Lecture 4.21. Singular value decomposition

A. Eremenko

April 16, 2020

The problem addressed here is how can one simplify a linear transformation by choosing two *different bases*, one in the domain and one in the image. Every linear transformation from R^n to R^m can be represented by an $m \times n$ matrix in the standard basis. So we just consider the transformation

$$x \mapsto Ax$$

for some arbitrary matrix A.

Theorem. *For every A, there exists an orthonormal basis $v_1, \ldots, v_n$ and an orthonormal basis $u_1, \ldots, u_m$ such that*

$$Av_j = \sigma_j u_j, \quad 1 \leq j \leq n, \tag{1}$$

where $\sigma_j \geq 0$.

Remark. In the case that $m < n$ this should be understood as $\sigma_j = 0$ for $j > m$.

Proof. Consider the matrix $A^T A$ (it is $n \times n$). This matrix is symmetric and positive semidefinite. Indeed

$$(A^T A)^T = A^T A,$$

and

$$x^T A^T Ax = (Ax)^T (Ax) \geq 0,$$

by the positivity of the dot product.

By the Spectral Theorem for symmetric matrices, there is an orthonormal basis $v_1, \ldots, v_n$ made of eigenvectors of $A^T A$. We take it as a basis in the domain of A, and we order this basis so that eigenvalues are listed in non-increasing order. Since $A^T A$ is positive semidefinite, eigenvalues are non-negative.

Now we prove that vectors Av_i are orthogonal:

$$(Av_i)^T Av_j = v_i^T A^T Av_j = \lambda_j v_i^T v_j.$$

This is zero when $i \neq j$. To convert Av_j into an *orthonormal* system we have to divide these vectors by square roots of λ_j. Recall that λ_j are non-negative, and denote

$$\sigma_j = \sqrt{\lambda_j} \geq 0.$$

Then we set $u_j = Av_j/\sigma_j$ when $\sigma_j \neq 0$, and obtain (1). If m is greater than the number of non-zero σ_j, complete $u_1, \ldots, u_n$ to an orthonormal basis. So we proved the theorem.

Now let us state it in terms of matrix factorization. Let $V = [v_1, \ldots, v_n]$ be the matrix whose columns are v_j. Let $U = [u_1, \ldots, u_m]$ be the matrix with columns u_j. These matrices are orthogonal:

$$V^T = V^{-1}, \quad U^T = U^{-1}.$$

Multiplying V by A from the left, we obtain using (1):

$$AV = A[v_1, \ldots, v_n] = [\sigma_1 u_1, \ldots, \sigma_n u_n] = U\Sigma,$$

where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$. In other words,

$$A = U\Sigma V^T. \tag{2}$$

The numbers σ_j are called the *singular values* of the matrix A, and the formula (2) is called the *singular value decomposition,* abbreviated as SVD. In the case that $m > n$, we have to extend Σ by adding zeros in the bottom so that it becomes an $m \times m$ matrix, and so that (2) makes sense.

The geometric meaning of SVD is the following. If we have any linear transformation $\mathbf{R}^n \to \mathbf{R}^m$, then we can make orthogonal changes of coordinates in $\mathbf{R}^n$ and $\mathbf{R}^m$ (two *different* new coordinate systems!) so that in these new coordinates the matrix of our transformation is diagonal with non-negative entries. In other words, in appropriate coordinate systems, any linear transformation is just stretching by different amounts along each coordinate.

This is in great contrast with the Jordan normal form of an operator: for operators (acting from a space to the *same* space, only the *same* coordinate change in the domain and in the image was allowed. While for linear transformations separate different changes are allowed. For this reason SD is simpler that the Jordan form.

To find the SVD for a given matrix, just find eigenvalues and eigenvectors of A^TA. Order the eigenvalues and eigenvectors so that eigenvalues decrease. Put eigenvectors v_j as columns of V, and vectors Av_j/σ_j as columns of U, where σ_j are positive square roots of positive λ_j. If $m > n$ add some columns to U using Gram–Schmidt process. The diagonal entries of Σ are positive square roots of eigenvalues of A^TA. Don't forget to add $m-n$ zeros rows if $m > n$, so that Σ has the correct size.

Remark. The columns of U are eigenvectors of AA^T: indeed, multiplying (1) from the left on AA^T we obtain

$$AA^TAv_j = \sigma_j AA^T u_j.$$

As $A^TAv_j = \lambda_j v_j$, we have

$$A\lambda_j v_j = \sigma AA^T u_j,$$

and using (1) again

$$\lambda_j \sigma_j u_j = \sigma_j AA^T u_j.$$

Dividing on σ_j we conclude that u_j are eigenvectors of AA^T with eigenvalues λ_j.

So A^TA *and* AA^T *always have the same eigenvalues with the same multiplicities, except for the zero eigenvalue.*

You can use this fact to save efforts when finding the SVD for a rectangular matrix: compute the eigenvalues of the *smaller* of the two matrices AA^T and A^TA.

Polar decompositions. We give several applications.

1. Every real square matrix A can be written as a product

$$A = SO,$$

where S is symmetric, positive semidefinite, and O is orthogonal. Indeed, we have

$$A = U\Sigma V^T = (U\Sigma U^{-1})(UV^T),$$

so we can define $S = U\Sigma U^{-1} = U\Sigma U^T$ which is symmetric and positive semidefinite, and $O = UV^T$ which is orthogonal.

Similarly we can write every real square matrix as

$$A = U\Sigma V^T = (UV^{-1})(V\Sigma V^T),$$

where UV^{-1} is orthogonal and $V\Sigma V^T$ is symmetric and positive semidefinite.

These two representations of arbitrary real matrix generalize the polar representation of a complex number.

2. Same arguments work for complex matrices, using Hermitian transpose instead of the usual one. We obtain that every complex matrix can we written in the form

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

where U and V are unitary and Σ diagonal, with non-negative entries. In the polar decompositions the matrix S will be symmetric positive definite and O will be unitary.

Example 1.

$$A = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

We compute AA^T, because it is of smaller size:

$$AA^T = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

The characteristic polynomial is

$$\lambda^2 - 4\lambda + 3,$$

and eigenvalues are $\lambda_1 = 3, \geq \lambda_2 = 1$. It is important to order them, so that λ_1 is bigger. Then

$$A^TA = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

It must have the same eigenvalues as AA^T, except zero eigenvalue. We find eigenvectors of A^TA:

For $\lambda_1 = 3$, we obtain $\mathbf{v}_1 = (1, -2, 1)^T$.

For $\lambda_2 = 1$, we obtain $\mathbf{v}_2 = (-1, 0, 1)^T$.

Since AA^T is 3×3, there is the third eigenvalue, and it must be 0.

For $\lambda_3 = 0$ we find $\mathbf{v}_3 = (1, 1, 1)^T$.

We also need eigenvectors of AA^T. They can be found in two ways: a) directly, since we know the eigenvalues, or b) as the images of $\mathbf{v}_j$ under A. Since it is easier to work with 2×2 matrices, rather than 3×3 we find them directly:

For $\lambda_1 = 3$, we obtain $\mathbf{u}_1 = (-1, 1)^T$.

For $\lambda_2 = 1$, we find $\mathbf{u}_2 = (1, 1)^T$.

All these eigenvectors $\mathbf{u}_i$ and $\mathbf{v}_j$ still have to be normalized. The final answer is

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$V^T = \begin{pmatrix} 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix},$$

the rows of V^T are the normalized vectors $\mathbf{v}_j$.

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

ad you may check that

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

Example 2.

$$A = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}.$$

The smaller matrix of AA^T and A^TA is

$$A^TA = (9).$$

Its only eigenvalue is 9, an eigenvector is $\mathbf{v}_1 = (1)$ and the only singular value is $\sigma_1 = \sqrt{9} = 3$. So

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1 = \begin{pmatrix} -1/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

We have $\|\mathbf{u}_1\| = 1$.

Since the matrix U must be 3×3 we need to complete the vector $\mathbf{u}_1$ to an orthonormal basis of $\mathbf{R}^3$, in other words, to find $\mathbf{u}_2, \mathbf{u}_3$ which are orthogonal to $\mathbf{u}_1$ and to each other, and have norm 1.

Orthogonality to $\mathbf{u}_1$ means that their coordinates must satisfy

$$-x_1 + 2x_2 + 2x_3 = 0.$$

This has linearly independent solutions $\mathbf{u}_1^* = (2, 0, 1)^T$ and $(2, 1, 0)^T$. They are not orthogonal, so we orthogonalize them, replacing the second vector by

$$\mathbf{u}_3^* = (2, 1, 0)^T - c(2.0.1)^T,$$

where c is found from the condition that $(\mathbf{u}_1^*, u_2^*) = 0$. We find $c = 4/5$, so

$$u_2^* = \begin{pmatrix} 2/5 \\ 1 \\ -4/5 \end{pmatrix}.$$

Now we normalize $\mathbf{u}_2^*$ and $\mathbf{u}_2^*$, and use the normalized vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ as columns of U. So

$$U = \begin{pmatrix} -1/3 & 2/5 & 2/(3\sqrt{5}) \\ 2/3 & 0 & \sqrt{5}/3 \\ 2/3 & 1/\sqrt{5} & -4/(3\sqrt{5}) \end{pmatrix}.$$

We also have

$$\Sigma = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \quad V^T = (1).$$

You can check that

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

Example 3. Let

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

Find the polar decomposition

$$A = SQ,$$

where S is symmetric, positive semidefinite, and Q is orthogonal.

Finding an SVD leads to quite complicated calculations. Here is a shortcut. We have $S = AQ^{-1}$, where Q is also orthogonal. The general form of a 2×2 orthogonal matrix (with determinant 1) is

$$Q^{-1} = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}, \quad p^2 + q^2 = 1,$$

(see, for example, Lecture 3.26, page 1, where this was proved).

So we have

$$S = \begin{pmatrix} 1 & -2 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} p & -q \\ q & p \end{pmatrix} = \begin{pmatrix} p - 2q & -q - 2p \\ 3p - q & -3q - p \end{pmatrix}.$$

Since this must be symmetric, we must have $-q - 2p = 3p - q$, so $p = 0$ and $q = \pm 1$. Taking $q = 1$ we obtain

$$Q^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = Q,$$

and the polar decomposition

$$\begin{pmatrix} 1 & -2 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

$q = -1$ does not lead to a solution (check!).

Application to signal processing. Suppose that we have a very large matrix, for example, a 1024×1024 matrix representing a picture on a computer screen. It contains more than 10^6 numbers, and the question is whether we can somehow compress this data when we wish to transmit this picture through some communication channel with limited capacity.

The idea is to approximate our big matrix by a matrix of small rank. Let A be our big matrix, and write its SVD decomposition

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

Remember that Σ is diagonal, with decreasing entries. We will obtain an approximation to A if we discard the small entries of Σ. Small in comparison with the largest entries. So we replace the last entries of Σ by zeros, and obtain a diagonal matrix with some k first non-zero elements. But then only first k rows of V^T matter: the rest are multiplied on zeros anyway. Similarly, only the first k columns of U matter, the rest are multiplied by 0. Therefore, we can transmit only the first k columns of U and V, and hopefully obtain an approximation of A.

This really works.