If A is diagonalizable, the equation $A = PDP^{-1}$ can be thought of as a factorization of the matrix A. If A isn't diagonalizable, we can use the ***singular value decomposition (SVD)*** instead.

Theorem (Singular Value Decomposition)

If A is an $m \times n$ matrix of rank k, then there exist matrices U, Σ, and V^T such that $A = U\Sigma V^T$ and

1. *$A^T A = VDV^{-1}$ for some diagonal matrix D.*
2. *Σ is a diagonal matrix. The first k diagonal entries of Σ are the square roots of the nonzero eigenvalues of $A^T A$ (the **singular values** of A), in decreasing order, and the rest are 0.*
3. *If $\vec{u}_i$ is the ith column of U, $\vec{v}_i$ is the ith row of V^T, and σ_i is the ith diagonal entry of Σ, then $\vec{u}_i = (1/\sigma_i)A\vec{v}_i$.*
4. *The first k columns of U form an orthonormal basis of $\operatorname{col}(A)$, and the first k columns of V form an orthonormal basis of $\operatorname{row}(A)$.*
5. *The columns of U form an orthonormal basis of $\mathbb{R}^m$, and the columns of V form an orthonormal basis of $\mathbb{R}^n$.*

Singular value decomposition (SVD)

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

Reduced SVD

If $\text{rank}(A) = r$, then we can take the first r columns of U and Σ and the first r rows of V in the SVD and it still works!

$$A = U_r \Sigma_r V_r^T$$

Rank k approximation of A

Taking fewer rows and columns of U, Σ, and V gives a good approximation of A with smaller rank.

$$A_k = U_k \Sigma_k V_k^T \text{ with } k < r$$

This is how the SVD is used for image compression.