Theorem 11.3.12: Let A be an $m \times n$ matrix, and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be its rows. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ be its right singular vectors, and let $\sigma_1, \ldots, \sigma_r$ be its singular values. For any positive integer $k \leq r$, Span $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is the k-dimensional vector space $\mathcal{V}$ that minimizes

$$(\text{distance from } \boldsymbol{a}_1 \text{ to } \mathcal{V})^2 + \cdots + (\text{distance from } \boldsymbol{a}_m \text{ to } \mathcal{V})^2$$

and the minimum sum of squared distances is $\|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2$.

Proof

By Lemma 11.3.11, the sum of squared distances for the space $\mathcal{V} = \text{Span } \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is

$$\|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2 \tag{11.1}$$

To prove that this is the minimum, we need to show that any other k-dimensional vector space $\mathcal{W}$ leads to a sum of squares that is no smaller.

Any k-dimensional vector space $\mathcal{W}$ has an orthonormal basis. Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ be such a basis. Plugging these vectors into Lemma 11.3.11, we get that the sum of squared distances from $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ to $\mathcal{W}$ is

$$\|A\|_F^2 - \|A\boldsymbol{w}_1\|^2 - \|A\boldsymbol{w}_2\|^2 - \cdots - \|A\boldsymbol{w}_k\|^2 \tag{11.2}$$

In order to show that $\mathcal{V}$ is the closest, we need to show that the quantity in 11.2 is no less than the quantity in 11.1. This requires that we show that $\|A\boldsymbol{w}_1\|^2 + \cdots + \|A\boldsymbol{w}_k\|^2 \leq \sigma_1^2 + \cdots + \sigma_k^2$. Let W be the matrix with columns $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$. Then $\|AW\|_F^2 = \|A\boldsymbol{w}_1\|^2 + \cdots + \|A\boldsymbol{w}_k\|^2$ by the column analogue of Lemma 11.1.1. We must therefore show that $\|AW\|_F^2 \leq \sigma_1^2 + \cdots + \sigma_k^2$.

By Theorem 11.3.10, A can be factored as $A = U\Sigma V^T$ where the columns of V are $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$, and where U and V are column-orthogonal and Σ is the diagonal matrix with diagonal elements $\sigma_1, \ldots, \sigma_r$. By substitution, $\|AW\|_F^2 = \|U\Sigma V^T W\|_F^2$. Since U is column-orthogonal, multiplication by U preserves norms, so $\|U\Sigma V^T W\|_F^2 = \|\Sigma V^T W\|_F^2$.

Let X denote the matrix $V^T W$. The proof makes use of two diferent interpretations of X, in terms of columns and in terms of rows.

First, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ denote the columns of X. For $j = 1, \ldots, k$, by the matrix-vector interpretation of matrix-matrix multiplication, $\boldsymbol{x}_j = V^T \boldsymbol{w}_j$. By the dot-product interpretation of matrix-vector multiplication, $\boldsymbol{x}_j = [\boldsymbol{v}_1 \cdot \boldsymbol{w}_j, \ldots, \boldsymbol{v}_r \cdot \boldsymbol{w}_j]$, which is the coordinate representation in terms of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ of the projection of $\boldsymbol{w}_j$ onto Span $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$. Therefore the projection itself is $V\boldsymbol{x}_j$. The projection of a norm-one vector onto a space has norm at most one, so $\|V\boldsymbol{x}_j\| \leq 1$. Since V is a column-orthogonal matrix, $\|V\boldsymbol{x}_j\| = \|\boldsymbol{x}_j\|$, so $\boldsymbol{x}_j$ has norm at most one. This shows that $\|X\|_F^2 \leq k$.

Second, let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ denote the rows of X. For $i = 1, \ldots, r$, by the vector-matrix interpretation of matrix-matrix multiplication, $\boldsymbol{y}_i = \boldsymbol{v}_i^T W$. By the dot-product interpretation of vector-matrix multiplication, $\boldsymbol{y}_i = [\boldsymbol{v}_i \cdot \boldsymbol{w}_1, \ldots, \boldsymbol{v}_i \cdot \boldsymbol{w}_k]$, which is the coordinate representation in terms of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ of the projection of $\boldsymbol{v}_i$ onto $\mathcal{W}$. Using the same argument as before, since $\boldsymbol{v}_i$ has norm one, the coordinate representation has norm at most one. This shows that each row $\boldsymbol{y}_i$ of X has norm at most one.

Now we consider ΣX. Since Σ is a diagonal matrix with diagonal elements $\sigma_1, \ldots, \sigma_r$, it follows that row i of ΣX is σ_i times row i of X, which is $\sigma_i \boldsymbol{y}_i$. Therefore the squared Frobenius norm of ΣX is $\sigma_1^2 \|\boldsymbol{y}_1\|^2 + \cdots \sigma_r^2 \|\boldsymbol{y}_r\|^2$. How big can that quantity be?

Imagine you have k dollars to spend on r products. Product i gives you value σ_i^2 per dollar you spend on it. Your goal is to maximize the total value you receive. Since $\sigma_1 \geq \cdots \geq \sigma_r$, it makes sense to spend as much as you can on product 1, then spend as much of your remaining money on product 2, and so on. You are not allowed to spend more than one dollar on each product. What do you do? You spend one dollar on product 1, one dollar on product 2, ..., one dollar on product k, and zero dollars on the remaining products. The total value you receive is then $\sigma_1^2 + \cdots + \sigma_r^2$.

Now we formally justify this intuition. Our goal is to show that $\sigma_1^2\|\boldsymbol{y}_1\|^2+\cdots+\sigma_r^2\|\boldsymbol{y}_r\|^2 \leq \sigma_1^2 + \cdots + \sigma_k^2$. We have shown that $\|\boldsymbol{y}_i\|^2 \leq 1$ for $i = 1, \ldots, k$. Since $\|X\|_F^2 \leq k$, we also know that $\|\boldsymbol{y}_1\|^2 + \cdots + \|\boldsymbol{y}_r\|^2 \leq k$.

Define $\beta_i = \begin{cases} \sigma_i^2 - \sigma_k^2 & \text{if } i \leq r \\ 0 & \text{otherwise} \end{cases}$

Then $\sigma_i^2 \leq \beta_i + \sigma_k^2$ for $i = 1, \ldots, r$ (using the fact that $\sigma_1, \ldots, \sigma_r$ are in nonincreasing order).

Therefore

$$\begin{aligned}
\sigma_1^2\|\boldsymbol{y}_1\|^2 + \cdots + \sigma_r^2\|\boldsymbol{y}_r\|^2 &\leq (\beta_1 + \sigma_k^2)\|\boldsymbol{y}_1\|^2 + \cdots + (\beta_r + \sigma_k^2)\|\boldsymbol{y}_r\|^2 \\
&= \left(\beta_1\|\boldsymbol{y}_1\|^2 + \cdots + \beta_r\|\boldsymbol{y}_r\|^2\right) + \left(\sigma_k^2\|\boldsymbol{y}_1\|^2 + \cdots + \sigma_k^2\|\boldsymbol{y}_r\|^2\right) \\
&\leq (\beta_1 + \cdots + \beta_r) + \sigma_k^2\left(\|\boldsymbol{y}_1\|^2 + \cdots + \|\boldsymbol{y}_r\|^2\right) \\
&\leq \left(\sigma_1^2 + \cdots + \sigma_k^2 - k\sigma_k^2\right) + \sigma_k^2 k \\
&= \sigma_1^2 + \cdots + \sigma_k^2
\end{aligned}$$

This completes the proof. □