

Lecture 19 - Notes

Rahul Arya

April 2019

1 Overview

Last lecture, we looked at the Schur decomposition of a matrix, where we showed that any matrix with purely real eigenvalues can be written in upper triangular form with respect to an orthogonal basis. As a consequence of the existence of this decomposition, we demonstrated that any real symmetric matrix with only real eigenvalues has a full set of real eigenvalues and eigenvectors, with the additional property that these eigenvectors are mutually orthogonal. In this lecture, we will strengthen both those results to work with complex eigenvalues. In addition, we will present a new matrix decomposition that applies even to general rectangular matrices, known as the *singular value decomposition*, and explore some of its properties.

2 Real Spectral Theorem

Before generalizing the Schur decomposition of an arbitrary matrix, let's first consider the simpler case of a given real symmetric matrix A . If all of A 's eigenvalues were real, then we could write it in Schur form and use our results from last lecture to produce an eigendecomposition into an orthogonal basis. As it turns out (though it may not yet be obvious), this is in fact always the case!

Let's try to prove it. Consider some eigenvalue λ of A . By the definition of eigenvectors, there exists some nonzero vector x such that

$$Ax = \lambda x.$$

To show λ 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 $\bar{x}^T x = 0$. Ideally, we'd disprove the second equality in order to prove the first.

3 Inner Products

Let's explore some of the properties of the strange expression: $\bar{x}^T x$. Clearly, if x had only real components, it would be equivalent to $x^T x = \|x\|^2$, which we know how to deal with. But what if x was complex? Let's look at a toy example where x has two components, so we can express it as

$$x = \begin{bmatrix} a + bj \\ c + dj \end{bmatrix}$$

for some reals a, b, c , and d . Then, we find that

$$\bar{x}^T x = a^2 + b^2 + c^2 + d^2,$$

which is clearly always a real nonnegative number. In fact, the only way for this number to be zero is if $x = \vec{0}$ itself.

As it turns out, the quantity $\bar{x}^T x$ is the generalization of the norm of a vector to the complex field. Recall from last lecture that our old definition of $\|x\| = \sqrt{x^T x}$ broke down, since it would equal zero for some nonzero (but complex) vectors. This new definition clearly lacks that drawback.

Furthermore, it seems natural to similarly generalize the notion of the inner product of two complex-valued vectors such that

$$\langle x, y \rangle = \bar{y}^T x,$$

based on the premise that $\|x\| = \sqrt{\langle x, x \rangle}$. While we will not prove it here, it can be shown that the above definition of inner products allows us to define a consistent notion of orthogonality, such that we can calculate the Schur decomposition of *any* matrix, even one with complex entries or eigenvalues.

4 Back to the Real Spectral Theorem

We now have the tools necessary to finish off our proof from before. Since x was an eigenvector of A , we know that $x \neq \vec{0}$. Consequently, $\bar{x}^T x = \|x\|^2 \neq 0$, so we have that $\lambda = \bar{\lambda}$, completing the proof.

Thus, as we have shown that symmetric matrices have real eigenvalues, we can apply the result from last time to conclude that any symmetric matrix can be diagonalized into a set of orthogonal eigenvectors with real eigenvalues. It is straightforward to demonstrate that these eigenvectors can also be chosen to be real, but we will omit the proof here.

However, there exists a somewhat simpler way of getting from the Schur decomposition to the real spectral theorem. In defining inner products over complex fields, we noted that the Schur decomposition could now be produced for arbitrary matrices A . That is to say, we can always write such a matrix in the form

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

where Q is an orthogonal matrix (as per our more general definition of inner products) and U is upper triangular.

When Q was real and orthogonal, we were able to show that $Q^T Q = I \implies Q^T = Q^{-1}$, which was the first step towards obtaining the real spectral theorem. Let's see if we can obtain a similar result in the complex case. Let the columns of Q be defined as the vectors \vec{q}_1 through \vec{q}_n , so

$$Q = \begin{bmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{bmatrix}.$$

Since Q is orthogonal, its columns are mutually orthonormal. Thus, $\overline{\vec{q}_i}^T \vec{q}_j = 0$ for $i \neq j$, and $\overline{\vec{q}_i}^T \vec{q}_i = 1$.

It is natural to conjecture that our result in the real case still holds, where $Q^T = Q^{-1}$. To test this, we can evaluate $Q^T Q$, to obtain

$$\begin{aligned} Q^T Q &= \begin{bmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{bmatrix}^T \begin{bmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} - & \vec{q}_1^T & - \\ \vdots & & \\ - & \vec{q}_n^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} \vec{q}_1^T \vec{q}_1 & \vec{q}_1^T \vec{q}_2 & \cdots & \vec{q}_1^T \vec{q}_n \\ \vec{q}_2^T \vec{q}_1 & \vec{q}_2^T \vec{q}_2 & \cdots & \vec{q}_2^T \vec{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{q}_n^T \vec{q}_1 & \vec{q}_n^T \vec{q}_2 & \cdots & \vec{q}_n^T \vec{q}_n \end{bmatrix}. \end{aligned}$$

Unfortunately, we have no idea what the resultant matrix could look like! In particular, it may or may not be I depending on Q , so it's clear that it's possible for $Q^T \neq Q^{-1}$ in the complex case.

What would we like the resultant matrix to look like. Well, we want it to be the identity matrix, so we should expect 1s along the diagonal, and 0s everywhere

else. In the real case, we were able to write $\vec{q}_i^T \vec{q}_i = 1$ and $\forall i \neq j (\vec{q}_i^T \vec{q}_j = 0)$ and get the desired result, but the conjugates in our new definition of the inner product mess things up somewhat. Ideally, we'd have conjugates over all the \vec{q}_i^T terms, so we could apply the orthonormality properties of our set of vectors.

But wait! If we want conjugates over the \vec{q}_i terms, all we have to do is take the conjugate of Q^T ! In other words,

$$\overline{Q^T} Q = \begin{bmatrix} - & \overline{\vec{q}_1^T} & - \\ & \vdots & \\ - & \overline{\vec{q}_n^T} & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \overline{\vec{q}_1^T} \vec{q}_1 & \overline{\vec{q}_1^T} \vec{q}_2 & \cdots & \overline{\vec{q}_n^T} \vec{q}_1 \\ \overline{\vec{q}_2^T} \vec{q}_1 & \overline{\vec{q}_2^T} \vec{q}_2 & \cdots & \overline{\vec{q}_n^T} \vec{q}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\vec{q}_n^T} \vec{q}_1 & \overline{\vec{q}_n^T} \vec{q}_2 & \cdots & \overline{\vec{q}_n^T} \vec{q}_n \end{bmatrix} = I,$$

so we find that $Q^{-1} = \overline{Q^T}$. Typically, the quantity $\overline{Q^T}$ is known as the *Hermitian transpose* of Q , and is written variously as Q^H , Q^\dagger (particularly in physics), or Q^* .

Thus, the existence of the Schur decomposition tells us that any symmetric matrix A can be written as

$$A = QUQ^{-1} = QUQ^*.$$

Last time, we took the transpose of A in order to show that U was not only upper triangular, but diagonal as well. This time, it should be intuitively clear that we need to take not just the transpose, but the Hermitian transpose. Since we are interested in real symmetric matrices A , we have $A = A^T = A^*$, so

$$A = A^* = (QUQ^*)^* = QU^*Q^*,$$

implying that $U = U^*$. Since U is upper triangular and U^* is lower triangular, we see that they both must be diagonal matrices, so A is diagonalizable.

But if U is diagonal, then $U = U^T$, so $U^* = \overline{U^T} = \overline{U}$. As we saw that $U = U^*$, we can substitute to find that $U = \overline{U}$, so all of its entries must be real. Since the entries of U are the eigenvalues (repeated according to their multiplicities) of A , we have therefore also shown that the eigenvalues of a real symmetric matrix A are real, without having to go through an additional proof.

Moreover, notice that we didn't actually require that A was both real and symmetric - instead, all we needed was $A = A^*$. This makes our result stronger as it applies to a more general class of matrices: for instance, letting A be

$$A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix},$$

we see that A is clearly neither real nor symmetric, but $A = A^*$, so this result still applies. These matrices are known as *Hermitian matrices*, and our result is a weaker form of what is known as the *complex spectral theorem*.

5 Singular Value Decomposition

We will now move on from the Schur decomposition to discuss another matrix decomposition, known as the *singular value decomposition*. Whereas Schur form was a strict generalization of matrix diagonalization, the singular value decomposition (or *SVD* for short) is an alternative way to factorize matrices.

As we did in the last two lectures, today we will first assert the existence of the SVD without proof and examine its properties, and next time we will prove its existence and derive a way to construct it. Unlike the decompositions we have seen so far, the SVD applies not only to square matrices, but in fact to arbitrary rectangular matrices.

Specifically, for a real $m \times n$ matrix A , we may express it as the product of a square $m \times m$ orthogonal matrix U , a “diagonal” $m \times n$ matrix Σ with nonnegative entries, and a square $n \times n$ orthogonal matrix V^T . Drawn visually, this product looks like:

$$\begin{bmatrix} A \\ \\ \end{bmatrix}_{m \times n} = \begin{bmatrix} U \\ \\ \end{bmatrix}_{m \times m} \begin{bmatrix} \Sigma \\ \\ \end{bmatrix}_{m \times n} \begin{bmatrix} V^T \\ \\ \end{bmatrix}_{n \times n}.$$

Notice that Σ is the same dimension as A , and V^T represents the transpose of some matrix V .

Now, what does it mean for a rectangular matrix Σ to be diagonal? Well, we define the diagonal entries of a matrix to be those whose row and column indices are equal, starting from the top left. So we can express Σ as

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

where the elements of the diagonal are the σ_i , and all elements off the diagonal are zero. This definition applies equally if Σ is “wide” (i.e. $n > m$) rather than “tall”, as was shown here.

Another interesting feature of the SVD is that, unlike the Schur decomposition and the eigenvalue decomposition, it is not simply the reexpression of A in a different basis, since in general $U^{-1} \neq V^T$. So what does it do? There are many possible interpretations of its behavior, but the one we will look at today involves a concept known as *outer products*.

6 Outer Products

Recall that, for real vectors \vec{x} and \vec{y} expressed as columns with n components, their inner product is defined as $\vec{y}^T \vec{x}$, which yields a 1×1 matrix typically treated as a scalar.

Similarly, we will define their *outer product* to be $\vec{x}\vec{y}^T$. Let's see what this means. Let

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^T$$

$$\vec{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T,$$

where it is possible that $m \neq n$. Then, by the definition of matrix multiplication,

$$\vec{x}\vec{y}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}.$$

So while the inner product took two vectors of the same dimension and produced a scalar, the outer product takes two vectors of possibly different dimensions and yields a matrix!

Furthermore, notice that this matrix cannot be any arbitrary matrix - since each of its columns are a scalar multiple of \vec{x} , it cannot be of a rank greater than 1. It is straightforward to show that any matrix of rank 0 or 1 can be produced by an outer product of two vectors, but we will not discuss the details here.

Now, why are we interested in the outer product? Well, recall that we can express real matrix multiplication in terms of inner products. Specifically, we know that

$$\begin{bmatrix} - & \vec{x}_1^T & - \\ \vdots & \vdots & \vdots \\ - & \vec{x}_m^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{y}_1 & \cdots & \vec{y}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{x}_1^T \vec{y}_1 & \vec{x}_1^T \vec{y}_2 & \cdots & \vec{x}_1^T \vec{y}_n \\ \vec{x}_2^T \vec{y}_1 & \vec{x}_2^T \vec{y}_2 & \cdots & \vec{x}_2^T \vec{y}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{x}_m^T \vec{y}_1 & \vec{x}_m^T \vec{y}_2 & \cdots & \vec{x}_m^T \vec{y}_n \end{bmatrix},$$

where the element at the i th row and j th column of the product of two matrices X and Y is the dot product of the i th row of X and the j th column of Y .

However, what if we were interested in the columns of X and the rows of Y instead? As it turns out, it is the case that

$$\begin{bmatrix} | & & | \\ \vec{x}_1 & \cdots & \vec{x}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \vec{y}_1^T & - \\ \vdots & \vdots & \vdots \\ - & \vec{y}_n^T & - \end{bmatrix} = \vec{x}_1 \vec{y}_1^T + \vec{x}_2 \vec{y}_2^T + \cdots + \vec{x}_n \vec{y}_n^T.$$

This result can be proved by applying the definition of matrix multiplication, but it is tedious and so will be omitted.

Instead, we will look at an example that demonstrates the main ideas behind the proof. Consider the product

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}.$$

From our knowledge of the matrix product as the composition of dot products representing each element in the result, we obtain

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 7 & 1 \cdot 6 + 2 \cdot 8 \\ 3 \cdot 5 + 4 \cdot 7 & 3 \cdot 6 + 4 \cdot 8 \end{bmatrix}.$$

We will not simplify this result, for reasons that will become clear in a moment. Now, calculating the product of these matrices using outer products, we obtain

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 6 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 & 1 \cdot 6 \\ 3 \cdot 5 & 3 \cdot 6 \end{bmatrix} + \begin{bmatrix} 2 \cdot 7 & 2 \cdot 8 \\ 4 \cdot 7 & 4 \cdot 8 \end{bmatrix}.$$

Notice that the terms in our matrix multiplication evaluated using dot products correspond exactly to the terms in our sums of outer products, so our outer product definition of matrix multiplication is consistent with our previous definitions, at least in this example. A slightly more rigorous form of this argument can be used to prove this result in general, but it is best to understand this result at an intuitive level.

7 Outer Product Form of the SVD

Now, we will use this new interpretation of matrix multiplication as a sum of outer products in order to obtain an alternative way of expressing the SVD. Recall that the SVD of a matrix A represented it as the product

$$A = U \Sigma V^T,$$

where Σ was a matrix with nonzero entries only along its main diagonal. Let A be an $m \times n$ matrix, so U is a square $m \times m$ matrix and V^T is a square $n \times n$ matrix. Additionally, without loss of generality, assume that $m \geq n$, so A is a “tall” matrix (the same results can be obtained if A is “wide” by considering A^T , which would become “tall”).

Let the columns of U be \vec{u}_1 through \vec{u}_m , the nonzero diagonal entries of Σ be σ_1 through σ_n (moving from the top-left to the bottom-right entry), and the rows of V^T be \vec{v}_1^T through \vec{v}_n^T . From our understanding of outer products, we know how to express UV^T in terms of the \vec{u}_i and \vec{v}_i . But how does Σ affect this interpretation?

Let’s consider just the first two terms of the SVD - the product $U\Sigma$. By the

outer product interpretation of matrix multiplication, we have that

$$\begin{aligned}
 U\Sigma &= \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \\
 &= \vec{u}_1 [\sigma_1 \ 0 \ \cdots 0] + \vec{u}_2 [0 \ \sigma_2 \ \cdots 0] + \cdots + \vec{u}_n [0 \ 0 \ \cdots \sigma_n] \\
 &\quad + \vec{u}_{n+1} [0 \ 0 \ \cdots 0] + \cdots + \vec{u}_m [0 \ 0 \ \cdots 0] \\
 &= \begin{bmatrix} | & | & & | \\ \sigma_1 \vec{u}_1 & \sigma_2 \vec{u}_2 & \cdots & \sigma_n \vec{u}_n \\ | & | & & | \end{bmatrix},
 \end{aligned}$$

since the inner products involving the zero row vector yield the zero matrix, and so can be disregarded.

Notice what has happened here. The σ_i have acted as coefficients for the first n columns of \vec{u} , with the subsequent columns vanishing entirely. Now, we can multiply by V and apply the outer-product interpretation of matrix multiplication to obtain

$$\begin{aligned}
 A &= U\Sigma V^T \\
 &= \begin{bmatrix} | & | & & | \\ \sigma_1 \vec{u}_1 & \sigma_2 \vec{u}_2 & \cdots & \sigma_n \vec{u}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & \vec{v}_1^T & - \\ - & \vec{v}_2^T & - \\ & \vdots & \\ - & \vec{v}_n^T & - \end{bmatrix} \\
 &= \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \cdots + \sigma_n \vec{u}_n \vec{v}_n^T,
 \end{aligned}$$

so any $m \times n$ matrix A (where $m \geq n$) can be expressed as the weighted sum of n rank-1 matrices of the form $\vec{u}_i \vec{v}_i^T$, where the set of \vec{u}_i and \vec{v}_i are each mutually orthonormal. This interpretation of the SVD is known as the *outer product form*.

8 Image Compression

This new form of the SVD has one practical application in image compression. Consider the Polish flag below:

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0.2 \\ 0.2 \\ \vdots \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}.$$

...what, that doesn't look like the Polish flag to you? Recall that any (digital) image can be considered to be a matrix of scalar values, representing pixels. When stored, a square $n \times n$ image would require n^2 scalar values to be saved, assuming no compression.

We'd like to do better, while not compromising our image quality too much. The key idea here is that rank-1 matrices can be represented as an outer product of two matrices each with n components, so they can be stored using only $\Theta(n)$ space. The (greyscale) Polish flag, consisting of two horizontal full-width stripes, is of rank 1, and so can be represented as an outer product, as shown above.

Ideally, we'd be able to store any image as a rank-1 matrix, to obtain this drastic space reduction. Of course, this is not possible - almost all $n \times n$ matrices are of full rank, so we cannot compress them all to take up $\Theta(n)$ space. This should make sense, of course, since it's clearly impossible to obtain a compression algorithm that reduces the size of *every* image with no loss of information.

Instead, the key idea is that many images can be represented well as the sum of a relatively small number (say k) of rank-1 matrices, so we can store them using only $\Theta(kn)$ space. How do we get these rank-1 matrices? Recall from the previous section that we can express any $n \times n$ matrix A as

$$A = \sigma_1 \vec{u}_1 \vec{v}_1 + \sigma_2 \vec{u}_2 \vec{v}_2 + \cdots + \sigma_n \vec{u}_n \vec{v}_n,$$

where without loss of generality we can choose our σ_i such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

It turns out (though we will not prove it) that for most matrices, including most matrices corresponding to real-world images, the magnitudes of the σ_i rapidly drop. That is to say, the later terms $\sigma_i \vec{u}_i \vec{v}_i$ for $i > k$ for some small value k tend to be significantly smaller than the first few terms. Consequently, we can approximate

$$A \approx \sigma_1 \vec{u}_1 \vec{v}_1 + \sigma_2 \vec{u}_2 \vec{v}_2 + \cdots + \sigma_k \vec{u}_k \vec{v}_k,$$

where k is much smaller than n . Therefore, we can represent A *approximately* as the sum of k rank-1 matrices, and so store it in only $\Theta(nk)$ space.

The lecture slides on Piazza and on the course site have some visual demonstrations of this, which I will not repeat here. Go check them out!

Next time, we will demonstrate the existence of the SVD, and consider another one of its interpretations.