



where the  $\lambda_i$  are eigenvalues of  $\tilde{A}$  and  $\tilde{A}$  is in upper-triangular form. Notice that, since for defective matrices there are fewer than  $n$  distinct eigenvalues, here we will repeat each eigenvalue in the diagonal in accordance with its *multiplicity*. The multiplicity of an eigenvalue  $\lambda_i$  of a matrix  $A$  represents the number of times the linear factor  $(\lambda - \lambda_i)$  appears in the characteristic polynomial of  $A$ . For instance, consider the defective matrix

$$D = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

whose characteristic polynomial was shown above to be

$$P_D(\lambda) = (\lambda - 1)^2.$$

Thus, we say that the eigenvalue  $\lambda = 1$  of  $D$  has a multiplicity of 2 (even though the corresponding eigenspace of  $D$  is only one-dimensional). Clearly, the sum of the multiplicities of all distinct eigenvalues of an  $n \times n$  matrix (such as  $A$ ) will be  $n$ , so we can indeed produce the  $\lambda_i$  that we need for our desired form to make sense.

### 3 Computing Upper-triangular Form

We wish to find a change of basis that converts an arbitrary  $n \times n$  square matrix  $A$  into the form  $\tilde{A}$ . Let this change of basis be represented by the columns  $\vec{v}_i$  of the matrix  $U$ , such that

$$A = U\tilde{A}U^{-1} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & ? & ? & \cdots & ? \\ 0 & \lambda_2 & ? & \cdots & ? \\ 0 & 0 & \lambda_3 & \cdots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}^{-1}.$$

Rearranging to get rid of the inverse, we obtain

$$A \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & ? & ? & \cdots & ? \\ 0 & \lambda_2 & ? & \cdots & ? \\ 0 & 0 & \lambda_3 & \cdots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Now, breaking this matrix down into a series of vector equations, we obtain the system

$$\begin{aligned} A\vec{v}_1 &= \lambda_1\vec{v}_1 \\ A\vec{v}_2 &= (?)\vec{v}_1 + \lambda_2\vec{v}_2 \\ A\vec{v}_3 &= (?)\vec{v}_1 + (?)\vec{v}_2 + \lambda_3\vec{v}_3 \\ &\vdots \\ A\vec{v}_n &= (?)\vec{v}_1 + (?)\vec{v}_2 + \dots + (?)\vec{v}_{n-1} + \lambda_n\vec{v}_n. \end{aligned}$$

Note that we use the symbol  $?$  to represent *different* unknown quantities each time it is written, not always the same value. Let's also temporarily forget that the  $\lambda_i$  are meant to be the eigenvalues of our system, and

instead just treat them as arbitrary scalar coefficients.

Thus, we see that a change of basis that puts a matrix  $A$  into upper-triangular form is equivalent to constructing a basis  $\{\vec{v}_i\}$  such that  $A\vec{v}_i$  can be written as a linear combination of the vectors  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_i\}$ , for all  $i$ .

One of these equations should immediately stand out - specifically,  $A\vec{v}_1 = \lambda_1\vec{v}_1$ , since this is of course an equation defining  $\vec{v}_1$  to be an eigenvector of  $A$  with eigenvalue  $\lambda_1$ . So if any square matrix can be written in an upper-triangular form, then any square matrix must have at least one eigenvector, even if the matrix isn't diagonalizable.

## 4 Existence of at least one eigenvector

Let's try to prove that any square matrix  $A$  has at least one eigenvector. Recall that we solve for eigenvalues and eigenvectors by considering the matrix  $A - \lambda I$ , and searching for eigenvalues  $\lambda$  that caused  $A - \lambda I$  to have a nontrivial nullspace. To do so, we viewed the determinant  $|A - \lambda I|$  as a polynomial  $P(\lambda)$  in  $\lambda$ , and searched for its roots.

However, the Fundamental Theorem of Algebra tells us that every polynomial must have at least one distinct (possibly complex) root! Thus, we will obtain at least one eigenvalue  $\lambda$  such that  $A - \lambda I$  has a nontrivial nullspace. By considering an element  $\vec{v} \in \text{Null}(A - \lambda I)$ , we see immediately that

$$\begin{aligned} (A - \lambda I)\vec{v} &= \vec{0} \\ \implies A\vec{v} &= \lambda\vec{v}, \end{aligned}$$

so we have obtained an eigenvalue-eigenvector pair  $(\lambda, \vec{v})$  for our matrix  $A$ , even if  $A$  were not diagonalizable.

## 5 Guessing a basis

So we know how to compute some  $\vec{v}_1$  such that  $A\vec{v}_1 = \lambda_1\vec{v}_1$ . But what about the remaining  $\vec{v}_i$  for  $i \geq 2$ ? To determine these  $\vec{v}_i$ , we will make a guess. We will make a guess that the  $\vec{v}_i$  not only form a basis, but in fact form an *orthonormal* basis - in other words, that  $\vec{v}_i \perp \vec{v}_j$  for all  $i \neq j$ , and that  $\|\vec{v}_i\| = 1$  for all  $i$ . Consider an arbitrary such orthonormal basis, starting with the known eigenvector  $\vec{v}_1$  (normalized to be of magnitude 1) and constructing the remaining vectors using the Gram-Schmidt process. First, let's place the  $n - 1$  arbitrarily chosen vectors, which we will denote as  $\vec{b}_i$  (defining  $\vec{b}_1 = \vec{v}_1$  for convenience), in a matrix  $R_{n-1}$  defined as

$$R_{n-1} = \begin{bmatrix} \vec{b}_2 & \vec{b}_3 & \cdots & \vec{b}_n \end{bmatrix},$$

so our full upper-triangular basis will look like

$$U_n = \begin{bmatrix} \vec{v}_1 & R_{n-1} \end{bmatrix},$$

using block matrix notation. Notice that as the  $\vec{b}_i$  are orthonormal,  $R_{n-1}^T R_{n-1} = I_{n-1}$ , where  $I_k$  represents the  $k$ -dimensional identity matrix.

Is this basis upper-triangular? Let's find out, by seeing how  $A$  acts on these basis vectors, working in this

upper triangular basis. Observe that

$$AU_n = A \begin{bmatrix} \vec{v}_1 & R_{n-1} \end{bmatrix} = \begin{bmatrix} A\vec{v}_1 & AR_{n-1} \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{v}_1 & AR_{n-1} \end{bmatrix}.$$

Expressing this result in the same basis, taking advantage of the fact that the inverse of an orthogonal matrix is its transpose, we see that

$$\begin{aligned} U_n^{-1}AU_n &= \begin{bmatrix} \vec{v}_1 & R_{n-1} \end{bmatrix}^T \begin{bmatrix} \lambda_1 \vec{v}_1 & AR_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \vec{v}_1^T \\ R_{n-1}^T \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{v}_1 & AR_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{v}_1^T AR_{n-1} \\ \lambda_1 R_{n-1}^T \vec{v}_1 & R_{n-1}^T AR_{n-1} \end{bmatrix}, \end{aligned}$$

What does this matrix look like? Ideally, we'd like it to be upper-triangular, and so of the form

$$\begin{bmatrix} \lambda_1 & ? & ? & \dots & ? \\ 0 & \lambda_2 & ? & \dots & ? \\ 0 & 0 & \lambda_3 & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

The top two blocks of  $U_n^{-1}AU_n$  look alright, since the top row of an upper triangular matrix does not have to contain any zeros.

The bottom two blocks, however, might pose more of an issue. Specifically, comparing the two matrices above, for  $U_n^{-1}AU_n$  to be upper triangular,  $\lambda_1 R_{n-1}^T = \vec{0}$ , and  $R_{n-1}^T AR_{n-1}$  must itself be an  $n-1$ -dimensional square upper triangular matrix.

Let's try to verify the first of our requirements. Recall that we chose the  $\vec{b}_i$  to be orthonormal to  $\vec{v}_1$ , so  $\vec{b}_i^T \vec{v}_1 = 0$  whenever  $i \neq 1$ . So then

$$\lambda_1 R_{n-1}^T \vec{b}_1 = \lambda_1 \begin{bmatrix} - & \vec{b}_2^T & - \\ - & \vec{b}_3^T & - \\ & \vdots & \\ - & \vec{b}_n^T & - \end{bmatrix} \vec{v}_1 = \lambda_1 \begin{bmatrix} \vec{b}_2^T \vec{v}_1 \\ \vec{b}_3^T \vec{v}_1 \\ \vdots \\ \vec{b}_n^T \vec{v}_1 \end{bmatrix} = \lambda_1 \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0},$$

as desired! Therefore, we can rewrite

$$U_n^{-1}AU_n = \begin{bmatrix} \lambda_1 & \vec{v}_1^T AR_{n-1} \\ \vec{0} & R_{n-1}^T AR_{n-1} \end{bmatrix}.$$

Unfortunately, recall that we chose the columns of  $R_{n-1}$  essentially arbitrarily, so long as together with  $\vec{v}_1$  they formed an orthonormal basis for all of  $n$ -dimensional space. So there is no guarantee that  $R_{n-1}^T AR_{n-1}$  is upper-triangular, meaning that we aren't quite done yet.

## 6 Low-dimensional cases

The above approach doesn't always take us to an upper-triangular form, but it certainly takes us closer to one. When *does* it take us to upper-triangular form? Exactly when

$$R_{n-1}^T A R_{n-1}$$

is itself upper-triangular. And in what case is that matrix upper-triangular? Well, if it's  $1 \times 1$ , then it is trivially upper-triangular, so we'd be done. In other words, we know that when  $n = 2$ , the above approach yields an upper-triangulation of the original matrix.

Let's take the time to work this out algebraically. Let  $M_2$  be a  $2 \times 2$  matrix. The above approach lets us construct an orthonormal basis

$$U_2 = \begin{bmatrix} \vec{v}_1 & R_1 \end{bmatrix}$$

such that

$$U_2^{-1} M_2 U_2 = \begin{bmatrix} \lambda_1 & \vec{v}_1^T M_2 R_1 \\ \vec{0} & R_1^T M_2 R_1 \end{bmatrix}.$$

But in the  $2 \times 2$  case,  $R_1$  is simply another unit vector, orthogonal to  $\vec{v}_1$ . Let this vector be  $\vec{v}_2$ , so

$$U_2 = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}.$$

Then we see that

$$U_2^{-1} M_2 U_2 = \begin{bmatrix} \lambda_1 & \vec{v}_1^T M_2 \vec{v}_2 \\ \vec{0} & \vec{v}_2^T M_2 \vec{v}_2 \end{bmatrix}.$$

Since all the components of the above matrix are in fact  $1 \times 1$ , we have expressed  $M_2$  in upper-triangular form!

This step was fairly trivial, since we were just applying our result in a special case. Let's look at the next, slightly less trivial case: when  $n = 3$ . Consider some  $3 \times 3$  matrix  $M_3$ , and construct an orthonormal basis

$$U_3 = \begin{bmatrix} \vec{v}_1 & R_2 \end{bmatrix}$$

such that

$$U_3^{-1} M_3 U_3 = \begin{bmatrix} \lambda_1 & \vec{v}_1^T M_3 R_2 \\ \vec{0} & R_2^T M_3 R_2 \end{bmatrix}.$$

As mentioned before, since  $R_2^T M_3 R_2$  is not necessarily upper-triangular, we run into a problem with our solution.

But wait!  $R_2^T M_3 R_2$  is a  $2 \times 2$  matrix. And we just saw how to upper-triangularize arbitrary  $2 \times 2$  matrices! Let

$$M_2 = R_2^T M_3 R_2$$

and upper-triangularize it as  $U_2^{-1} M_2 U_2$  for some orthogonal basis  $U_2$ .

Ideally, we'd be able to combine our "partial" upper-triangularization of  $M_3$  with this complete upper-triangularization of  $M_2$  in order to obtain an upper-triangularization of  $M_3$  itself. Making substitutions,

we might conjecture that an upper-triangularization might look something like

$$\begin{bmatrix} \lambda_1 & \vec{q}^T \\ \vec{0} & U_2^{-1}M_2U_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \vec{q}^T \\ \vec{0} & U_2^T R_2^T M_3 R_2 U_2 \end{bmatrix}.$$

Notice that we don't really care about the values of the elements above the diagonal, since they don't affect whether or not our result is upper-triangular, so we just denote them as ?.

Can we construct a change of basis to write  $M_3$  in the above form? Well, observe that the above form can be further rewritten as

$$\begin{bmatrix} \lambda_1 & \vec{q}^T \\ \vec{0} & (R_2U_2)^T M_3 (R_2U_2) \end{bmatrix}.$$

In other words, it looks very much like how  $U_3$  acted on  $M_3$ , except with  $R_2U_2$  instead of just  $R_2$ . Thus, based on what the above  $U_3$  looked like, we can conjecture that the alternative change of basis

$$U_3 = \begin{bmatrix} \vec{v}_1 & R_2U_2 \end{bmatrix}$$

will rewrite  $A$  in upper-triangular form. Let's check this out and see if it works. Recall that we constructed

$$U_2 = \begin{bmatrix} \vec{v}_2 & \vec{v}_3 \end{bmatrix},$$

where  $\vec{v}_2$  is an eigenvector of  $M_2$  with eigenvalue  $\lambda_2$  and  $\vec{v}_3 \perp \vec{v}_2$ . (Notice that we have incremented subscript indices to avoid ambiguity.) Thus, our new  $U_3$  looks like

$$U_3 = \begin{bmatrix} \vec{v}_1 & R_2U_2 \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & R_2\vec{v}_2 & R_2\vec{v}_3 \end{bmatrix}.$$

We wish to compute  $U_3^{-1}M_3U_3$  and verify that it is upper-triangular. To do so, we need to compute  $U_3^{-1}$ .

Previously, we could simply write  $U_3^{-1} = U_3^T$ , because  $U_3$  was obviously orthogonal by the construction of  $R_2$  using Gram-Schmidt. However, now, we need to prove that our newly constructed  $U_3$  is still orthogonal. To do so, we can look at its columns. Observe that the second and third columns of  $U_3$ ,  $R_2\vec{v}_2$  and  $R_2\vec{v}_3$ , both lie in the column space of  $R_2$ , which by construction is orthogonal to the first column  $\vec{v}_1$ . Furthermore, we see that the inner product of the second and third columns is

$$\begin{aligned} (R_2\vec{v}_2)^T (R_2\vec{v}_3) &= \vec{v}_2^T (R_2^T R_2) \vec{v}_3 \\ &= \vec{v}_2^T I_2 \vec{v}_3 \\ &= 0, \end{aligned}$$

relying on the fact that  $R_2^T R_2 = I_2$  as  $R_2$  is orthonormal, and  $\vec{v}_2^T \vec{v}_3 = 0$  as the two vectors were constructed to be orthogonal. Thus, all the columns of  $U_3$  are mutually orthogonal. To verify that they are of unit magnitude, we can simply compute their squared magnitude through inner products, where we see that

$$\begin{aligned} \vec{v}_1^T \vec{v}_1 &= 1 \\ (R_2\vec{v}_2)^T (R_2\vec{v}_2) &= \vec{v}_2^T R_2^T R_2 \vec{v}_2 = \vec{v}_2^T \vec{v}_2 = 1 \\ (R_2\vec{v}_3)^T (R_2\vec{v}_3) &= \vec{v}_3^T R_2^T R_2 \vec{v}_3 = \vec{v}_3^T \vec{v}_3 = 1, \end{aligned}$$

since  $\vec{v}_1$ ,  $\vec{v}_2$ , and  $\vec{v}_3$  were all constructed to be of unit magnitude.

Therefore, we have shown that  $U_3$  forms an orthogonal basis, so  $U_3^{-1} = U_3^T$ . Thus, using similar techniques

to what we did in the “partial” upper-triangularization, we can rewrite  $M_3$  in this new basis as

$$\begin{aligned}
 U_3^{-1}M_3U_3 &= U_3^T M_3 U_3 \\
 &= \begin{bmatrix} \vec{v}_1^T \\ (R_2 U_2)^T \end{bmatrix} M_3 \begin{bmatrix} \vec{v}_1 & R_2 U_2 \end{bmatrix} \\
 &= \begin{bmatrix} \vec{v}_1^T \\ (R_2 U_2)^T \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{v}_1 & M_3 R_2 U_2 \end{bmatrix} \\
 &= \begin{bmatrix} \lambda_1 & \vec{?} \\ U_2^T (R_2^T \lambda_1 \vec{v}_1^T) & U_2^T R_2^T M_3 R_2 U_2 \end{bmatrix} \\
 &= \begin{bmatrix} \lambda_1 & \vec{?} \\ \vec{0} & U_2^T M_2 U_2 \end{bmatrix} \\
 &= \begin{bmatrix} \lambda_1 & ? & ? \\ 0 & \lambda_2 & ? \\ 0 & 0 & ? \end{bmatrix},
 \end{aligned}$$

so we have successfully placed  $M_3$  in upper-triangular form!

## 7 Induction

Let’s take a quick step back. What have we done so far?

First, we developed a fairly intuitive change of basis that took an arbitrary  $n \times n$  matrix to a “partial” upper-triangular form. Then, we saw that in the  $2 \times 2$  case, this change of basis actually took our matrix to the final upper-triangular form. And now we’ve just seen that we can use our technique to upper-triangularize  $2 \times 2$  matrices to upper-triangularize arbitrary  $3 \times 3$  matrices as well.

What’s next? Well, this approach is practically crying out for induction to be applied! Let’s try to show, given the ability to upper-triangularize arbitrary  $(n - 1) \times (n - 1)$  matrices using an orthogonal change of basis, that we can upper-triangularize arbitrary  $n \times n$  matrices using an orthogonal change of basis as well.

Consider an arbitrary  $n \times n$  matrix  $A$ . We know that we can produce an unit eigenvector of  $A$   $\vec{v}_1$  with eigenvalue  $\lambda_1$  that, along with the columns of the matrix  $R_{n-1}$  (produced using Gram-Schmidt), form an orthogonal basis of  $n$ -dimensional space.

In analogy with our approach for the case of  $n = 3$ , we then consider the  $(n - 1) \times (n - 1)$  matrix

$$M_{n-1} = R_{n-1}^T A R_{n-1}$$

and, using our inductive hypothesis, produce an orthogonal matrix  $U_{n-1}$  such that

$$U_{n-1}^{-1} M_{n-1} U_{n-1}$$

is upper-triangular.

Then, we let

$$U_n = \begin{bmatrix} \vec{v}_1 & R_{n-1} U_{n-1} \end{bmatrix}.$$

Before, we showed that  $U_3$  was orthogonal by considering each pair of its columns. Clearly, that approach

is not going to be as easy now for  $U_n$ , since there are  $n$  columns, not just 3. Instead, we will show that  $U_n^T U_n = I_n$ , which clearly implies that it is orthogonal.

This can be done as follows:

$$\begin{aligned} U_n^T U_n &= \begin{bmatrix} \vec{v}_1^T \\ (R_{n-1} U_{n-1})^T \end{bmatrix} \begin{bmatrix} \vec{v}_1 & R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \vec{v}_1^T \vec{v}_1 & \vec{v}_1^T R_{n-1} U_{n-1} \\ U_{n-1}^T R_{n-1}^T \vec{v}_1 & U_{n-1}^T R_{n-1}^T R_{n-1} U_{n-1} \end{bmatrix}. \end{aligned}$$

Let's look at each of the components of the above matrix. Since it was chosen to be a unit vector,  $\vec{v}_1^T \vec{v}_1 = 1$ . As  $R_{n-1}$  was constructed using Gram-Schmidt to have its columns be orthogonal to  $\vec{v}_1$ , we have that  $\vec{v}_1^T R_{n-1} = \vec{0}^T$  and that  $R_{n-1}^T \vec{v}_1 = \vec{0}$ .

Looking at the bottom-right term, we recall that  $R_{n-1}$  was constructed to be an orthogonal matrix, so  $R_{n-1}^T R_{n-1} = I_2$ . Furthermore,  $U_{n-1}$  was constructed to be an orthogonal basis for  $M_{n-1}$ , so  $U_{n-1}^T U_{n-1} = I_{n-1}$ . Thus,

$$U_{n-1}^T (R_{n-1}^T R_{n-1}) U_{n-1} = U_{n-1}^T I_{n-1} U_{n-1} = U_{n-1}^T U_{n-1} = I_{n-1},$$

canceling out the middle terms first. Putting all of this together, we see that

$$U_n^T U_n = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & I_{n-1} \end{bmatrix} = I_n,$$

so  $U_n$  is indeed orthogonal, as we expected.

Now that we have shown  $U_n$  is orthogonal, we can write  $U_n^{-1} = U_n^T$ . Reexpressing  $A$  in this change of basis, we see that

$$\begin{aligned} U_n^{-1} A U_n &= U_n^T A U_n \\ &= \begin{bmatrix} \vec{v}_1^T \\ (R_{n-1} U_{n-1})^T \end{bmatrix} A \begin{bmatrix} \vec{v}_1 & R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \vec{v}_1^T \\ (R_{n-1} U_{n-1})^T \end{bmatrix} \begin{bmatrix} A \vec{v}_1 & A R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \vec{v}_1^T \\ (R_{n-1} U_{n-1})^T \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{v}_1 & A R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \vec{v}_1^T \vec{v}_1 & \vec{?} \\ U_{n-1}^T R_{n-1}^T \lambda_1 \vec{v}_1 & U_{n-1}^T R_{n-1}^T A R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{?} \\ \lambda_1 U_{n-1}^T (R_{n-1}^T \vec{v}_1) & U_{n-1}^T R_{n-1}^T A R_{n-1} U_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{?} \\ \vec{0} & U_{n-1}^T M_{n-1} U_{n-1} \end{bmatrix}. \end{aligned}$$

Since  $U_{n-1}^T M_{n-1} U_{n-1}$  is upper-triangular, so is  $U_n^{-1} A U_n$ , so we have successfully upper-triangularized an arbitrary  $n \times n$  matrix  $A$  with an orthogonal change of basis by applying the inductive hypothesis.



By induction, we can therefore upper-triangularize *arbitrary* square matrices using orthogonal changes of basis, which is what we had aimed to prove! Awesome!

## 8 Schur Decomposition

There's are still a couple loose ends to clear up, however. Recall that we had initially hoped for the elements along the main diagonal of  $\tilde{A}$  to be the eigenvalues of  $A$ . But though our construction made  $\lambda_1$  an eigenvalue, the remaining  $\lambda_i$  were eigenvalues of different matrices, and we have not yet seen whether they are also eigenvalues of  $A$  itself.

To see that that is in fact the case, recall that we have just shown that we can write

$$A = U\tilde{A}U^{-1} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & ? & ? & \dots & ? \\ 0 & \lambda_2 & ? & \dots & ? \\ 0 & 0 & \lambda_3 & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_n \\ | & & | \end{bmatrix}^{-1},$$

where the  $\lambda_i$  are not necessarily the eigenvalues of  $A$ . Consider a particular  $\lambda_i$ . To show that it is an eigenvalue of  $A$ , we must show that  $|A - \lambda_i I| = 0$ . Recalling that  $|AB| = |A||B|$ , we have that

$$\begin{aligned} |A - \lambda_i I| &= |U\tilde{A}U^{-1} - \lambda_i U U^{-1}| \\ &= |U(\tilde{A} - \lambda_i I)U^{-1}| \\ &= |U| \cdot |\tilde{A} - \lambda_i I| \cdot |U^{-1}|. \end{aligned}$$

Let's look at each of the elements of this product individually. First, observe that

$$|U| \cdot |U^{-1}| = |UU^{-1}| = |I| = 1,$$

so we can cancel the determinant of  $U$  with the determinant of its inverse, to write

$$|A - \lambda_i I| = |\tilde{A} - \lambda_i I|.$$

In other words, we have shown that the characteristic polynomial of  $A$  remained unchanged under a change of basis. Now, observe that

$$\tilde{A} - \lambda_i I = \begin{bmatrix} \lambda_1 - \lambda_i & ? & ? & \dots & ? \\ 0 & \lambda_2 - \lambda_i & ? & \dots & ? \\ 0 & 0 & \lambda_3 - \lambda_i & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n - \lambda_i \end{bmatrix}.$$

Thus, the  $i$ th pivot element of the above matrix must be 0, since it will equal  $\lambda_i - \lambda_i = 0$ . If we have an upper-triangular matrix with only  $n - 1$  nonzero pivots, it is clear that it has linearly dependent columns, so its determinant is zero. Thus,

$$|\tilde{A} - \lambda_i I| = 0 \implies |A - \lambda_i I| = 0,$$

so each  $\lambda_i$  is an eigenvalue of  $A$ , as expected! This way of representing  $A$  is known as the *Schur Decompo-*

sition of  $A$ .

## 9 Complex Inner Products

There's one subtlety that we skipped over in the above proof. Specifically, we assumed throughout that the notions of orthogonality and inner products were defined on our vector spaces. But how do you take the inner product of two complex-valued vectors? If you try to reuse the definition of the dot product, you'll get some weird results that cause our understanding of these concepts to break down. For instance,

$$\left\| \begin{bmatrix} i \\ 1 \end{bmatrix} \right\|^2 = \begin{bmatrix} i \\ 1 \end{bmatrix} \cdot \begin{bmatrix} i \\ 1 \end{bmatrix} = -1 + 1 = 0,$$

which doesn't seem to make sense, since only the zero vector should have a norm of zero.

For now, we should assume that we are working in the vector space of reals  $\mathbb{R}^n$  using the standard definition of the dot product. This, however, requires all the  $\lambda_i$  to be real, which may not always be the case. For now, we will make that assumption, though in future lectures we will see a small generalization of the dot product which will ensure our above result is true in all cases.

## 10 The Spectral Theorem

Finally, we will use our decomposition to obtain some interesting results about diagonalizing real, symmetric matrices.

First, we claim that the eigenvalues of a real, symmetric matrix are all themselves real. Why might we suspect this to be true? Well, after working with many real, symmetric matrices, we might observe this fact to be true. Alternatively, by looking at the characteristic polynomial of a symmetric  $2 \times 2$  matrix

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

we see that it is of the form

$$(a - \lambda)(c - \lambda) - b^2 = \lambda^2 - (a + c)\lambda + (ac - b^2).$$

But the determinant of the above quadratic is

$$\Delta = (a + c)^2 - 4(ac - b^2) = (a - c)^2 + 4b^2$$

which is clearly always positive, so  $2 \times 2$  real symmetric matrices always have real eigenvalues.

Thus, our conjecture is a reasonable one to make. Let's try to prove it. Consider a real, symmetric  $n \times n$  matrix  $A = A^T$  and an eigenvalue  $\lambda$  of  $A$ . By the definition of eigenvectors, there exists some nonzero vector  $x$  such that

$$Ax = \lambda x.$$

To show  $\lambda$  is real, we'd like to get a result that looks like  $\lambda = \bar{\lambda}$ . So, striking out blindly, we can take the conjugate to get  $\bar{\lambda}$  involved somehow, to obtain

$$A\bar{x} = \bar{\lambda}\bar{x}.$$

Notice that  $A = \bar{A}$ , since we assumed  $A$  was real. Now, let's try to take advantage of  $A$ 's symmetric nature, by taking the transpose and using the fact that  $A = A^T$ , to obtain

$$\bar{x}^T A = \bar{x}^T \bar{\lambda}.$$

At this point, it should be fairly clear that we can post-multiply both sides by  $x$  to obtain

$$\begin{aligned} \bar{x}^T A x &= \bar{\lambda} \bar{x}^T x \\ \implies \lambda \bar{x}^T x &= \bar{\lambda} \bar{x}^T x \\ \implies (\lambda - \bar{\lambda}) \bar{x}^T x &= 0. \end{aligned}$$

So by basic arithmetic, either  $\lambda = \bar{\lambda}$ , or  $\bar{x}^T x = 0$ . But since we chose  $x$  to be a nonzero vector, the former equality must be the one that is true, so  $\lambda$  is real, as desired.

Now, we will make a stronger claim. We assert that, not only are all the eigenvalues of  $A$  real, but that  $A$  can be *diagonalized*, meaning that it has  $n$  linearly independent eigenvectors. Furthermore, we claim that these eigenvectors can be chosen such that they are all orthogonal.

Again, this can be seen by experimenting with the  $2 \times 2$  case or by just looking at the eigendecomposition of some random larger symmetric matrices. We omit the demonstration of this fact here, and will proceed straight to a proof.

What is our goal? We wish to show that we can express

$$A = Q^T \Lambda Q,$$

where  $Q$  is an orthogonal matrix whose columns are the eigenvectors of  $A$ , and  $\Lambda$  is a diagonal matrix containing the eigenvalues of  $A$ .

Notice that, since  $\Lambda$  is a diagonal matrix, it is also upper-triangular, with the elements along its diagonal clearly being the eigenvalues of  $A$ . Thus, our desired diagonalization of  $A$  is also a Schur decomposition of  $A$ .

So the first question to ask should be: can we construct a Schur decomposition of an arbitrary real, symmetric matrix  $A$ ? The critical assumption needed when doing so was that all the  $\lambda_i$  were real produced during the induction, as otherwise our arguments related to orthogonality broke down. Let's look at our procedure and try to show that this is the case.

First, consider  $\lambda_1$ .  $\lambda_1$  was chosen to be an eigenvalue of  $A$ . But since all the eigenvalues of  $A$  are real (since  $A$  is symmetric, using the result from above), we know that  $\lambda_1$  is real. Great!

Next, let's look at  $\lambda_2$ . Looking at the induction,  $\lambda_2$  was chosen to be an eigenvalue of  $M_{n-1} = R_{n-1}^T A R_{n-1}$ , where  $R_{n-1}$  was an  $n \times (n-1)$  orthogonal matrix constructed in a particular fashion. Right now, the exact construction of  $R_{n-1}$  isn't super important. What *is* important is that, as  $A = A^T$ ,

$$M_{n-1}^T = (R_{n-1}^T A R_{n-1})^T = R_{n-1}^T A^T R_{n-1} = R_{n-1}^T A R_{n-1} = M_{n-1},$$

so  $M_{n-1}$  is symmetric. And so, as  $\lambda_2$  is an eigenvalue of the symmetric matrix  $M_{n-1}$ , it is itself real. This looks promising!

In a similar manner,  $\lambda_3$  is an eigenvalue of  $M_{n-2} = R_{n-2}^T M_{n-1} R_{n-2}$ , which is symmetric, so  $\lambda_3$  is itself real. And this recursive argument can be continued in the natural manner to show that *all* the  $\lambda_i$  are real!

Awesome<sup>1</sup>!

Since all the  $\lambda_i$  are real, the induction involved in the Schur form proof works out, so we can write

$$A = U\tilde{A}U^T,$$

where  $U$  is an orthogonal matrix and  $\tilde{A}$  is upper-triangular. Our goal is to show that  $\tilde{A} = \Lambda$  - in other words, that  $\tilde{A}$  is in fact diagonal! Since we know that  $\tilde{A}$  is already upper-triangular, one way of doing this would be to show that

$$\tilde{A} = \tilde{A}^T,$$

so the “upper” part of  $\tilde{A}$  consists entirely of zeros as well, so it is diagonal. How can we get  $\tilde{A}^T$  from our upper-triangular decomposition? We might as well just blindly take the transpose of the entire equation, just so we get the desired term *somewhere*. Doing so, we obtain

$$A^T = U\tilde{A}^T U^T.$$

But since  $A$  is symmetric,  $A = A^T$ , so we can write

$$A = U\tilde{A}^T U^T.$$

So we have shown that, when working in the basis of  $U$ ,  $A$  becomes both  $\tilde{A}$  and  $\tilde{A}^T$ ? How can this be true? Clearly, the only way for this to be possible is if

$$\tilde{A} = \tilde{A}^T,$$

as desired! Thus, we have shown that we can write

$$A = U\tilde{A}U^T,$$

where  $\tilde{A}$  is a diagonal matrix made up of  $A$ 's eigenvalues, and  $U$  is an orthogonal matrix. So we have diagonalized  $A$ ! This completes the proof of what is known as the *real spectral theorem*.

### Contributors:

- Rahul Arya.
- Anant Sahai.

---

<sup>1</sup>We could alternatively frame this argument using induction, like we did when deriving Schur form, if you're more comfortable with that. But both approaches are fundamentally equivalent and lead to the same result. Indeed, very soon, when we consider how inner products generalize to the complex numbers, we'll arrive at an even simpler proof of the spectral theorem that uses neither this iterative argument nor induction, but instead falls right out of Schur form!