_{\text{rem}_2} \\ & & \lambda_1 & & & & \\ \hline & & & \lambda_2 & & & \\ & 0 & & & \ddots & & A_{\text{rem}_3} \\ & & & & & \lambda_2 & \\ \hline & & & & & & \\ & 0 & & & 0 & & A_3 \\ & & & & & & \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} k_1 \text{ entries} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} k_2 \text{ entries} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n - k_1 - k_2 \text{ entries} \end{array}$$

7. Now, do it again to A_3! In fact, keep repeating this process until we cannot continue. What does our matrix under our resuting look like then? Well: we've filled its diagonal with these diagonal $\begin{bmatrix} \lambda_i & & \\ & \ddots & \\ & & \lambda_i \end{bmatrix}$ matrices, which have only 0's below them. In other words, our matrix is upper-triangular!

So we're done! $\square$

2 The Schur Decomposition: Why We Care

The above decomposition is incredibly useful in certain situations, like (as we often study) raising a matrix to some large power! We study an example here:

Example. Consider the following matrix:

$$A = \begin{bmatrix} 13 & 8 & 8 \\ -1 & 7 & -2 \\ -1 & -2 & 7 \end{bmatrix}$$

What is A^{50}?

Proof. We could calculate this directly, given enough computing time and power. However, that seems...bad. How can we do this faster? Well: we might hope that we could do the trick that we've used before, and write A in the form UDU^{-1}, where D is a diagonal matrix!

That doesn't work. In fact, there is no pair of matrices U, D, such that U is invertible and D is diagonal, and $A = UDU^{-1}$. (In this sense, A is like $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, which we studied on HW#2.

However, we can do something almost as good: find A's Schur decomposition! We first find this decomposition, and then talk about why it is useful for finding A^{50}.

First, we try to find an eigenvalue with corresponing nontrivial eigenvector for A. We do this, as always, by brute-force: if we have a vector (x, y, z) and constant λ such that

$$A = \begin{bmatrix} 13 & 8 & 8 \\ -1 & 7 & -2 \\ -1 & -2 & 7 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 13x + 8y + 8z \\ -x + 7y - 2z \\ -x - 2y + 7z \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

This gives us three linear equations, which we can solve:

$$\begin{aligned} 13x + 8y + 8z &= \lambda x, \\ -x + 7y - 2z &= \lambda y, \\ -x - 2y + 7z &= \lambda z. \end{aligned}$$

In particular, adding four copies of the second to the first gives us

$$9x + 36y = \lambda x + 4\lambda y \Leftrightarrow (9 - \lambda)x = (4\lambda - 36)y.$$

Similarly, adding four copies of the third to the first gives us

$$9x + 36z = \lambda x + 4\lambda z \Leftrightarrow (9 - \lambda)x = (4\lambda - 36)z.$$

Finally, subtracting the third from the second gives

$$9y - 9z = \lambda y - \lambda z \Leftrightarrow (9 - \lambda)y = (9 - \lambda)z.$$

So: um, these three equations seem to be basically shouting "try setting $\lambda = 9$," because that satisfies **all** of our equations! If we do this, then all three of our equations simplify to the same constraint, which is

$$x + 2y + 2z = 0.$$

There are lots of solutions to this equation! One that comes to mind is $(2, -2, 1)$. Checking shows that this is indeed an eigenvector for the eigenvalue 9:

$$A = \begin{bmatrix} 13 & 8 & 8 \\ -1 & 7 & -2 \\ -1 & -2 & 7 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 18 \\ -18 \\ 9 \end{bmatrix}$$

So: we have an eigenvalue! To find the Schur decomposition of A, we now need to find **all** of the eigenvectors corresponding to the eigenvalue 9. As shown above, this is just the space

$$E_9 = \{(x, y, z) \mid x + 2y + 2z = 0\}.$$

We want an orthonormal basis for this space. To do so, we first find a basis, and then use Gram-Schmidt. The basis is easy enough: we know that $(2, -2, 1)$ is in our space, and we can also immediately see that $(0, 1, -1)$ is also in our space, as it satisfies the constraint $x + 2y + 2z = 0$. If we perform Gram-Schmidt on these two vectors, setting $\vec{u_1} = (2, -2, 1)$, and

$$\begin{aligned} \vec{u_2} &= (0, 1, -1) - \text{proj}((0, 1, -1) \text{ onto } (2, -2, 1)) \\ &= (0, 1, -1) - \frac{(0, 1, -1) \cdot (2, -2, 1)}{(2, -2, 1) \cdot (2, -2, 1)}(2, -2, 1) \\ &= (0, 1, -1) + \frac{3}{9}(2, -2, 1) \\ &= \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right). \end{aligned}$$

Finally, we scale these vectors by their length to get a basis for $E_9 : \left\{\left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right)\right\}$. We now need to extend this to a basis for all of $\mathbb{C}^3$: to do this, we simply take some vector not in the span of these two vectors, like $(0, 0, 1)$ (check for yourself: why is this not in their span?), and perform Gram-Schmidt on this third vector:

$$\begin{aligned} \vec{u_3} &= (0, 0, 1) - \text{proj}\left((0, 0, 1) \text{ onto } \left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right)\right) - \text{proj}\left((0, 0, 1) \text{ onto } \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right)\right) \\ &= (0, 0, 1) - \frac{1/3}{1}\left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right) + \frac{2/3}{1}\left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right) \\ &= \left(\frac{2}{9}, \frac{4}{9}, \frac{4}{9}\right). \end{aligned}$$

Scaling this vector by its length, which is $\sqrt{36/81} = \frac{2}{3}$, gives us $\left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$. So we have an orthonormal basis

$$B = \left\{\vec{b_{1,\lambda}} = \left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right), \vec{b_{2,\lambda}}\left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right), \vec{c} = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)\right\}$$

Now we need to write our matrix A in the basis B. To do this, simply note that because $\vec{b_{1,\lambda}}, \vec{b_{2,\lambda}}$ are both eigenvectors for the eigenvalue $\lambda = 9$, we know exactly where A sends these vectors: specifically, we have $A\vec{b_{1,\lambda}} = 9\vec{b_{1,\lambda}} = (9,0,0)_B$, and $A\vec{b_{2,\lambda}} = 9\vec{b_{2,\lambda}} = (0,9,0)_B$.

Finally, we can just calculate what A does to $\vec{c}$:

$$\begin{bmatrix} 13 & 8 & 8 \\ -1 & 7 & -2 \\ -1 & -2 & 7 \end{bmatrix} \cdot \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 15 \\ 3 \\ 3 \end{bmatrix} = 9(\vec{b_{1,\lambda}} + \vec{b_{2,\lambda}} + \vec{c}) = (9,9,9)_B.$$

Therefore, our matrix A under the basis B is just the matrix with these three vectors as its columns: i.e.

$$A_B = \begin{bmatrix} 9 & 0 & 9 \\ 0 & 9 & 9 \\ 0 & 0 & 9 \end{bmatrix}_B .$$

In other words, under the orthonormal basis we've found, A is upper-triangular! (We were lucky here and only had to go one step along the Schur process to get an upper-triangular matrix. However, if the lower-right-hand block was bigger — if we were looking at a 4×4 matrix, say —we might have had to repeat this process of finding an eigenvalue, extending it to a basis of the relevant space, etc. on the lower block.)

So: how can we use this to raise A to a large power? Well: notice that if we write

$$A_B^{50} = \left(\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B + \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B \right)^{50},$$

we can use the binomial theorem to write this as the following sum:

$$\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{50} + \binom{50}{1} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{49} \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B^{1} + \binom{50}{2} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{48} \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B^{2}$$
$$+ \ldots + \binom{50}{49} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{1} \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B^{49} + \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_{B.}^{50}$$

Cool observation:

$$\begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B^{2} = \text{the all-zeroes matrix!}$$

This lets us simplify the above dramatically! In particular, every term with a $\begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B^{k}$ goes away for $k \geq 2$; so we have

$$
\begin{aligned}
A_B^{50} &= \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{50} + \binom{50}{1} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{49} \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B \\
&= \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{50} + 50 \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}_B^{49} \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B \\
&= \begin{bmatrix} 9^{50} & 0 & 0 \\ 0 & 9^{50} & 0 \\ 0 & 0 & 9^{50} \end{bmatrix}_B + \begin{bmatrix} 50 \cdot 9^{49} & 0 & 0 \\ 0 & 50 \cdot 9^{49} & 0 \\ 0 & 0 & 50 \cdot 9^{49} \end{bmatrix}_B \begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 9 \\ 0 & 0 & 0 \end{bmatrix}_B \\
&= \begin{bmatrix} 9^{50} & 0 & 0 \\ 0 & 9^{50} & 0 \\ 0 & 0 & 9^{50} \end{bmatrix}_B + \begin{bmatrix} 0 & 0 & 50 \cdot 9^{50} \\ 0 & 0 & 50 \cdot 9^{50} \\ 0 & 0 & 0 \end{bmatrix}_B \\
&= \begin{bmatrix} 9^{50} & 0 & 50 \cdot 9^{50} \\ 0 & 9^{50} & 50 \cdot 9^{50} \\ 0 & 0 & 9^{50} \end{bmatrix}_{B.}
\end{aligned}
$$

If we convert back to the standard basis, we have

$$
\begin{aligned}
A &= \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 9^{50} & 0 & 50 \cdot 9^{50} \\ 0 & 9^{50} & 50 \cdot 9^{50} \\ 0 & 0 & 9^{50} \end{bmatrix}_B \cdot \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \\
&= 9^{49} \cdot \begin{bmatrix} 209 & 400 & 400 \\ -50 & -91 & -100 \\ -50 & -100 & -91 \end{bmatrix}.
\end{aligned}
$$

Solution! With relatively little work, as compared to calculating A^{50} by hand! □

3 The Schur Decomposition: How People Actually Find It

It bears noting that in practice, people don't use repeated Gram-Schmidt to find this Schur decomposition! Instead, they use the following absolutely beautiful application of the QR-decomposition to find the Schur decomposition:

Theorem. (Francis 1961, Kublanovskaja 1962, Huang and Tam 2005): Take some $n \times n$ matrix A with entries in $\mathbb{C}$. Suppose that A has n distinct eigenvalues, and the Schur decomposition $Q_s^T R_s Q_s$. Consider the following process:

1. Set $A_1 = A$.
2. Find a QR-decomposition $Q_1 R_1$ of A.
3. Define $A_2 = R_1 Q_1$.
4. Now, find a QR-decomposition $Q_2 R_2$ of A_2.

5. Repeat this process! I.e. if A_k has the QR-decomposition Q_kR_k, use this to define $A_{k+1} = R_kQ_k$.

Notice that because

$$A_{k+1} = R_kQ_k = Q_k^T(Q_kR_k)Q_k = (Q_kR_k)_{\text{in the basis } Q_k^T},$$

the eigenvalues of A_k and A_{k+1} don't change.

Furthermore, the following very surprising fact holds: if we look at the sequence $A_1, A_2, A_3, \ldots$, the diagonals of these matrices converge to the eigenvalues of A! In fact, in many situations we actually have something stronger: the sequence of the A_i's will converge to R_s, the upper-triangular part of the Schur decomposition! In other words, we often have the following result:

$$\lim_{n\to\infty} A_n = R_s.$$

A proof of this is beyond this course, but it's cool to know about this theorem nonetheless!