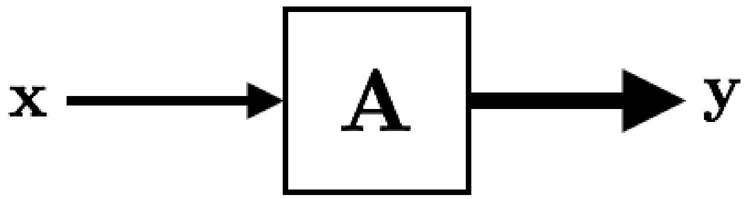
Singular Value Decomposition
Proof: Operator Norm is the Largest Singular Value

A matrix $\boldsymbol{A}$ may be viewed as an "operator" that acts on a vector $\boldsymbol{x}$ to produce a new vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ as shown below.



The operator or two norm of a matrix measures the largest possible amplification a given matrix $\boldsymbol{A}$ applies to any vector $\boldsymbol{x}$. That is, the operator norm is the maximum of the ratio $||\boldsymbol{y}||_2/||\boldsymbol{x}||_2$. This is written formally as

$$||\boldsymbol{A}||_2 = ||\boldsymbol{A}||_{op} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{||\boldsymbol{A}\boldsymbol{x}||_2}{||\boldsymbol{x}||_2}$$

We use the fact that $||c\boldsymbol{x}||_2 = c||\boldsymbol{x}||_2$ to rewrite the operator norm in the more convenient form

$$||\boldsymbol{A}||_2 = ||\boldsymbol{A}||_{op} = \max_{||\boldsymbol{x}||_2=1} ||\boldsymbol{A}\boldsymbol{x}||_2$$

Theorem: $||\boldsymbol{A}||_{op} = \sigma_1$ where σ_1 is the largest singular value of the matrix $\boldsymbol{A}$.

Proof: Substitute the singular value decomposition for $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ where $\boldsymbol{U}$ and $\boldsymbol{V}$ are square matrices (the non-economy or non-skinny SVD) to write

$$||\boldsymbol{A}||_2 = ||\boldsymbol{A}||_{op} = \max_{||\boldsymbol{x}||_2=1} ||\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}||_2$$

Now let $\boldsymbol{z} = \boldsymbol{V}^T\boldsymbol{x}$ and note that $||\boldsymbol{z}||_2^2 = \boldsymbol{x}^T\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x} = ||\boldsymbol{x}||_2^2$ because the right singular vectors in $\boldsymbol{V}$ are orthonormal. That is, $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$. Hence we may write

$$||\boldsymbol{A}||_2 = ||\boldsymbol{A}||_{op} = \max_{||\boldsymbol{z}||_2=1} ||\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{z}||_2$$

We may use the properties of the left singular vectors in $\boldsymbol{U}$ to eliminate the dependence on $\boldsymbol{U}$. We have $||\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{z}||_2^2 = \boldsymbol{z}^T\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{z} = ||\boldsymbol{\Sigma}\boldsymbol{z}||_2^2$ since the left singular vectors are also orthonormal, that is, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$. This results in

$$||\boldsymbol{A}||_2^2 = ||\boldsymbol{A}||_{op}^2 = \max_{||\boldsymbol{z}||_2=1} ||\boldsymbol{\Sigma}\boldsymbol{z}||_2^2$$

Let $\boldsymbol{A}$ have rank p so there are p nonzero singular values in the diagonal matrix $\boldsymbol{\Sigma}$. We thus may write

$$||\boldsymbol{\Sigma z}||_2^2 = \sum_{i=1}^{p} \sigma_i^2 z_i^2$$

where σ_i are the singular values and z_i is the i^{th} element of $\boldsymbol{z}$. We may rewrite the squared norm as

$$||\boldsymbol{A}||_2^2 = ||\boldsymbol{A}||_{op}^2 = \max_{z_1^2+z_2^2+\cdots+z_M^2=1} \sum_{i=1}^{p} \sigma_i^2 z_i^2$$

The unit norm constraint on $\boldsymbol{z}$ implies that an increase in the magnitude of any one element z_k must be offset by a decrease in the magnitudes of the other elements of $\boldsymbol{z}$. Clearly we should set $z_{p+1} = \cdots = z_M = 0$ since these elements do not contribute to the cost function. Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$, the best element to allocate the unit energy in $\boldsymbol{z}$ is z_1. To see this, consider the case where $p = M = 2$ and $\sigma_1 > \sigma_2$. In this case we require $z_2^2 = 1 - z_1^2$ and

$$\sum_{i=1}^{p} \sigma_i^2 z_i^2 = \sigma_1^2 z_1^2 + \sigma_2^2(1 - z_1^2) = z_1^2(\sigma_1^2 - \sigma_2^2) + \sigma_2^2$$

Since $\sigma_1^2 - \sigma_2^2 > 0$, this quantity is maximized by choosing z_1^2 to be as large or possible, that is, $z_1^2 = 1$. Any energy allocated to z_2 is multiplied by σ_2^2, which results in a lower value than had that same energy been allocated to z_1 because $\sigma_1 > \sigma_2$. A similar argument applies to the case $p > 2$. The strategy that maximizes $\sum_{i=1}^{p} \sigma_i^2 z_i^2$ allocates all of the unit energy in $\boldsymbol{z}$ to z_1, since σ_1 is the largest singular value.

Thus, we've shown that

$$||\boldsymbol{A}||_2 = ||\boldsymbol{A}||_{op} = \max_{||\boldsymbol{x}||_2=1} ||\boldsymbol{Ax}||_2 = \sigma_1$$

and this maximum value is obtained when $\boldsymbol{x} = \boldsymbol{V z} = \boldsymbol{v}_1$ where $\boldsymbol{v}_1$ is the right singular vector corresponding to the largest singular value.