Principal Component Analysis
Proof: Left Singular Vector is the First Principal Component

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_M \end{bmatrix}$ be an N-by-M ($N \geq M$) matrix with columns $\boldsymbol{a}_i$. The expression $\boldsymbol{f}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{f}$ represents the sum of squares of the elements of the vector $\boldsymbol{A}^T\boldsymbol{f}$, whose elements are the inner product between $\boldsymbol{f}$ and each column of $\boldsymbol{A}$. The solution to the problem

$$\max_{||\boldsymbol{f}||_2^2=1} \sum_{i=1}^{M} |\boldsymbol{f}^T\boldsymbol{a}_i|^2 = \max_{||\boldsymbol{f}||_2^2=1} \boldsymbol{f}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{f}$$

gives the direction $\boldsymbol{f}$ containing the maximum variability or variance across the columns of $\boldsymbol{A}$, that is, the direction that best fits the set of vectors $\boldsymbol{a}_i, i = 1, 2, \ldots, M$. The vector $\boldsymbol{f}$ is called the first principal component of the data $\boldsymbol{a}_i, i = 1, 2, \ldots, M$. Let the $\boldsymbol{A}$ have singular value decomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ where $\boldsymbol{U}, \boldsymbol{V}$ are square matrices, that is, the full singular value decomposition.

Theorem:

$$\max_{||\boldsymbol{f}||_2^2=1} \boldsymbol{f}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{f} = \sigma_1^2$$

is obtained by setting $\boldsymbol{f} = \boldsymbol{u}_1$, the left singular vector corresponding to the largest singular value.

Proof: Substitute the singular value decomposition for $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ to write

$$\boldsymbol{f}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{f} = \boldsymbol{f}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T\boldsymbol{f} = \boldsymbol{f}^T\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T\boldsymbol{f}$$

where the second equality follows from the orthonormality of the columns of $\boldsymbol{V}$: $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$. Now let $\boldsymbol{z} = \boldsymbol{U}^T\boldsymbol{f}$ and note that $||\boldsymbol{z}||_2^2 = \boldsymbol{f}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{f} = \boldsymbol{f}^T\boldsymbol{f} = ||\boldsymbol{f}||_2^2$ because the right singular vectors in $\boldsymbol{U}$ are orthonormal. That is, $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$. Hence we may rewrite the maximization problem as

$$\max_{||\boldsymbol{z}||_2=1} \boldsymbol{z}^T\boldsymbol{\Sigma}^2\boldsymbol{z} = \max_{\sum_{i=1}^{M} z_i^2=1} \sum_{j=1}^{M} \sigma_i^2 z_i^2$$

where z_i is the i^{th} element of $\boldsymbol{z}$.

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 $\boldsymbol{z}$. Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$, the best strategy is to allocate all of the unit energy in $\boldsymbol{z}$ to z_1. To see this, consider the case where $M = 2$ and $\sigma_1 > \sigma_2$. In this case we require $z_2^2 = 1 - z_1^2$ and

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

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 $M > 2$. 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, so $\boldsymbol{z} = \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \end{array}\right]$. The orthonormality of the columns of $\boldsymbol{U}$ thus imply that $\boldsymbol{f} = \boldsymbol{U}^T \boldsymbol{z}$ or $\boldsymbol{f} = \boldsymbol{u}_1$ where $\boldsymbol{u}_1$ is the first column of $\boldsymbol{U}$, the first left singular vector of $\boldsymbol{A}$.