Notes on singular value decomposition for Math 54

Recall that if A is a symmetric $n \times n$ matrix, then A has real eigenvalues $\lambda_1, \ldots, \lambda_n$ (possibly repeated), and $\mathbb{R}^n$ has an orthonormal basis $v_1, \ldots, v_n$, where each vector v_i is an eigenvector of A with eigenvalue λ_i. Then

$$A = PDP^{-1}$$

where P is the matrix whose columns are $v_1, \ldots, v_n$, and D is the diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_n$. Since the vectors $v_1, \ldots, v_n$ are orthonormal, the matrix P is orthogonal, i.e. $P^T P = I$, so we can alternately write the above equation as

$$A = PDP^T. \tag{1}$$

A singular value decomposition (SVD) is a generalization of this where A is an $m \times n$ matrix which does not have to be symmetric or even square.

1 Singular values

Let A be an $m \times n$ matrix. Before explaining what a singular value decomposition is, we first need to define the singular values of A.

Consider the matrix $A^T A$. This is a symmetric $n \times n$ matrix, so its eigenvalues are real.

Lemma 1.1. *If λ is an eigenvalue of $A^T A$, then $\lambda \geq 0$.*

Proof. Let x be an eigenvector of $A^T A$ with eigenvalue λ. We compute that

$$\|Ax\|^2 = (Ax) \cdot (Ax) = (Ax)^T Ax = x^T A^T Ax = x^T (\lambda x) = \lambda x^T x = \lambda \|x\|^2.$$

Since $\|Ax\|^2 \geq 0$, it follows from the above equation that $\lambda \|x\|^2 \geq 0$. Since $\|x\|^2 > 0$ (as our convention is that eigenvectors are nonzero), we deduce that $\lambda \geq 0$. $\square$

Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of $A^T A$, with repetitions. Order these so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Let $\sigma_i = \sqrt{\lambda_i}$, so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Definition 1.2. The numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ defined above are called the **singular values** of A.

Proposition 1.3. *The number of nonzero singular values of A equals the rank of A.*

Proof. The rank of any square matrix equals the number of nonzero eigenvalues (with repetitions), so the number of nonzero singular values of A equals the rank of $A^T A$. By a previous homework problem, $A^T A$ and A have the same kernel. It then follows from the "rank-nullity" theorem that $A^T A$ and A have the same rank. □

Remark 1.4. In particular, if A is an $m \times n$ matrix with $m < n$, then A has at most m nonzero singular values, because $\text{rank}(A) \leq m$.

The singular values of A have the following geometric significance.

Proposition 1.5. *Let A be an $m \times n$ matrix. Then the maximum value of $\|Ax\|$, where x ranges over unit vectors in $\mathbb{R}^n$, is the largest singular value σ_1, and this is achieved when x is an eigenvector of $A^T A$ with eigenvalue σ_1^2.*

Proof. Let $v_1, \ldots, v_n$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$ with eigenvalues σ_i^2. If $x \in \mathbb{R}^n$, then we can expand x in this basis as

$$x = c_1 v_1 + \cdots + c_n v_n \tag{2}$$

for scalars $c_1, \ldots, c_n$. Since x is a unit vector, $\|x\|^2 = 1$, which (since the vectors $v_1, \ldots, v_n$ are orthonormal) means that

$$c_1^2 + \cdots + c_n^2 = 1.$$

On the other hand,

$$\|Ax\|^2 = (Ax) \cdot (Ax) = (Ax)^T (Ax) = x^T A^T A x = x \cdot (A^T A x).$$

By (2), since v_i is an eigenvalue of $A^T A$ with eigenvalue σ_i^2, we have

$$A^T A x = c_1 \sigma_1^2 v_1 + \cdots + c_n \sigma_n^2 v_n.$$

Taking the dot prodoct with (2), and using the fact that the vectors $v_1, \ldots, v_n$ are orthonormal, we get

$$\|Ax\|^2 = x \cdot (A^T A x) = \sigma_1^2 c_1^2 + \cdots + \sigma_n^2 c_n^2.$$

Since σ_1 is the largest singular value, we get

$$\|Ax\|^2 \leq \sigma_1^2 (c_1^2 + \cdots + c_n^2).$$

Equality holds when $c_1 = 1$ and $c_2 = \cdots = c_n = 0$. Thus the maximum value of $\|Ax\|^2$ for a unit vector x is σ_1^2, which is achieved when $x = v_1$. □

One can similarly show that σ_2 is the maximum of $\|Ax\|$ where x ranges over unit vectors that are orthogonal to v_1 (exercise). Likewise, σ_3 is the maximum of $\|Ax\|$ where x ranges over unit vectors that are orthogonal to v_1 and v_2; and so forth.

2 Definition of singular value decomposition

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Let r denote the number of nonzero singular values of A, or equivalently the rank of A.

Definition 2.1. A **singular value decomposition** of A is a factorization

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

where:

- U is an $m \times m$ orthogonal matrix.
- V is an $n \times n$ orthogonal matrix.
- Σ is an $m \times n$ matrix whose i^{th} diagonal entry equals the i^{th} singular value σ_i for $i = 1, \ldots, r$. All other entries of Σ are zero.

Example 2.2. If $m = n$ and A is symmetric, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, ordered so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. The singular values of A are given by $\sigma_i = |\lambda_i|$ (exercise). Let $v_1, \ldots, v_n$ be orthonormal eigenvectors of A with $Av_i = \lambda_i v_i$. We can then take V to be the matrix whose columns are $v_1, \ldots, v_n$. (This is the matrix P in equation (1).) The matrix Σ is the diagonal matrix with diagonal entries $|\lambda_1|, \ldots, |\lambda_n|$. (This is almost the same as the matrix D in equation (1), except for the absolute value signs.) Then U must be the matrix whose columns are $\pm v_1, \ldots, \pm v_n$, where the sign next to v_i is $+$ when $\lambda_i \geq 0$, and $-$ when $\lambda_i < 0$. (This is almost the same as P, except we have changed the signs of some of the columns.)

3 How to find a SVD

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and let r denote the number of nonzero singular values. We now explain how to find a SVD of A.

Let $v_1, \ldots, v_n$ be an orthonormal basis of $\mathbb{R}^n$, where v_i is an eigenvector of A^TA with eigenvalue σ_i^2.

Lemma 3.1. *(a)* $\|Av_i\| = \sigma_i$.

(b) If $i \neq j$ *then* Av_i *and* Av_j *are orthogonal.*

Proof. We compute

$$(Av_i) \cdot (Av_j) = (Av_i)^T (Av_j) = v_i^T A^T A v_j = v_i^T \sigma_j^2 v_j = \sigma_j^2 (v_i \cdot v_j).$$

If $i = j$, then since $\|v_i\| = 1$, this calculation tells us that $\|Av_i\|^2 = \sigma_j^2$, which proves (a). If $i \neq j$, then since $v_i \cdot v_j = 0$, this calculation shows that $(Av_i) \cdot (Av_j) = 0$. □

Theorem 3.2. *Let A be an $m \times n$ matrix. Then A has a (not unique) singular value decomposition $A = U\Sigma V^T$, where U and V are as follows:*

- *The columns of V are orthonormal eigenvectors $v_1, \ldots, v_n$ of $A^T A$, where $A^T A v_i = \sigma_i^2 v_i$.*
- *If $i \leq r$, so that $\sigma_i \neq 0$, then the i^{th} column of U is $\sigma_i^{-1} A v_i$. By Lemma 3.1, these columns are orthonormal, and the remaining columns of U are obtained by arbitrarily extending to an orthonormal basis for $\mathbb{R}^m$.*

Proof. We just have to check that if U and V are defined as above, then $A = U\Sigma V^T$. If $x \in \mathbb{R}^n$, then the components of $V^T x$ are the dot products of the rows of V^T with x, so

$$V^T x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_n \cdot x \end{pmatrix}.$$

Then

$$\Sigma V^T x = \begin{pmatrix} \sigma_1 v_1 \cdot x \\ \sigma_2 v_2 \cdot x \\ \vdots \\ \sigma_r v_r \cdot x \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

When we multiply on the left by U, we get the sum of the columns of U, weighted by the components of the above vector, so that

$$\begin{aligned} U\Sigma V^T x &= (\sigma_1 v_1 \cdot x)\sigma_1^{-1} A v_1 + \cdots + (\sigma_r v_r \cdot x)\sigma_r^{-1} A v_r \\ &= (v_1 \cdot x) A v_1 + \cdots + (v_r \cdot x) A v_r. \end{aligned}$$

Since $Av_i = 0$ for $i > r$ by Lemma 3.1(a), we can rewrite the above as

$$\begin{aligned} U\Sigma V^T x &= (v_1 \cdot x)Av_1 + \cdots + (v_n \cdot x)Av_n \\ &= Av_1 v_1^T x + \cdots + Av_n v_n^T x \\ &= A(v_1 v_1^T + \cdots v_n v_n^T)x \\ &= Ax. \end{aligned}$$

In the last line, we have used the fact that if $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$, then $v_1 v_1^T + \cdots + v_n v_n^T = I$ (exercise). $\square$

Example 3.3. (from Lay's book) *Find a singular value decomposition of*

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}.$$

Step 1. We first need to find the eigenvalues of $A^T A$. We compute that

$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

We know that at least one of the eigenvalues is 0, because this matrix can have rank at most 2. In fact, we can compute that the eigenvalues are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$. Thus the singular values of A are $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$, and $\sigma_3 = 0$. The matrix Σ in a singular value decomposition of A has to be a 2×3 matrix, so it must be

$$\Sigma = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}.$$

Step 2. To find a matrix V that we can use, we need to solve for an orthonormal basis of eigenvectors of $A^T A$. One possibility is

$$v_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}.$$

(There are seven other possibilities in which some of the above vectors are multiplied by -1.) Then V is the matrix with v_1, v_2, v_3 as columns, that is

$$V = \begin{pmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}.$$

Step 3. We now find the matrix U. The first column of U is

$$\sigma_1^{-1}Av_1 = \frac{1}{6\sqrt{10}}\begin{pmatrix}18\\6\end{pmatrix} = \begin{pmatrix}3/\sqrt{10}\\1/\sqrt{10}\end{pmatrix}.$$

The second column of U is

$$\sigma_2^{-1}Av_2 = \frac{1}{3\sqrt{10}}\begin{pmatrix}3\\9\end{pmatrix} = \begin{pmatrix}1/\sqrt{10}\\-3/\sqrt{10}\end{pmatrix}.$$

Since U is a 2×2 matrix, we do not need any more columns. (If A had only one nonzero singular value, then we would need to add another column to U to make it an orthogonal matrix.) Thus

$$U = \begin{pmatrix}3/\sqrt{10} & 1/\sqrt{10}\\1/\sqrt{10} & -3/\sqrt{10}\end{pmatrix}.$$

To conclude, we have found the singular value decomposition

$$\begin{pmatrix}4 & 11 & 14\\8 & 7 & -2\end{pmatrix} = \begin{pmatrix}3/\sqrt{10} & 1/\sqrt{10}\\1/\sqrt{10} & -3/\sqrt{10}\end{pmatrix}\begin{pmatrix}6\sqrt{10} & 0 & 0\\0 & 3\sqrt{10} & 0\end{pmatrix}\begin{pmatrix}1/3 & -2/3 & 2/3\\2/3 & -1/3 & -2/3\\2/3 & 2/3 & 1/3\end{pmatrix}^T.$$

4 Applications

Singular values and singular value decompositions are important in analyzing data.

One simple example of this is "rank estimation". Suppose that we have n data points $v_1,\ldots,v_n$, all of which live in $\mathbb{R}^m$, where n is much larger than m. Let A be the $m\times n$ matrix with columns $v_1,\ldots,v_n$. Suppose the data points satisfy some linear relations, so that $v_1,\ldots,v_n$ all lie in an r-dimensional subspace of $\mathbb{R}^m$. Then we would expect the matrix A to have rank r. However if the data points are obtained from measurements with errors, then the matrix A will probably have full rank m. But only r of the singular values of A will be large, and the other singular values will be close to zero. Thus one can compute an "approximate rank" of A by counting the number of singular values which are much larger than the others, and one expects the measured matrix A to be close to a matrix A' such that the rank of A' is the "approximate rank" of A.

For example, consider the matrix

$$A' = \begin{pmatrix}1 & 2 & -2 & 3\\-4 & 0 & 1 & 2\\3 & -2 & 1 & -5\end{pmatrix}$$

The matrix A' has rank 2, because all of its columns are points in the subspace $x_1 + x_2 + x_3 = 0$ (but the columns do not all lie in a 1-dimensional subspace). Now suppose we perturb A' to the matrix

$$A = \begin{pmatrix} 1.01 & 2.01 & -2 & 2.99 \\ -4.01 & 0.01 & 1.01 & 2.02 \\ 3.01 & -1.99 & 1 & -4.98 \end{pmatrix}$$

This matrix now has rank 3. But the eigenvalues of $A^T A$ are

$$\sigma_1^2 \approx 58.604, \quad \sigma_2^2 \approx 19.3973, \quad \sigma_3^2 \approx 0.00029, \quad \sigma_4^2 = 0.$$

Since two of the singular values are much larger than the others, this suggests that A is close to a rank 2 matrix.

For more discussion of how SVD is used to analyze data, see e.g. Lay's book.

5 Exercises (some from Lay's book)

1. (a) Find a singular value decomposition of the matrix $A = \begin{pmatrix} 2 & -1 \\ 2 & 2 \end{pmatrix}$.

   (b) Find a unit vector x for which $\|Ax\|$ is maximized.

2. Find a singular value decomposition of $A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$.

3. (a) Show that if A is an $n \times n$ symmetric matrix, then the singular values of A are the absolute values of the eigenvalues of A.

   (b) Give an example to show that if A is a 2×2 matrix which is not symmetric, then the singular values of A might not equal the absolute values of the eigenvalues of A.

4. Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Let v_1 be an eigenvector of $A^T A$ with eigenvalue σ_1^2. Show that σ_2 is the maximum value of $\|Ax\|$ where x ranges over unit vectors in $\mathbb{R}^n$ that are orthogonal to v_1.

5. Show that if $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$, then

$$v_1 v_1^T + \cdots + v_n v_n^T = I.$$

6. Let A be an $m \times n$ matrix, and let P be an orthogonal $m \times m$ matrix. Show that PA has the same singular values as A.