

Proof of the Singular Value Decomposition

I walk the reader carefully through Gilbert Strang's existence proof of the singular value decomposition.

PUBLISHED

20 December 2018

The existence claim for the singular value decomposition (SVD) is quite strong: “Every matrix is diagonal, provided one uses the proper bases for the domain and range spaces” ([Trefethen & Bau III, 1997](#)). MIT professor Gilbert Strang has a [wonderful lecture](#) on the SVD, and he includes an existence proof for the SVD. The goal of this post is to force myself to walk through his proof carefully. For legibility, I break the proof into two sections: an overview and details. I hope this allows the reader to get the big picture of the proof while consulting details as needed.

Note that while the SVD holds for complex matrices, we restrict ourselves to real-valued matrices in this proof. Trefethen and Bau have a proof for the existence and uniqueness of the SVD for complex matrices, but I found Strang's proof more instructive.

If you're unfamiliar with the SVD, please see my previous post on a [geometrical interpretation of it](#).

Proof

Consider an $m \times n$ matrix A with rank r . The matrix $A^\top A$ is therefore symmetric and positive semi-definite (PSD) ([Details, Section 1](#)). This means the matrix is diagonalizable with an eigendecomposition of the form:

$$A^\top A = V\Lambda V^\top = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top = \sum_{i=1}^n (\sigma_i)^2 \mathbf{v}_i \mathbf{v}_i^\top$$

where V is an orthonormal matrix whose columns are the eigenvectors of $A^\top A$ and where $r \leq n$ and $r = \text{rank}(A) = \text{rank}(A^\top A)$ ([Details, Section 2](#)). The second equality above, which is a sum of matrices, holds [because \$\Lambda\$ is diagonal](#).

We have defined a quantity σ_i (the *singular values*) as the square root of the i -th eigenvalue; we know we can take the square root of our eigenvalues because PSD matrices can be equivalently characterized as matrices with non-negative eigenvalues ([Details, Section 3](#)).

For the i -th eigenvector-eigenvalue pair, we have

$$A^\top A \mathbf{v}_i = (\sigma_i)^2 \mathbf{v}_i.$$

Next comes what is, at least in my mind, the critical step in the proof. For now, assume that we have a full-rank matrix ($\sigma_i > 0$ for all i). Define a new vector \mathbf{u}_i such that,

$$\mathbf{u}_i = \frac{A \mathbf{v}_i}{\sigma_i}.$$

By construction, \mathbf{u}_i is a unit eigenvector of AA^\top ([Details, Section 4](#)). Now let V be an $n \times n$ matrix—because $A^\top A$ is $n \times n$ —where the i -th column is \mathbf{v}_i ; let U be an $m \times m$ matrix—because $A \mathbf{v}_i$ is an m -vector—where the i -th column is \mathbf{u}_i ; and let Σ be a diagonal matrix whose i -th element is σ_i . Then we can express the relationships we have so far in matrix form as:

$$\begin{aligned} U &= AV\Sigma^{-1} \\ U\Sigma &= AV \\ A &= U\Sigma V^\top \end{aligned}$$

where we use the fact that $VV^\top = I$ and Σ^{-1} is a diagonal matrix where the i -th value is the reciprocal of σ_i . And we're done. Note that the first r columns of V are an orthonormal basis for the row space of A , while the first r columns of U are an orthonormal basis for the column space of A .

In the low-rank scenario, some $\sigma_i = 0$. Provided the σ_i are sorted, we can complete U by adding additional column vectors that span \mathbb{R}^m and then add rows of 0-vectors to Σ . See Figure 7 [here](#).

Details

1. Gram matrices as positive semi-definite

Gram matrices are PSD. Consider an arbitrary Gram matrix $G = M^\top M$. Then we have:

$$\begin{aligned} \mathbf{x}^\top G \mathbf{x} &= \mathbf{x}^\top M^\top M \mathbf{x} \\ &= (M \mathbf{x})^\top M \mathbf{x} \\ &= (M \mathbf{x})^2 \\ &\geq 0 \end{aligned}$$

If that last step is not obvious, let $\mathbf{z} = M \mathbf{x}$ and note that

$$\mathbf{z}^\top \mathbf{z} = \sum_{i=1}^N (z_i)^2.$$

In general, positive-definiteness is required for any operation to be an inner product.

2. A and $A^\top A$ have the same rank

To show that $\text{rank}(A) = \text{rank}(A^\top A)$, it is sufficient to show that $A\mathbf{x} = \mathbf{0}$ and $A^\top A\mathbf{x} = \mathbf{0}$ have the same solutions, i.e. $A\mathbf{x} = \mathbf{0} \iff A^\top A\mathbf{x} = \mathbf{0}$. This makes sense because *rank* is just the maximal number of linearly independent columns, and a set of vectors

$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ are linearly *dependent* if there exists scalars a_1, a_2, \dots, a_k , not all zero, such that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_k\mathbf{x}_k = \mathbf{0}.$$

Now if $A\mathbf{x} = \mathbf{0}$, then $A^\top(A\mathbf{x}) = A^\top\mathbf{0} = \mathbf{0}$.

If $A^\top A\mathbf{x} = \mathbf{0}$, then

$$\begin{aligned} A^\top A\mathbf{x} &= \mathbf{0} \\ \mathbf{x}^\top A^\top A\mathbf{x} &= \mathbf{0} \\ (A\mathbf{x})^\top A\mathbf{x} &= \mathbf{0}. \end{aligned}$$

Note that for any vector \mathbf{v} , by definition of the inner product, $\mathbf{v}^\top \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0}$.

3. The eigenvalues of PSD matrices are all non-negative

An equivalent characterization of a PSD matrix is that all its eigenvalues are non-negative. First, consider a real symmetric matrix A . Since it is real and symmetric, it has an eigendecomposition of the form:

$$A = Q\Lambda Q^\top = \sum_{n=1}^N \mathbf{q}_n \lambda_n \mathbf{q}_n^\top$$

And therefore:

$$\begin{aligned} \mathbf{x}^\top A\mathbf{x} &= \mathbf{x}^\top \left(\sum_{n=1}^N \mathbf{q}_n \lambda_n \mathbf{q}_n^\top \right) \mathbf{x} \\ &= \sum_{n=1}^N \lambda_n \mathbf{x}^\top \mathbf{q}_n \mathbf{q}_n^\top \mathbf{x} \\ &= \sum_{n=1}^N \lambda_n (\mathbf{x}^\top \mathbf{q}_n)^2 \end{aligned}$$

This final expression is greater or equal to zero if all the eigenvalues λ_n are non-negative. So *if* that is true, the matrix A is PSD.

Next, consider a matrix that is PSD, so:

$$\mathbf{x}^\top A \mathbf{x} \geq 0$$

Now consider an arbitrary eigenvector of A , \mathbf{v}_i . Since A is PSD, we know:

$$\begin{aligned} \mathbf{v}_i^\top A \mathbf{v}_i &= \mathbf{v}_i^\top \lambda_i \mathbf{v}_i \\ &= \lambda_i \mathbf{v}_i^\top \mathbf{v}_i \\ &\geq 0 \end{aligned}$$

Since $\mathbf{v}_i^\top \mathbf{v}_i$ is necessarily non-negative, then λ_i must be non-negative for $\lambda_i \mathbf{v}_i^\top \mathbf{v}_i \geq 0$ to hold.

4. \mathbf{u}_i is a unit eigenvector of AA^\top

To see that \mathbf{u}_i is an eigenvector of AA^\top , note that

$$\begin{aligned} AA^\top \mathbf{u}_i &= AA^\top \left(\frac{A \mathbf{v}_i}{\sigma_i} \right) \\ &= AA^\top A \mathbf{v}_i \frac{1}{\sigma_i} \\ &= A(\sigma_i)^2 \mathbf{v}_i \frac{1}{\sigma_i} \\ &= (\sigma_i)^2 \mathbf{u}_i \end{aligned}$$

where step 2-3 applies the definition of \mathbf{v}_i as an eigenvector of $A^\top A$ and step 3-4 applies the definition of \mathbf{u}_i .

To see that \mathbf{u}_i is unit, note that

$$\begin{aligned} \mathbf{u}_i^\top \mathbf{u}_i &= \left(\frac{A \mathbf{v}_i}{\sigma_i} \right)^\top \frac{A \mathbf{v}_i}{\sigma_i} \\ &= \left(\frac{\mathbf{v}_i^\top A^\top}{\sigma_i} \right) \frac{A \mathbf{v}_i}{\sigma_i} \\ &= \frac{\mathbf{v}_i^\top A^\top A \mathbf{v}_i}{(\sigma_i)^2} \\ &= \frac{\mathbf{v}_i^\top (\sigma_i)^2 \mathbf{v}_i}{(\sigma_i)^2} \\ &= \mathbf{v}_i^\top \mathbf{v}_i \\ &= 1 \end{aligned}$$

where the last step is because \mathbf{v}_i is a unit vector as well.

Conclusion

I found this proof instructive. Importantly, it makes it clear where the relationship between singular values and eigenvalues comes from. The right singular vectors of A , the columns of V , are the set of orthonormal eigenvectors of $A^\top A$. The left singular vectors of A , the columns of U , are the set of orthonormal eigenvectors of AA^\top . And the non-zero singular values of A are the square roots of the eigenvalues of both $A^\top A$ and AA^\top .

Acknowledgements

I thank James D., Carl A., and José M. for pointing out a few mistakes in earlier drafts.

1. Trefethen, L. N., & Bau III, D. (1997). *Numerical linear algebra* (Vol. 50). Siam.