Singular value decomposition

A. Eremenko

April 12, 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

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

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. As $A^T A$ 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 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 They 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 singular values and the formula (2) is called the singular value decomposition. 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.

To find the SVD for a given matrix, just find eigenvalues and eigenvectors of $A^T A$. 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 $A^T A$. 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

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

As $A^T Av_j = \lambda_j v_j$, 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 $A^T A$ and AA^T always have the same eigenvalues with the same multiplicity, except the zero eigenvalue.

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

$$A = SO,$$

where S is symmetric positive semidefinite, and O is orthogonal. Undeed, 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 square real 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.

This generalizes he polar representation of a complex number.

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 symmetric and positive definite. In the polar decompositions the matrix S will be symmetric positive definite and O will be unitary.