Transition to Higher Mathematics Project

Jeffrey Chan
Perm: 8764375

University of California Santa Barbara
Department of Mathematics
Math 8

Contents

1 Definitions

First I will recall some linear algebra definitions and facts: (the source for this section is Axler, Linear Algebra Done Right)

- An *orthonormal basis* of a complex vector space V is a linearly independent set of vectors $\{v_1, \dots v_n\}$ that span V and have $v_i^* v_j = \delta_{ij}$.
- The above $*$ notation is the *Hermitian conjugate* of an $m \times n$ matrix $M \in \mathbb{C}^{m\times n}$. It is a $n \times m$ matrix formed from M by transposing M and taking the complex conjugate of each entry. It is denoted M^*. The entries of M^* are
$$(M^*)_{ij} = \overline{M_{ji}}$$
- An $n \times n$ matrix U is called *unitary* if it is invertible and $U^* = U^{-1}$. It is a theorem that if we make a matrix by taking its columns to be the vectors of an orthonormal basis then that matrix is unitary. We can think of unitary matrices as a special change of basis matrix that keeps the magnitude of each vector the same. This is called an *isometry*.
- An $n \times n$ matrix M is *symmetric* if $M^* = M$. It is easy to see that for any $m \times n$ matrix M the $n \times n$ matrix M^*M is symmetric.

Now I define the relevant vocabulary for SVD. It's helpful to recall what an eigenvalue is first:

Definition 1.1. An *eigenvalue* of an $m \times m$ matrix $M \in \mathbb{C}^{m\times m}$ is a complex number $\lambda \in \mathbb{C}$ such that there exists a non-zero vector $v \in \mathbb{C}^m$ with $Mv = \lambda v$. We call v an *eigenvector* associated with the eigenvalue λ. (We can always take v to be a unit vector by rescaling it.)

The definition of a singular value is similar, but cannot be exactly the same because if $M \in \mathbb{C}^{m\times n}$ with $m \neq n$ (if M is not square) then the input vector v would belong to $\mathbb{C}^n$ and the output vector Mv would belong to $\mathbb{C}^m$. So we generalize the definition of eigenvalues and eigenvectors like this:

Definition 1.2. A *singular value* of $M \in \mathbb{C}^{m\times n}$ is a non-negative real number σ such that there exist unit length vectors $u \in \mathbb{C}^m$ and $v \in \mathbb{C}^n$ with

$$Mv = \sigma u \quad \text{and} \quad M^*u = \sigma v.$$

We call u a *left-singular vector* corresponding to σ and v a *right-singular vector* corresponding to σ.

Now remember that for some matrices $M \in \mathbb{C}^{m\times m}$ it's possible to diagonalize, meaning

Definition 1.3. A matrix $M \in \mathbb{C}^{m\times m}$ is *diagonalizable* if there exists an invertible matrix $P \in \mathbb{C}^{m\times m}$ such that $M = PDP^{-1}$ with D a diagonal matrix.

Notice that this is a decomposition of M that can only be done if M is a square matrix because we have to use P and P^{-1}, which will have the same dimensions. It's important to remember that being square is not enough: there are square matrices that are not diagonalizable. Also, even when a square matrix is diagonalizable it may not be possible to find a *unitary* matrix P that performs the diagonalization. One might say that these are two "shortcomings" of diagonalization.

Theorem 1.4. *A symmetric matrix is diagonalizable and it has only non-negative eigenvalues. In fact we can use a unitary matrix for the diagonalization: so if M symmetric then*

$$M = UDU^*$$

for some unitary matrix U and some diagonal matrix D with all diagonal entries ≥ 0.

The singular value decomposition is also a way of writing M as a product of three matrices with the middle matrix diagonal. But it is more general because it can be done for any matrix.

Definition 1.5. The *singular value decomposition* of $M \in \mathbb{C}^{m\times n}$ consists of unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ and a matrix $\Sigma \in \mathbb{R}^{m\times n}$ with all entries $\Sigma_{ij} = 0$ if $i \neq j$ and $\Sigma_{ii} \geq 0$ such that $M = U\Sigma V^*$. Such a decomposition exists for any $M \in \mathbb{C}^{m\times n}$ (see Theorem 2.1). This decomposition can always be chosen to have the non-zero entries of Σ be decreasing: $\Sigma_{11} \geq \Sigma_{22} \geq \ldots \geq \Sigma rr \geq 0$.

2 Singular Decomposition Proof

The following proof idea is from Wikipedia, and I am expanding the idea of it.[Wikipedia]

Theorem 2.1. *Given any matrix $M \in \mathbb{C}^{m\times n}$ with complex entries there exist unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ and a matrix $\Sigma \in \mathbb{R}^{m\times n}$ with all entries $\Sigma_{ij} = 0$ if $i \neq j$ and $\Sigma_{ii} \geq 0$ such that $M = U\Sigma V^*$.*

Proof. We will assume this important result from linear algebra: A symmetric matrix is diagonalizable and has non-negative eigenvalues.

Say that M is an $m \times n$ matrix. We can make a symmetric matrix from M by taking M^*M. This is an $n \times n$ matrix. It is symmetric and so we can diagonalize it using an unitary matrix V:

$$M^*M = V\tilde{D}V^*. \tag{1}$$

Here V is an $n\times n$ unitary matrix, meaning that $V^* = V^{-1}$. The matrix $\tilde{D}$ is a diagonal matrix. It may have some 0's on its diagonal, which we would like to separate out. This can be done by re-arranging the columns of V, which does not change the fact that V is unitary (it's like re-indexing a basis). So we write $\tilde{D}$ as a block-diagonal matrix

$$\tilde{D} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}.$$

where D is a diagonal matrix with no zeros on the diagonal:

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{pmatrix}.$$

Since the eigenvalues of M^*M are non-negative and D has only the non-zero eigenvalues, D must have only positive elements on its diagonal: each $d_i > 0$. That means that we can take a square root of D: $D^{1/2}$ is a diagonal matrix whose diagonal elements are the square roots of the corresponding diagonal elements of D:

$$D^{1/2} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_r} \end{pmatrix}. \tag{2}$$

It will turn out that $D^{1/2}$ has all the singular values of M on its diagonal, so we are going to give them new names: $\sqrt{d_i} = s_i > 0$. With these new names we write

$$D^{1/2} = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_r \end{pmatrix}.$$

Also since $D^{1/2}$ is diagonal with all positive elements on its diagonal it has an inverse, $D^{-1/2}$, whose elements are

$$D^{-1/2} = \begin{pmatrix} s_1^{-1} & & 0 \\ & \ddots & \\ 0 & & s_r^{-1} \end{pmatrix}.$$

Now look back to when we wrote

$$M^*M = V \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V^*.$$

So the first r columns of V get multiplied by the non-zero part D and the last $n-r$ columns of V get multiplied by 0. So it might be interesting to separate out V into two pieces: its first r columns will form a matrix V_1 and its last r columns will form a matrix V_2: $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix}$ in block matrix form. Then our equation looks like

$$M^*M = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^* \\ V_2^* \end{pmatrix}. \tag{3}$$

We can say a bit more about these blocks V_1 and V_2 because we know that $V^*V = I_n$. Writing it out in block form we have

$$\begin{pmatrix} V_1^* \\ V_2^* \end{pmatrix} \begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & I_{n-r} \end{pmatrix}.$$

which implies that

$$V_1^*V_1 = I_r \tag{4}$$

$$V_2^*V_2 = I_{n-r} \tag{5}$$

We also have the relation $VV^* = I_n$ which gives us

$$\begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} V_1^* \\ V_2^* \end{pmatrix} = V_1 V_1^* + V_2 V_2^* = I_n. \tag{6}$$

Those relations that we listed will be useful in our next step:

We claim that $D^{1/2}V^*$ will be the last two matrices in our SVD $M = U\Sigma V^*$. We'll prove it like this. Define $U_1 = MV_1 D^{-1/2}$. Then

$$\begin{aligned} U_1 D^{1/2} V_1^* &= (MV_1 D^{-1/2}) D^{1/2} V_1^* \\ &= M(V_1 V_1^*) \\ &= M(I_n - V_2 V_2^*) \text{ by the relation } V_1 V_1^* = I_n - V_2 V_2^* \text{ above} \\ &= M - (MV_2)V_2^*. \end{aligned}$$

We're close, but we need to deal with the MV_2 term.

Go back to the equation $M^*M = V\tilde{D}V^*$ and rewrite it as $V^*M^*MV = \tilde{D}$. In block form that says

$$\begin{aligned} \begin{pmatrix} V_1^* \\ V_2^* \end{pmatrix} M^*M \begin{pmatrix} V_1 & V_2 \end{pmatrix} &= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} V_1^*M^* \\ V_2^*M^* \end{pmatrix} \begin{pmatrix} MV_1 & MV_2 \end{pmatrix} &= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} V_1^*M^*MV_1 & V_1^*M^*MV_2 \\ V_2^*M^*MV_1 & V_2^*M^*MV_2 \end{pmatrix} &= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Comparing the entries of the block matrix on the LHS and the RHS tells us a few important things:

$$\begin{aligned} V_2^*M^*MV_2 &= 0 \\ \implies (MV_2)^*(MV_2) &= 0 \\ \implies MV_2 &= 0. \end{aligned}$$

Because $A^*A = 0$ implies $A = 0$ for any matrix A.

Using that together with $U_1 D^{1/2} V_1^* = M - (MV_2)V_2^*$ from above we finally get

$$U_1 D^{1/2} V_1^* = M. \tag{7}$$

This is very close to a singular value decomposition. The only problem is that the matrices U_1 and V_1^* are not square and therefore are not unitary. We'll make them unitary by adding more columns to each of them, starting with U_1: What size is U_1? From its definition $U_1 = MV_1D^{-1/2}$ we see it must have as many rows as M, so it has m rows, and it must have as many columns as $D^{-1/2}$, which has r columns. Recall that r was the number of non-zero eigenvalues of M^*M. This cannot be larger than the rank of M, so U_1 is $m \times r$ with $r \leq m$. It has more rows than columns, so we need to add columns to make it square. The columns of U_1 are actually orthonormal vectors because

$$\begin{aligned}
U_1^*U_1 &= (MV_1D^{-1/2})^*MV_1D^{-1/2} \\
&= D^{-1/2}V_1^*M^*MV_1D^{-1/2} \\
\text{from above we have } V_1^*M^*MV_1 &= D \text{ so} \\
&= D^{-1/2}DD^{-1/2} \\
&= I_r.
\end{aligned}$$

So the columns of U_1 are the first r vectors of an orthonormal basis for $\mathbb{C}^m$. It's a theorem from linear algebra that we can find $m - r$ more vectors to complete these columns to an orthonormal basis. We'll put those $m - r$ vectors together to form an $m \times (m - r)$ matrix U_2 and then let $U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$. It will turn out not to matter what U_2 is for the sake of SVD; we are just adding it in so that we can replace U_1 by an orthogonal matrix.

Take a look back at V from above: this was where we got V_1 from. We had written $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix}$ and we'll use this V for our other matrix in the SVD. It is also unitary.

For Σ we'll use $D^{1/2}$, and then we will add on rows and/or columns of 0's to make Σ have the same size as M. So Σ will look like

$$\Sigma = \begin{pmatrix} D^{1/2} & 0 \\ 0 & 0 \end{pmatrix}. \tag{8}$$

where those blocks of zeros are of appropriate size to give the matrix dimensions

$m \times n$. Keep in mind that $D^{1/2}$ is an $r \times r$ square matrix. Then we have

$$\begin{aligned} U\Sigma V^* &= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} D^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^* \\ V_2^* \end{pmatrix} \\ &= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} D^{1/2} V_1^* \end{pmatrix} \\ &= U_1 D^{1/2} V_1^* \\ &= M. \end{aligned}$$

$\square$

References

[Wikipedia] https://en.wikipedia.org/wiki/Singular_value_decomposition

[Axler] Axler, Linear Algebra Done Right