Lectures - Week 9
The Singular Value Decomposition Theorem

We know that eigenvalues are only defined for a square matrix. However, in this section we want to define an analogue of eigenvalues for a rectangular matrix. This will lead us to our final decomposition, the Singular Vale Decomposition (SVD), of a matrix; recall that we have seen two other decompositions: LU (and its variants) and QR. The SVD yields a diagonal (but not square if A is rectangular) matrix. Recall that if an $n \times n$ square matrix has n linearly independent eigenvectors then there is a matrix P whose columns are these eigenvectors such that $P^{-1}AP = \Lambda$ where Λ is a diagonal matrix. If the eigenvectors of A are orthogonal (i.e., meet at right angles) then we can normalize these eigenvectors and choose orthonormal eigenvectors to obtain $Q^T AQ = \Lambda$. For example, if A is symmetric then it is orthogonally similar to a diagonal matrix. If a square matrix does NOT have orthogonal eigenvectors then we need two different orthogonal matrices to diagonalize A. The SVD provides this decomposition for a matrix which does not have orthogonal eigenvectors and it holds for rectangular matrices as well. You will discover in ACS I that the SVD is extremely useful in many applications.

Definition The *singular values* of an $m \times n$ matrix A are the square roots of the eigenvalues of the $n \times n$ matrix $A^T A$.

We first note that $A^T A$ is symmetric and at least positive semi-definite so its eigenvalues are real and non-negative so it makes sense to talk about their square root. If A is square and symmetric then the singular values are related to the eigenvalues of A. When A is symmetric the eigenvalues of $A^T A$ are the eigenvalues of A^2 which are the squares of the eigenvalues of A so when we take the square root we get the magnitude of the eigenvalues of A.

Our third decomposition, the singular value decomposition (SVD) is given in the next theorem. We will see that it gives us information about the four fundamental spaces associated with a matrix plus additional information on the relative importance of the columns of A.

Theorem The Singular Value Decomposition Theorem (SVD). Let A be an $m \times n$ matrix. Then A can be factored as

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

where

U is an $m \times m$ orthogonal matrix
Σ is an $m \times n$ diagonal matrix ($\Sigma_{ij} = 0$ for $i \neq j$) with entries $\sigma_i \geq 0$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$, $k = \min\{m, n\}$.
V is an $n \times n$ orthogonal matrix

We first remark that Σ is square if A is, but rectangular in general. The following matrices illustrate some possible forms for Σ; the first is when A is 3 x 3 with rank 3, the second is when A is 3 x 3 with rank 2, the third is when A is 3 x 2 with rank 2, the next is when A is 3 x 2 with rank 1 and the last when A is 2 x 3 with rank 2.

$$\begin{pmatrix} 10 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 10 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 10 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 10 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 10 & 0 & 0 \\ 0 & 4 & 0 \end{pmatrix}$$

We note that the SVD implies

$$A = U\Sigma V^T \Rightarrow U^T A = \Sigma V^T \Rightarrow U^T AV = \Sigma$$

so we have diagonalized A using two different orthogonal matrices. However, this is not a similarity transformation so the eigenvalues are not preserved. The diagonal entries of Σ are the singular values of A rather than its eigenvalues.

We will not prove this result but rather investigate what U, Σ and V tell us about A. To determine the importance of U, V and Σ we first consider an expression for each of the two matrices $A^T A$ and AA^T using the SVD of A.

$$A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T \, .$$

Now U is orthogonal and thus $U^T U = I_{m\times m}$ implies

$$A^T A = V\Sigma^T U^T U\Sigma V^T = V(\Sigma^T\Sigma)V^T \, .$$

This says that $A^T A$ is orthogonally similar to $\Sigma^T\Sigma$; here $\Sigma^T\Sigma$ is an $n \times n$ diagonal matrix with entries σ_i^2, 0.

Now consider the matrix AA^T

$$AA^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U(\Sigma\Sigma^T)U^T$$

which says that AA^T is orthogonally similar to $\Sigma\Sigma^T$; here $\Sigma\Sigma^T$ is an $m \times m$ diagonal matrix with entries $\sigma_i^2, 0$.

We now want to use these results to interpret the meaning of each matrix in the SVD.

(1.) *The diagonal entries of Σ, denoted σ_i, are the singular values of A.*

From our expression for the $n \times n$ matrix $A^T A$ (or equivalently for AA^T) we have seen that $A^T A = V(\Sigma^T\Sigma)V^T$ so that the eigenvalues of $A^T A$ and $\Sigma^T\Sigma$ are the same because they are similar.. The eigenvalues of $\Sigma^T\Sigma$ are σ_i^2 and so the eigenvalues of $A^T A$ are too. Thus the singular values of A are $\sqrt{\sigma_i^2} = \sigma_i$

(2.) *The columns of V are the orthonormal eigenvectors of $A^T A$.*

Clearly $A^T A$ is symmetric and thus has a complete set of orthonormal eigenvectors. Also $A^T A = V(\Sigma^T\Sigma)V^T$ which implies $V^T(A^T A)V = \Sigma^T\Sigma$ so V is the matrix which diagonalizes $A^T A$ and thus the columns of V are the orthonormal eigenvectors.

(3.) *The columns of U are the orthonormal eigenvectors of AA^T.*

Clearly AA^T is symmetric and thus has a complete set of orthonormal eigenvectors. From above $AA^T = U(\Sigma\Sigma^T)U^T$ which implies $U^T(AA^T)U = \Sigma\Sigma^T$ so U is the matrix which diagonalizes AA^T and each of the columns of U is an orthonormal eigenvector of AA^T.

(4.) *The SVD provides an orthonormal basis for the null space of A, $\mathcal{N}(A)$, and its orthogonal complement, $\mathcal{R}(A^T)$ (which is the row space of A).*

We know that if $\vec{x} \in \mathcal{N}(A)$ then $A\vec{x} = \vec{0}$. Now partition the $n \times n$ matrix V as $V = (V_1|V_2)$ where V_1 is $n \times p$ and V_2 is $n \times (n-p)$ where p is the index of the last nonzero singular value σ_i; i.e., $\sigma_{p+1} = \sigma_{p+2} = \cdots = \sigma_k = 0$ where $k = \min\{m, n\}$. Our claim is that the columns of V_2 form a basis for $\mathcal{N}(A)$ and the columns of V_1 form a basis for $\mathcal{R}(A^T)$. To see this note that $A = U\Sigma V^T$ implies $AV = U\Sigma$. Because the last $(p+1)$ through n columns of V correspond to the diagonal entries of Σ which are zero, then $AV_2 = 0$ and the columns of V_2 are in $\mathcal{N}(A)$ and are orthonormal because V is orthogonal. Now the first p columns of V denoted by V_1 are orthonormal to the columns of V_2 and form a basis for $\mathcal{R}(A^T)$ because $\mathcal{R}(A^T)$ is the orthogonal complement of $\mathcal{N}(A)$.

(5.) *The SVD provides an orthonormal basis for the range of A, $\mathcal{R}(A)$, and its orthogonal complement $\mathcal{N}(A^T)$.*

Now partition the $m \times m$ orthogonal matrix U as $U = (U_1|U_2)$ where U_1 is $m \times p$. As above, the SVD implies

$$AV = U\Sigma \Rightarrow A(V_1|V_2) = \Sigma(U_1|U_2)$$

The first p columns of U denoted by U_1 form a basis for the range of A and the remaining columns denoted by U_2 are orthogonal and form a basis for the orthogonal complement $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.

(6.) *The rank of A is given by the number of nonzero singular values in Σ.*

If we multiply a matrix by an orthogonal matrix it does not change its rank.

In the sequel we will assume that the first p diagonal entries of Σ, σ_i, $i = 1, \ldots, p$ are > 0.

Example Find the SVD for the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 7 & 7 \end{pmatrix}$$

Clearly A is a square matrix with rank one. We first form A^TA and find its eigenvectors to get V.

$$A^TA = \begin{pmatrix} 50 & 50 \\ 50 & 50 \end{pmatrix}$$

which has a characteristic equation $(50-\lambda)^2 - 50^2 = 0$ which implies $\lambda = 0$ and $\lambda = 100$. The orthonormal eigenvectors corresponding to these can be determined as

$$[100, \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}], \quad [0, \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}]$$

so we take

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} .$$

The matrix Σ is a 2×2 diagonal matrix with entries given as the singular values of A which are the square root of the eigenvalues of $A^T A$. Thus

$$\Sigma = \begin{pmatrix} 10 & 0 \\ 0 & 0 \end{pmatrix} .$$

Now we form AA^T and find its eigenvectors to get U.

$$AA^T = \begin{pmatrix} 2 & 14 \\ 14 & 98 \end{pmatrix}$$

whose eigenpairs are given by

$$[100, \frac{1}{\sqrt{50}} \begin{pmatrix} 1 \\ 7 \end{pmatrix}], \quad [0, \frac{1}{\sqrt{50}} \begin{pmatrix} -7 \\ 1 \end{pmatrix}]$$

Thus

$$U = \frac{1}{\sqrt{50}} \begin{pmatrix} 1 & -7 \\ 7 & 1 \end{pmatrix}$$

As a check we form $U\Sigma V^T$ and see if we get A

$$\frac{1}{\sqrt{50}} \begin{pmatrix} 1 & -7 \\ 7 & 1 \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^T = \frac{1}{\sqrt{100}} \begin{pmatrix} 1 & -7 \\ 7 & 1 \end{pmatrix} \begin{pmatrix} 10 & 10 \\ 0 & 0 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 10 & 10 \\ 70 & 70 \end{pmatrix} = A$$

Example For the matrix in the previous example, find a basis for the four fundamental spaces in the usual way and then compare with those found by using the SVD.

Clearly A has rank one and the dimension of the null space is one. Clearly $(1,-1)^T \in \mathcal{N}(A)$. The dimension of the $\mathcal{R}(A)$ is one and a basis can be taken as the first column of A $(1,7)^T$. The dimension of $\mathcal{R}(A^T)$ which is the row space of A is one and a basis is just $(1,1)^T$. Lastly the left null space of A, $\mathcal{N}(A^T)$ has dimension one and a basis is just $(-7,1)^T$. Comparing with the SVD we see that there is one nonzero singular value so the rank of A is one. The last column of V $(1/\sqrt{2})(-1,1)^T$ is a basis for the null space of A and is just $-1/\sqrt{2}$ times our basis vector. The first column of V will give the orthogonal complement of $\mathcal{N}(A)$ which is $\mathcal{R}(A^T)$ and is just $1/\sqrt{2}$ times our vector. The first column of U is $(1/\sqrt{50}(1,7)$ which is just $1/\sqrt{50}$ times our basis above. The basis for the orthogonal complement of the range of A, $\mathcal{N}(A^T)$ is just the second column of U which is a constant times our vector.

Example Let

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix} = U\Sigma V^T$$

Find its rank, a basis for the range and the null space in the usual manner as well as form A^TA and find its eigenvalues and eigenvectors. Then use the SVD of A to calculate each of these and compare.

Clearly A has rank two because it has two linearly independent columns. The range is all of R^2. Its null space has dimension 3-2=1 and a basis for the null space is given by $(-3,-6,15)^T$ because

$$\begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3 & 1 & 1 \\ 0 & 10/3 & 4/3 \end{pmatrix} \Rightarrow z_3 = 3, z_2 = -6/5, z_1 = -3/5$$

Note if we multiply the last column of V, i.e., the last row of V^T by $-3\sqrt{30}$ then we get this vector so the last column of V^T is in the null space of A. Because the range is all of R^2 we can use the standard basis. Also note that the first two columns of U form an orthonormal basis for R^2 and thus for the range of A.
Forming A^TA gives

$$\begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

By looking at the characteristic equation for A^TA we find that its eigenpairs are

$$[12, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}], \quad [10, \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}], \quad [0, \frac{1}{\sqrt{30}} \begin{pmatrix} -1 \\ -2 \\ 5 \end{pmatrix}]$$

Clearly these orthonormal eigenvectors are the columns of V and the singular values of A are the square roots of the eigenvalues $12, 10$, i.e., the diagonal entries of Σ.

Recall that our decomposition $U\Sigma V^T$ gives us $AV = U\Sigma$ so if $\vec{v}_i$ are the columns of V and $\vec{u}_i$ the columns of U, then $A\vec{v}_i = \sigma_i \vec{u}_i$. The columns $\vec{v}_i$ are eigenvectors of A^TA and we call them the right singular vectors of A. Also if we take the transpose of the SVD of A we have $A^T = V\Sigma^T U^T$ which implies $A^T U = V\Sigma^T$ and therefore $A^T\vec{u}_i = \sigma_i \vec{v}_i$. The columns of U are eigenvectors of AA^T and we call them the left singular vectors of A.

The 2-condition number can be extended to rectangular matrices. Recall that for a square matrix

$$\|A\|_2 = \sqrt{\rho(A^TA)}$$

which is just the largest singular value of A. Let A be an $m \times n$ matrix and let the singular values of A be denoted $\sigma_1 \geq \sigma_2 \geq \sigma_k > 0$ for $k = \min\{m, n\}$. Call $\sigma_{\max} = \sigma_1$ and $\sigma_{\min} = \sigma_k > 0$. Then

$$\mathcal{K}_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}} \geq 1$$

i.e., the ratio of the largest singular value to the smallest. If $\sigma_{\min} = 0$ then clearly this is not defined; in this case A is not full rank and we say the condition number is infinite (just as in the case of a singular matrix we have a zero eigenvalue).

Lastly, the SVD provides us with information about the relative importance of the columns of A. Suppose that we have n vectors in R^m and we form an $m \times n$ matrix A whose rank is m. Suppose further that this vectors contain redundant information and we want to represent the information contained in the matrix by a matrix of lower rank. The SVD allows us to do this. We first note that from the SVD for A we have

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \cdots + \sigma_k \vec{u}_k \vec{v}_k^T$$

where $\vec{u}_i$ represents the ith column of U and $\vec{v}_i^T$ the ith column of V^T and k is the last nonzero singular value of A. Now suppose we want to approximate our matrix A by a matrix of rank $\ell < m$; call it A_ℓ. Then we take

$$A_\ell = \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^T \, .$$

Example Consider the following rank 3 matrix and its SVD.

$$A = \begin{pmatrix} 2. & 1. & 1. \\ 10. & 3. & 4. \\ 8. & 1. & 4. \\ 6. & 0. & 8. \\ 4. & 6. & 8. \end{pmatrix}$$

where the $U \Sigma V^T$ is

$$\begin{pmatrix} -0.122 & 0.045 & 0.141 & 0.268 & -0.944 \\ -0.552 & 0.468 & 0.415 & 0.469 & 0.289 \\ -0.448 & 0.400 & -0.057 & -0.783 & -0.154 \\ -0.486 & -0.125 & -0.821 & 0.272 & 0.012 \\ -0.493 & -0.777 & 0.361 & -0.149 & 0.038 \end{pmatrix} \begin{pmatrix} 19.303 & 0.000 & 0.000 \\ 0.000 & 6.204 & 0.000 \\ 0.000 & 0.000 & 4.111 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{pmatrix} \begin{pmatrix} -0.738 & 0.664 & 0.121 \\ -0.269 & -0.453 & 0.850 \\ -0.619 & -0.595 & -0.512 \end{pmatrix}$$

Determine a rank one approximation to A

From above we have

$$A_1 = 19.3 \begin{pmatrix} -.122 \\ -.552 \\ -.448 \\ -.486 \\ -.493 \end{pmatrix} \left(-.7238 \quad -.269 \quad -.619 \right) = \begin{pmatrix} 1.743 & 0.635 & 1.464 \\ 7.864 & 2.864 & 6.603 \\ 6.379 & 2.323 & 5.356 \\ 6.920 & 2.520 & 5.811 \\ 7.021 & 2.557 & 5.895 \end{pmatrix}$$

A rank two approximation can be found to be

$$\begin{pmatrix} 1.930 & 0.508 & 1.297 \\ 9.794 & 1.548 & 4.875 \\ 8.028 & 1.199 & 3.880 \\ 6.407 & 2.870 & 6.270 \\ 3.821 & 4.738 & 8.760 \end{pmatrix}$$

The rank 3 approximation is A itself because A is rank three.