The Singular Value Decomposition

Let $A \in \mathbb{R}^{m \times n}$. Then there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, and a diagonal matrix of *singular values* $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$, where $p = \min(m, n)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_p \geq 0$, such that

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

So what?

Recall the two fundamental subspaces associated with any matrix (or linear transformation) A: The range of A is the subspace of $\mathbb{R}^m$ defined as

$$\mathrm{Range}(A) = \{y \in \mathbb{R}^m : y = Ax, \text{ for some } x \in \mathbb{R}^n\},$$

and the nullspace of A is the subspace of $\mathbb{R}^n$ defined as

$$\mathrm{Nullsp}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

The rank of a matrix A is the dimension of the range of A, and the nullity of A is the dimension of the nullspace of A. One of the fundamental properties of an $m \times n$ matrix A is

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n.$$

In an inner product space, this result should be seen as a corollary to another fundamental result which says that the range of A is the orthogonal complement of the nullspace of A^T:

$$\mathrm{Range}(A) = [\mathrm{Nullsp}(A^T)]^\perp.$$

Applying this result to A^T gives

$$\mathrm{Range}(A^T) = [\mathrm{Nullsp}(A)]^\perp.$$

Back to the SVD: If $r = \mathrm{rank}(A)$, then $\sigma_r > 0$ and $\sigma_{r+1} = 0$. If we write $U = [U_1, U_2]$, and $V = [V_1, V_2]$, where $U_1 \in \mathbb{R}^{m \times r}$ and $V_1 \in \mathbb{R}^{n \times r}$, then (the columns of) U_1 form an orthonormal basis (O.B.) for $\mathrm{Range}(A)$, U_2 an O.B. for $\mathrm{Nullsp}(A^T)$, V_1 an O.B. for $\mathrm{Range}(A^T)$, and V_2 an O.B. for $\mathrm{Nullsp}(A)$.

It's all there in the SVD. And more. A matrix of rank s which best approximates A in the 2-norm is

$$A_s \equiv \sum_{j=1}^{s} \sigma_j u_j v_j^T.$$

This implies that the singular values tell us about how close A is to matrices of a given rank (e.g. "how close to singular is this square matrix?"), and helps us to quantify the uncertainties in our data.