Math 108B	Professor: Padraic Bartlett

Lecture 6: Lies, Inner Product Spaces, and Symmetric Matrices

Week 6 — *UCSB 2014*

1 Lies

Fun fact: I have deceived[1] you somewhat with these last few lectures!

Let me clarify. For the past week, we've been working with the Schur decomposition, a result which says that for any complex-valued matrix A, there is some orthonormal basis B for $\mathbb{C}^n$ under which A is upper-triangular!

This is certainly true. Moreover, the proof we used to show this result is also completely accurate: our six-step process of

1. finding an eigenvalue λ,
2. finding an orthonormal basis B for E_λ, the collection of all eigenvectors for λ,
3. extending this basis to an orthonormal basis $B \cup C$ for all of $\mathbb{C}^n$,
4. writing A in this basis $B \cup C$,
5. noticing that A in this basis is a matrix of the form

$$A_B = \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & A_{\text{rem}} \\ & & \lambda_1 & \\ \hline & 0 & & A_2 \end{array}\right],$$

6. and finally repeating this whole 1-6 process again on A_2— this works exactly as stated.

Rather, I claim that I tricked you in a far subtler and more fundamental way: not in terms of proof techniques or methods, but rather in terms of the very basic concepts that you've been working with! To understand this deception, try to answer the following question:

Question. What is the length of the vector $(1, i) \in \mathbb{C}^2$? What is a vector that is orthogonal to $(1, i)$?

Naively, you might hope that you can solve this problem by just using our existing formula for the length of a real-valued vector:

$$\forall \vec{v} \in \mathbb{R}^n, \ ||\vec{v}|| = \sqrt{v_1^2 + \ldots v_n^2}.$$

[1] For good reasons.

Doing this here, however, gives us a fairly distressing answer:

$$\sqrt{1^2 + i^2} = \sqrt{1 - 1} = 0.$$

That seems. . . wrong. This is a clearly nonzero vector; surely it should have nonzero length!

A more subtle yet equally wrong mistake comes up when trying to find vectors orthogonal to $(1, i)$. You might believe that to find a vector orthogonal to $(1, i)$, we just need to find some $\alpha, \beta \in \mathbb{C}^2$ such that

$$(\alpha, \beta) \cdot (1, i) = 0.$$

If you were interpret the above as the normal dot product, this is just asking us for α, β such that

$$\alpha \cdot 1 + \beta \cdot i = 0.$$

One immediate solution to the above is to set $\alpha = 1, \beta = i$: then we get $1^2 + i^2 = 0$.

But what does this say? We've just argued that $(1, i)$ is orthogonal to. . . $(1, i)$. In other words, we're claiming that a vector is orthogonal to itself!

This is kind of horrible for us; pretty much all of our work earlier has very heavily relied on the idea that orthogonal vectors are linearly independent and have a large host of other "nice" properties that are **not** working in the above example. So: what do we do? Do we give up? Is math broken forever?

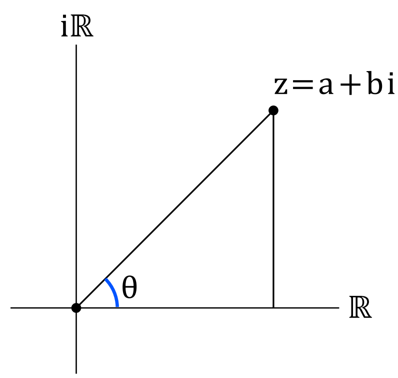
2 No.

Math is not broken forever.

Rather, our definitions just need some clarification for $\mathbb{C}^n$!

3 Length and Orthogonality in $\mathbb{C}^n$

To get a better idea of how we should be dealing with length and orthogonality, let's look at the simplest case possible, i.e. $\mathbb{C}^1$. For a complex number $a + bi$, we have the following:



The above picture suggests that the length of $a + bi$ is just $\sqrt{a^2 + b^2}$, the hypotenuse of the triangle in the above picture. Denote this quantity as $|z|$.

This idea is related to the following definition:

Definition. Given a complex number $z \in \mathbb{C}$, write $z = a + bi$, where $a, b \in \mathbb{R}$. We define the **complex conjugate**, $\overline{z}$, of z as the complex number $a - bi$. Notice that for any real number a, the complex conjugate of a is just a itself, and furthermore that real numbers are the only numbers with this property. Also, notice that for any complex number $z = a + bi$, we have

$$z \cdot \overline{z} = (a + bi)(a - bi) = a^2 + b^2 = |z|^2,$$

the length of z squared.

This setting suggests to us that a better definition for the "length" of a complex vector might be the following:

Definition. Take any vector $\vec{v} = (v_1, \ldots v_n) \in \mathbb{C}^n$. The **length** of $\vec{v}$ is the following quantity:

$$||\vec{v}|| = \sqrt{|v_1|^2 + \ldots |v_n|^2}.$$

This avoids the issue we had earlier, where nonzero vectors had nonzero length; here, if any component of $\vec{v}$ is nonzero, then (because the length quantities $\sqrt{a_j^2 + b_j^2}$ are all positive for any $v_j = a_j + ib_j i$) the entire quantity $||\vec{v}||$ is nonzero. So this fixes our first problem!

Moreover: I claim that this solution to our first problem — a notion of "length" for complex vectors — suggests an answer to our second problem, which is a notion for "orthogonality" of complex vectors!

Initially, we had hoped to say that two complex vectors are orthogonal whenever their dot product is 0. This, however, was a disaster: we wound up with vectors being orthogonal to themselves, which is very far away from the "meeting at right angles" idea that orthogonality had for us in $\mathbb{R}^n$. This was caused by the same phenomena that made our original guess for how to define length fail; in both cases, we had complex multiplication creating things that "canceled" when we didn't actually want these things to cancel! So, if we want to fix this notion, we need to start by fixing our notion of **dot product** for complex numbers! And in particular, this fix is suggested by our earlier work with length. Consider the following proposition, which we've proven in past lectures for real-valued vectors:

Proposition. Given any vector $\vec{v} \in \mathbb{R}^n$, we have $||\vec{v}||^2 = \vec{v} \cdot \vec{v}$.

This suggests that maybe our fix for the "length" problem earlier may be a good fix for our "orthogonality" problem here. In particular, suppose that we want to define a "product" on complex vectors with the following property: for any vector $\vec{v} \in \mathbb{C}^n$, we should have $\vec{v}$ **prod** $\vec{v} = ||\vec{v}||^2$. What should we do?

Well, if we make the following observation on the idea of length:

$$||\vec{v}||^2 = |v_1|^2 + \ldots |v_n|^2 = v_1\overline{v_1} + v_2\overline{v_2} + \ldots v_n\overline{v_n} = \sum_{i=1}^{n} v_i\overline{v_i}.$$

this strongly suggests the following definition for a "product:"

Definition. Suppose that $\vec{v}, \vec{w}$ are a pair of vectors in $\mathbb{C}$. We define the **inner product** of $\vec{v}$ and $\vec{w}$ as follows:

$$\langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i \overline{w_i}.$$

This idea of inner product lets us create a corresponding notion of orthogonality, like we did on HW#3 for some other notions of inner products:

Definition. Given two nonzero complex vectors $\vec{v}, \vec{w} \in \mathbb{C}$, we say that $\vec{v}$ and $\vec{w}$ are **orthogonal** if and only if $\langle \vec{v}, \vec{w} \rangle = 0$.

Notice that under this definition, we never say that a vector is orthogonal to itself, as for any nonzero vector $\vec{v}$ we have $\langle \vec{v}, \vec{v} \rangle = ||\vec{v}||^2$, which is nonzero.

This definition lets us go back and answer our earlier question: a vector (α, β) is now orthogonal to $(1, i)$ if and only if

$$\langle (\alpha, \beta), (1, i) \rangle = \alpha \cdot 1 + \beta \cdot \bar{i} = \alpha - i\beta.$$

So, one sample vector that is orthogonal to $(1, i)$ is therefore $(1, -i)$.

This notion of inner product on complex vectors is pretty similar to our old notion of dot product over real-valued vectors. In particular, it shares a number of properties with the dot product:

1. **Positive-definite**: for any vector $\vec{v}$, the inner product of $\vec{v}$ with itself corresponds to its length squared, and thus in particular is nonzero whenever $\vec{v} \neq 0$.

2. **Linear in first slot**: for any vectors $\vec{u}, \vec{v}, \vec{w}$ and coefficient α, the inner product $\langle \vec{u} + \alpha\vec{v}, \vec{w} \rangle$ is equal to the sum $\langle \vec{u}, \vec{w} \rangle + \alpha \langle \vec{v}, \vec{w} \rangle$.

3. **Conjugate-symmetric**: for any two vectors $\vec{v}, \vec{w}$, we have $\langle \vec{v}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{v} \rangle}$. (For the real-valued dot product, this property trivially held because $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$ and because both sides were real-valued (and thus equal to their own conjugate.))

There are many different notions of inner products; we say that any map that takes in two vectors in a space and sends them to a scalar is an **inner product** if it satisfies the three properties above. In practice, however, we will usually restrict ourselves to just working with this idea of inner product for the complex numbers, and only work with other inner products when explicitly stated.

More excitingly, basically every old result we've ever had on orthogonality? It carries forward to this notion of inner product! We list some highlights here:

Theorem. If a collection of vectors $\vec{v_1}, \ldots \vec{v_k}$ are all orthogonal to each other with respect to some inner product, then they are all linearly independent.

Proof. Take any linear combination of these vectors that combines to $\vec{0}$:

$$a_1\vec{v_1} + \ldots + a_n\vec{v_n} = \vec{0}$$

Now, take the inner product of this linear combination with any $\vec{v_i}$. By linearity, this is equal to

$$a_1\langle\vec{v_1}, \vec{v_i}\rangle + \ldots + a_n\langle\vec{v_n}, \vec{v_i}\rangle = \langle\vec{v_i}, \vec{0}\rangle.$$

Notice that because of linearity, we have that $\langle\vec{v_i}, \vec{0}\rangle = \langle\vec{v_i}, \vec{0}+\vec{0}\rangle = \langle\vec{v_i}, \vec{0}\rangle + \langle\vec{v_i}, \vec{0}\rangle$, and therefore that $\langle\vec{v_i}, \vec{0}\rangle = 0$. So this tells us the right-hand-side. The left-hand-side, by orthogonality, is just $a_i\langle\vec{v_i}, \vec{v_i}\rangle$, which is $a_i||\vec{v_i}||^2$. In particular, the $||\vec{v_i}||^2$ portion is nonzero; so we can divide both sides through by it to get

$$a_i = 0.$$

We can perform this manipulation for any i; therefore, all of the coefficients in our linear combination must be 0. In other words, our set is linearly independent, because there is no nontrivial combination of our vectors that creates the all-zero vector! $\square$

Definition. Let $\vec{v}, \vec{w}$ be a pair of vectors in a space equipped with an inner product. The **projection** of $\vec{v}$ onto $\vec{w}$ with respect to that inner product is the following object:

$$\text{proj}(\vec{v} \text{ onto } \vec{w}) = \frac{\langle\vec{v}, \vec{w}\rangle}{\langle\vec{w}, \vec{w}\rangle}\vec{w}.$$

Similarly, the **orthogonal part** of $\vec{v}$ over $\vec{w}$ with respect to that inner product, denoted $\text{orth}(\vec{v} \text{ over } \vec{w})$, is the following vector:

$$\text{orth}(\vec{v} \text{ over } \vec{w}) = \vec{v} - \text{proj}(\vec{v} \text{ onto } \vec{w})$$

Theorem. Given two vectors $\vec{v}, \vec{w}$ from a space equipped with some inner product, if we form the vector $\text{orth}(\vec{v} \text{ over } \vec{w})$ with respect to our inner product, it is orthogonal to the vector $\vec{w}$.

Proof. This proof works exactly how it did for dot products, except there are $\langle -, - \rangle$'s where we had $\cdot$'s before: we just calculate $\langle \text{orth}(\vec{v} \text{ over } \vec{w}), \vec{w}\rangle$, and get

$$\begin{aligned}
\langle \text{orth}(\vec{v} \text{ over } \vec{w}), \vec{w}\rangle &= \langle \vec{v} - \text{proj}(\vec{v} \text{ onto } \vec{w}), \vec{w}\rangle \\
&= \langle\vec{v}, \vec{w}\rangle - \left\langle \frac{\langle\vec{v}, \vec{w}\rangle}{\langle\vec{w}, \vec{w}\rangle}\vec{w}, \vec{w} \right\rangle \\
&= \langle\vec{v}, \vec{w}\rangle - \frac{\langle\vec{v}, \vec{w}\rangle}{\langle\vec{w}, \vec{w}\rangle}\langle\vec{w}, \vec{w}\rangle \\
&= \langle\vec{v}, \vec{w}\rangle - \langle\vec{v}, \vec{w}\rangle \\
&= 0.
\end{aligned}$$

So these two vectors are orthogonal! $\square$

In past lectures, we generalized the above ideas to the **Gram-Schmidt process**. We do this again here:

Theorem. Consider the following process (called the Gram-Schmidt process, formally), to create a set of k nonzero orthogonal vectors $\{\vec{u_1}, \dots \vec{u_k}\}$ out of combinations of some basis $B = \{\vec{b_1}, \dots \vec{b_k}\}$:

- $\vec{u_1} = \vec{b_1}$.
- $\vec{u_2} = \vec{b_2} - \text{proj}(\vec{b_2} \text{ onto } \vec{u_1})$.
- $\vec{u_3} = \vec{b_3} - \text{proj}(\vec{b_3} \text{ onto } \vec{u_1}) - \text{proj}(\vec{b_3} \text{ onto } \vec{u_2})$.
- $\vec{u_4} = \vec{b_4} - \text{proj}(\vec{b_4} \text{ onto } \vec{u_1}) - \text{proj}(\vec{b_4} \text{ onto } \vec{u_2}) - \text{proj}(\vec{b_4} \text{ onto } \vec{u_3})$.

  $\vdots$

- $\vec{u_k} = \vec{b_k} - \text{proj}(\vec{b_k} \text{ onto } \vec{u_1}) - \dots - \text{proj}(\vec{b_k} \text{ onto } \vec{u_{k-1}})$.

This process works: i.e. the result above is a set of orthogonal vectors that forms a new basis for whatever space B spanned! In particular, it works for any fixed inner product that you use to calculate all of the projection maps above.

We omit the proof here, because it's literally the same idea as before: just replace all of the instances of dot products with inner products, and everything goes through just like it did in our earlier example.

4 Applications: Real Symmetric Matrices

With the above clarifications made, everything we did with Schur's theorem still works: throughout our proof of Schur's theorem, the only technique we used to get orthogonal vectors was repeated Gram-Schmidt, and this works for any space equipped with an inner product! Math is unbroken, life is great, we can decompose matrices, yay.

...what else can we do with this new idea? Well, consider the following definition:

Definition. A matrix A is called **symmetric** if $A^T = A$: i.e. if for any (i, j), the (i, j)-th entry of A is equal to the (j, i)-th entry of A. For example, the matrix $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is symmetric, while the matrix $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is not.

On the homework, you proved the following result:

Theorem. Suppose that A is a real symmetric matrix, and let A's Schur decomposition be something of the form URU^{-1}, where U is an orthogonal matrix of real numbers, and R is an upper-triangular matrix. Then R is in fact a diagonal matrix.

Using our new ideas about inner products, we can strengthen this result here:

Theorem. (The spectral theorem.) Suppose that A is a $n \times n$ real symmetric matrix (i.e. don't make any assumptions about what U is like we did above.) Then in A's Schur decomposition URU^{-1}, R is a diagonal real-valued matrix! Furthermore, we can insure in our construction of U that it is a real-valued orthogonal matrix.

In other words, we don't need to make any assumptions about A itself: as long as A is a real-valued symmetric matrix, it has the form UDU^T, where U is an orthogonal matrix!

Consequently, notice that if we multiply UDU^T by any column $\vec{u_{c_i}}$ of U, we get

$$UDU^T \cdot \vec{u_{c_i}} = UD \cdot \vec{e_i} = Ud_{ii}\vec{e_i} = d_{ii} \cdot \vec{u_{c_i}}.$$

In other words, we have that the columns of U are all eigenvectors of $A = UDU^T$, with eigenvalues corresponding to the entries d_{ii} on the diagonal of D! This gives us the following remarkably useful result for free:

Theorem. Suppose that A is a $n \times n$ real symmetric matrix. Then there is an orthonormal basis $\{\vec{u_{c_1}}, \ldots \vec{u_{c_n}}\}$ for $\mathbb{R}^n$, made out of n orthogonal (and thus linearly independent) eigenvectors for A.

This is incredibly useful: as we've seen often in this course, many matrices (like $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$) cannot be written in this form, and whenever we **can** write a matrix in this form it makes lots of tasks, like raising the matrix to large powers, remarkably trivial.

How do we prove this?

5 Real Symmetric Matrices: The Concepts

First, we need to understand the following pair of definitions:

Definition. Take a complex-valued $m \times n$ matrix A. The **conjugate transpose** of A, denoted A^*, is the $n \times m$ matrix formed as follows: set the (i, j)-th entry of A^* to be the complex conjugate of the (j, i)-th entry of A. In other words, to get A^*, first take A's transpose, then replace each entry with its complex conjugate. Again, notice that for real-valued matrices A, we have $A^* = A^T$.

Definition. A complex-valued $n \times n$ matrix is called **unitary** if $U^* \cdot U = I$. For example, the matrix $\begin{bmatrix} 3/5 & 4i/5 & 0 \\ 4i/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is unitary, because

$$\begin{bmatrix} 3/5 & 4i/5 & 0 \\ 4i/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}^* \cdot \begin{bmatrix} 3/5 & 4i/5 & 0 \\ 4i/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3/5 & -4i/5 & 0 \\ -4i/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3/5 & 4i/5 & 0 \\ 4i/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The reason we mention these concepts is because they will help us understand the matrices U that we get from Schur decompositions URU^{-1} of matrices! Specifically, we have the following theorem:

Theorem. Take any orthonormal basis $\{\vec{u_{c_1}}, \ldots \vec{u_{c_n}}\}$ for $\mathbb{C}^n$. Form the $n \times n$ matrix U given by using these basis vectors as the columns of U. Then U is unitary: i.e. $U^*U = I$.

Proof. To see this, simply look at the product U^*U. Denote the columns $\vec{u_{ci}}$ as the vectors $(u_{1i}, \dots u_{ni})$; then we have

$$U^* \cdot U = \begin{bmatrix} \overline{u_{11}} & \overline{u_{21}} & \dots & \overline{u_{n1}} \\ \overline{u_{12}} & \overline{u_{22}} & \dots & \overline{u_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{u_{1n}} & \overline{u_{2n}} & \vdots & \overline{u_{nn}} \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \vdots & u_{nn} \end{bmatrix}.$$

The diagonal entries (i,i) on this product matrix are the following expressions:

$$(\overline{u_{1i}}, \dots \overline{u_{ni}}) \cdot (u_{1i}, \dots u_{ni}) = \sum_{k=1}^{n} \overline{u_{ki}} u_{ki} = \sum_{k=1}^{n} |u_{ki}|^2 = ||\vec{u_{c_i}}||^2 = 1,$$

because the $\vec{u_{c_i}}$'s are all unit length.

Conversely, the off-diagonal entries $(i,j), i \neq j$ are just

$$(\overline{u_{1i}}, \dots \overline{u_{ni}}) \cdot (u_{1j}, \dots u_{nj}) = \sum_{k=1}^{n} \overline{u_{ki}} u_{kj} = \langle \vec{u_{c_i}}, \vec{u_{c_j}} \rangle = 0.$$

Therefore, we have that $U^*U = I$, as claimed. ☐

In particular, this tells us that the matrices U from Schur decompositions URU^{-1} of matrices are unitary, because those matrices come from orthonormal bases for $\mathbb{C}^n$! So we can actually write Schur decompositions in the form URU^*, which is nice.

Now, note the following quick lemmas, one of which we have proven before:

Lemma 1. *Take any string of matrices $A_1, \dots A_n$, such that $A_1 \cdot \ldots \cdot A_n$ is well-defined. Then*

$$(A_1 \cdot A_2 \cdot \ldots \cdot A_n)^T = A_n^T \cdot A_{n-1}^T \cdot \ldots \cdot A_2^T \cdot A_1^T.$$

Proof. First, notice that for any two $n \times n$ matrices A, B, we have $(A \cdot B)^T = (B^T) \cdot (A^T)$. This is not too hard to show: simply notice that due to the definition of matrix multiplication, the entry in (i,j) of $A \cdot B$ is

$$(\text{the } i\text{-th row of } A) \cdot (\text{the } j\text{-th column of } B),$$

which means that the (j,i)-th entry of $(A \cdot B)^T$ is precisely that dot product.

On the other hand, notice that the j-th row of B^T is just $(b_{1,j}, \dots b_{n,j})$, in other words the j-th column of B. Similarly, the i-th column of A^T is $(a_{i,1}, \dots a_{i,n})$, the i-th row of A.

Therefore, we have that the (j,i)-th entry of $(B^T) \cdot (A^T)$ is just

$$\begin{aligned} &(\text{the } j\text{-th row of } B^T) \cdot (\text{the } i\text{-th column of } A^T) \\ =&(\text{the } j\text{-th column of } B) \cdot (\text{the } i\text{-th row of } A) \\ =&(\text{the } i\text{-th row of } A) \cdot (\text{the } j\text{-th column of } B). \end{aligned}$$

Therefore the two matrices $(A \cdot B)^T$, $(B^T) \cdot (A^T)$ have the same entries, and are therefore the same matrices!

In particular, this tells us that if we look at the transpose of a product of n matrices together, we have that it's just the product of their transposes in reverse order! Just repeatedly apply the above result. $\square$

Lemma 2. *Take any string of matrices $A_1, \ldots A_n$, such that $A_1 \cdot \ldots \cdot A_n$ is well-defined. Let $\overline{A_i}$ denote the matrix formed by replacing each entry of A_i with its conjugate. Then $\overline{A_1 \cdot \ldots \cdot A_n} = \overline{A_1} \cdot \ldots \cdot \overline{A_n}$.*

Proof. It suffices to prove this claim for the product of two matrices, as by induction (like we did above) the result will extend to arbitrary products of matrices. To do this, take any two matrices A, B. An arbitrary entry (i, j) in $\overline{A} \cdot \overline{B}$ looks like the dot product of the conjugate of A's i-th row with the conjugate of B's j-th column: i.e.

$$\sum_{k=1}^{n} \overline{a_{i,k}} \cdot \overline{b_{k,j}}.$$

(Note that even though we have complex numbers here, we still use the dot product to calculate matrix multiplication. This is because the origins of how we defined matrix multiplication came out of how we compose linear maps; in this sense, the connection with the dot product was somewhat "accidental.")

Similarly, an arbitrary entry in $\overline{AB}$ is just the conjugate of the dot product of A's i-th row and B's j-th column: i.e.

$$\sum_{k=1}^{n} \overline{a_{i,k} \cdot b_{k,j}}.$$

But for any two complex numbers $a + bi, x + yi$, we have

$$\overline{(a+bi) \cdot (x+yi)} = \overline{(ax-by) + (ay+bx)i} = (ax-by) - (ay+bx)i = (a-bi) \cdot (x-yi) = \overline{(a+bi)} \cdot \overline{(x+yi)}.$$

So these two quantities are equal, and therefore their corresponding matrices are equal. $\square$

Combining these two lemmas gives us the following result:

Lemma 3. *Take any string of matrices $A_1, \ldots A_n$, such that $A_1 \cdot \ldots \cdot A_n$ is well-defined. Then*

$$(A_1 \cdot A_2 \cdot \ldots \cdot A_n)^* = A_n^* \cdot A_{n-1}^* \cdot \ldots \cdot A_2^* \cdot A_1^*.$$

This gives us our first step on the way to the spectral theorem:

Theorem. Suppose that A is a $n \times n$ real symmetric matrix. Then in A's Schur decomposition URU^*, R is a diagonal real-valued matrix.

Proof. First, notice that because A is real-valued and symmetric, then $A^* = A$. Therefore, if A's Schur decomposition is URU^*, we have

$$\begin{aligned} A &= URU^*, \\ A = A^* &= (URU^*)^* = (U^*)^* R^* U^* = UR^*U^* \\ \Rightarrow URU^* &= UR^*U^* \\ \Rightarrow U^*URU^*U &= U^*UR^*U^*U \\ \Rightarrow R &= R^*. \end{aligned}$$

Because R is upper-triangular, R^* is lower-triangular; therefore, the identity above tells us that all of the off-diagonal entries must be zero! Furthermore, because each of the diagonal entries of R is equal to its own conjugate, we must have that all of these values are real. So we have proven our theorem! $\square$

Our last step is to show that the entries in the matrix U itself must also be real-valued. This, however, is not hard:

Theorem. Suppose that A is a $n \times n$ real symmetric matrix. Then it is possible to create A's Schur decomposition UDU^{-1} in such a way that U is an orthogonal real-valued matrix.

Proof. Take a Schur decomposition UDU^* of A, where we know that D is diagonal and real-valued from our earlier work.

Take any column $\vec{u_{c_i}}$ of U. Notice that because A written in the basis U is the diagonal matrix D, we must have that $\vec{u_{c_i}}$ is an eigenvector — one with corresponding eigenvalue d_{ii} in the matrix D. In particular, this tells us that A has n eigenvalues with n linearly-independent corresponding eigenvectors.

Now, these eigenvectors might be complex-valued. That's sad. Let's fix that!

Pick any eigenvalue λ. Let $\vec{u_1}, \ldots \vec{u_k}$ be the columns of U that are complex-valued eigenvectors corresponding to λ. For each vector $\vec{u}$, let $Re(\vec{u})$ denote the real-valued vector formed by taking the real parts of the vector $\vec{u}$, and let $Im(\vec{u})$ denote the real-valued vector formed by taking the coefficients from the imaginary parts of the vector $\vec{u}$. Notice that because $A\vec{u} = \lambda\vec{u}$, we have in fact that $A \cdot Re(\vec{u}) = \lambda Re(\vec{u})$ and $A \cdot Im(\vec{u}) = \lambda \cdot Im(\vec{u})$, by just looking at the real and imaginary components of $\vec{u}$!

Look at the collection

$$\{Re(\vec{u_1}), Im(\vec{u_1}), \ldots Re(\vec{u_k}), Im(\vec{u_k})\}.$$

I claim that there is some subset of k of these vectors that are linearly independent. To see why, notice that these vectors span the same k-dimensional complex vector space spanned by the k orthogonal vectors $\{\vec{u_1}, \ldots \vec{u_k}\}$, because we can create each of the $\vec{u_i}$'s as a complex linear combination $Re(\vec{u_i}) + i \cdot Im(\vec{u_i})$. Therefore, over $\mathbb{C}$ there must be a subset of our collection that is a basis for this k-dimensional space — i.e. a set of k linearly independent vectors!

What have we made here? Well: we have made a set of k linearly independent eigenvectors for our eigenvalue λ! By using Gram-Schmidt on this set, we can make it into an orthonormal basis for this eigenvalue.

Do this sequentially for all of the eigenvalues of A. I claim that these eigenvectors form an orthonormal basis for $\mathbb{R}^n$! If this is true, then we're done with our proof: if we write A as a matrix under this basis, we've written $A = UDU^T$, for U an orthogonal matrix and D the diagonal matrix made out of A's eigenvalues!

$\square$

So: to properly finish our proof, we need one final lemma:

Lemma 4. *Let A be a real symmetric matrix, λ, μ a pair of distinct real-valued eigenvalues, and $\vec{v}, \vec{w}$ a pair of corresponding real-valued eigenvectors. Then $\vec{v}$ and $\vec{w}$ are orthogonal.*

Proof. This proof relies on the following trivial-looking but confusingly powerful linear algebra technique: for a $n \times n$ matrix A and vector $\vec{v} \in F^n$, we have

$$A\vec{x} = \begin{bmatrix} \vec{a_{r_1}} \cdot \vec{x} \\ \vdots \\ \vec{a_{r_n}} \cdot \vec{x} \end{bmatrix} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \cdot A^T = \vec{x}A^T.$$

Using this, we have the following:

$$\lambda(\vec{v} \cdot \vec{w}) = (\lambda \cdot \vec{v}) \cdot \vec{w} = (A\vec{v}) \cdot \vec{w} = (\vec{v} \cdot A^T) \cdot \vec{w} = \vec{v} \cdot (A \cdot \vec{w}) = \vec{v} \cdot (\mu\vec{w}) = \mu(\vec{v} \cdot \vec{w}).$$

Therefore, either these two vectors are orthogonal or $\lambda = \mu$. Note that if you replaced the dot product above with an inner product, you'd get the same result for complex-valued matrices. $\square$

Done! We've now proven the spectral theorem, one of the crown jewels of linear algebra! Isn't it cool?