Singular value decomposition (SVD) generalizes matrix diagonalization to non-square matrices. It is powerful because it applies to **any** $m \times n$ matrix.

First we give the recipe:

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

Compute AA^T. This product is guaranteed to have m linearly independent eigenvectors by the spectral theorem. Collate those eigenvectors into the columns of a matrix. That matrix is U.

Similarly, compute A^TA, which is guaranteed to have n linearly independent eigenvectors that you can stick into V.

Next compute the non-zero eigenvalues λ_i of A^TA, take their square roots $\sigma_i = \sqrt{\lambda_i}$, and stick them into the diagonal of Σ.

Now Σ will consist of $\text{rank}(A)$ σ_i's on the diagonal and zeroes elsewhere. Add more zeroes on the diagonal to form a $\min(m, n) \times \min(m, n)$ square matrix. Then add the sufficient amount of rows/columns to form a $m \times n$ matrix—obviously the rows/columns consist entirely of zeroes.

We are almost done. Conventionally, the singular values in Σ should be ordered along the diagonal from greatest to least. This orderering of singular values should correspond to the ordering of eigenvectors in U and V (i.e. $Av_i = \sigma_i u_i$).

Next we look at the intuition. The first perspective we'll look at is SVD as a way of encoding linear transformations. For example, imagine you had any square defined by the two vectors at its sides—say $\vec{v_1}$ and $\vec{v_2}$. Left-multiplying these vectors by some A would then apply some linear transformation to the unit square. Here's the crucial claim of SVD: **any such linear transformation is equivalent to a rotation and a scaling, provided you are able to rotate the vector first.** More specifically the claim is

$$\begin{aligned} MR_1 &= R_2\Sigma \Rightarrow \\ M &= R_2\Sigma R_1^T \end{aligned}$$

which comes from right multiplication by R_1^T (which is orthogonal). This seems obvious for many transformations, such as rotations, scalings, and translations, but surprisingly it is true for shearing as well. Even more surprisingly, this holds true for transformations that change the dimension of the input vectors.

I like the other perspective as well. The SVD of A represents the action of A in the basis consisting of the eigenvectors of AA^T.