Detailed Proofs of the Eckart-Young-Mirsky Theorem

Low-Rank Matrix Approximation

Statement of the Theorem

Setup

Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank r. Its Singular Value Decomposition (SVD) is $A = U\Sigma V^T$, with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Low-Rank Approximation A_k

For any integer $k < r$, the truncated SVD gives the rank-k matrix:

$$A_k = \sum_{i=1}^{k} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$$

Statement of the Theorem

Theorem (Eckart-Young-Mirsky)

For any matrix $B \in \mathbb{R}^{m \times n}$ with $\text{rank}(B) \leq k$, A_k is the best approximation to A:

1. ***Spectral Norm:*** $\|A - A_k\|_2 = \min_{rank(B) \leq k} \|A - B\|_2 = \sigma_{k+1}$
2. ***Frobenius Norm:*** $\|A - A_k\|_F = \min_{rank(B) \leq k} \|A - B\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}$

Proof for the Spectral Norm - Part 1: Error of A_k

Calculating the error $\|A - A_k\|_2$

The difference matrix is:

$$A - A_k = \sum_{i=1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T - \sum_{i=1}^{k} \sigma_i \mathrm{u}_i \mathrm{v}_i^T = \sum_{i=k+1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$$

This is the SVD of $A - A_k$.

Proof for the Spectral Norm - Part 1: Error of A_k

Calculating the error $\|A - A_k\|_2$

The difference matrix is:

$$A - A_k = \sum_{i=1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T - \sum_{i=1}^{k} \sigma_i \mathrm{u}_i \mathrm{v}_i^T = \sum_{i=k+1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$$

This is the SVD of $A - A_k$.
The largest singular value of this matrix is σ_{k+1}. By definition of the spectral norm (which is the largest singular value), we have:

$$\|A - A_k\|_2 = \sigma_{k+1}$$

Proof for the Spectral Norm - Part 1: Error of A_k

Goal

Now, we must show that for any other matrix B with $\text{rank}(B) \leq k$, the error is at least this large:

$$\|A - B\|_2 \geq \sigma_{k+1}$$

Proof for the Spectral Norm - Part 2: Subspace Intersection

- Let B be any matrix with $\text{rank}(B) \leq k$.

Proof for the Spectral Norm - Part 2: Subspace Intersection

- Let B be any matrix with $\mathrm{rank}(B) \leq k$.
- The null space of B, $\mathcal{N}(B)$, has dimension:

$$\dim(\mathcal{N}(B)) = n - \mathrm{rank}(B) \geq n - k$$

Proof for the Spectral Norm - Part 2: Subspace Intersection

- Let B be any matrix with $\text{rank}(B) \leq k$.
- The null space of B, $\mathcal{N}(B)$, has dimension:

$$\dim(\mathcal{N}(B)) = n - \text{rank}(B) \geq n - k$$

- Let S_{k+1} be the subspace spanned by the first $k+1$ right singular vectors of A:

$$S_{k+1} = \text{span}\{\text{v}_1, \dots, \text{v}_{k+1}\}$$

  This subspace has $\dim(S_{k+1}) = k+1$.

Proof for the Spectral Norm - Part 2: Subspace Intersection

- Let B be any matrix with $\text{rank}(B) \leq k$.
- The null space of B, $\mathcal{N}(B)$, has dimension:

$$\dim(\mathcal{N}(B)) = n - \text{rank}(B) \geq n - k$$

- Let S_{k+1} be the subspace spanned by the first $k+1$ right singular vectors of A:

  $$S_{k+1} = \text{span}\{\mathrm{v}_1, \ldots, \mathrm{v}_{k+1}\}$$

  This subspace has $\dim(S_{k+1}) = k + 1$.
- Since $\dim(\mathcal{N}(B)) + \dim(S_{k+1}) \geq (n-k) + (k+1) = n + 1 > n$, their intersection is non-trivial.

Proof for the Spectral Norm - Part 2: Subspace Intersection

- Let B be any matrix with $\text{rank}(B) \leq k$.
- The null space of B, $\mathcal{N}(B)$, has dimension:

$$\dim(\mathcal{N}(B)) = n - \text{rank}(B) \geq n - k$$

- Let S_{k+1} be the subspace spanned by the first $k+1$ right singular vectors of A:

$$S_{k+1} = \text{span}\{\mathrm{v}_1, \ldots, \mathrm{v}_{k+1}\}$$

  This subspace has $\dim(S_{k+1}) = k + 1$.
- Since $\dim(\mathcal{N}(B)) + \dim(S_{k+1}) \geq (n-k) + (k+1) = n + 1 > n$, their intersection is non-trivial.
- Therefore, there must exist a non-zero unit vector z such that:

$$\mathrm{z} \in \mathcal{N}(B) \quad \text{and} \quad \mathrm{z} \in S_{k+1}$$

Proof for the Spectral Norm - Part 3: The Lower Bound

- We know $Bz = 0$ and $\mathrm{z} = \sum_{i=1}^{k+1} c_i \mathrm{v}_i$ for some scalars c_i with $\sum c_i^2 = 1$.

Proof for the Spectral Norm - Part 3: The Lower Bound

- We know $Bz = 0$ and $z = \sum_{i=1}^{k+1} c_i v_i$ for some scalars c_i with $\sum c_i^2 = 1$.
- The squared norm of the error is:

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2 = \|Az\|_2^2$$

Proof for the Spectral Norm - Part 3: The Lower Bound

- We know $Bz = 0$ and $z = \sum_{i=1}^{k+1} c_i v_i$ for some scalars c_i with $\sum c_i^2 = 1$.
- The squared norm of the error is:

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2 = \|Az\|_2^2$$

- We compute Az:

$$Az = \left(\sum_{j=1}^{r} \sigma_j u_j v_j^T\right) \left(\sum_{i=1}^{k+1} c_i v_i\right) = \sum_{i=1}^{k+1} c_i \sigma_i u_i$$

Proof for the Spectral Norm - Part 3: The Lower Bound

- We know $Bz = 0$ and $z = \sum_{i=1}^{k+1} c_i v_i$ for some scalars c_i with $\sum c_i^2 = 1$.
- The squared norm of the error is:

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2 = \|Az\|_2^2$$

- We compute Az:

$$Az = \left(\sum_{j=1}^{r} \sigma_j u_j v_j^T\right)\left(\sum_{i=1}^{k+1} c_i v_i\right) = \sum_{i=1}^{k+1} c_i \sigma_i u_i$$

- Taking the squared norm:

$$\|Az\|_2^2 = \left\|\sum_{i=1}^{k+1} c_i \sigma_i u_i\right\|_2^2 = \sum_{i=1}^{k+1} c_i^2 \sigma_i^2$$

Proof for the Spectral Norm - Part 3: The Lower Bound

- Since $\sigma_i \geq \sigma_{k+1}$ for $i \leq k+1$:

$$\sum_{i=1}^{k+1} c_i^2 \sigma_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} c_i^2 = \sigma_{k+1}^2$$

Proof for the Spectral Norm - Part 3: The Lower Bound

- Since $\sigma_i \geq \sigma_{k+1}$ for $i \leq k+1$:

$$\sum_{i=1}^{k+1} c_i^2 \sigma_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} c_i^2 = \sigma_{k+1}^2$$

- So, $\|A - B\|_2^2 \geq \sigma_{k+1}^2$, which completes the proof.

Proof for Frobenius norm. Key Tool: Weyl's Inequality

The proof relies on a powerful result from matrix analysis.

Theorem (Weyl's Inequality for Singular Values)

For any two matrices $X, Y \in \mathbb{R}^{m \times n}$, the following inequality holds for their singular values:

$$\sigma_{i+j-1}(X + Y) \leq \sigma_i(X) + \sigma_j(Y)$$

Our Application of Weyl's Inequality

Let B be any matrix of rank k. This means its $(k+1)$-th singular value and all subsequent ones are zero: $\sigma_{k+1}(B) = 0$.
We apply the inequality with $X = A - B$ and $Y = B$. Let $j = k + 1$. Then for any $i \geq 1$:

$$\sigma_{i+k}(A) = \sigma_{i+(k+1)-1}(A) \leq \sigma_i(A - B) + \sigma_{k+1}(B)$$

Since $\sigma_{k+1}(B) = 0$, we get a crucial relationship:

$$\sigma_{i+k}(A) \leq \sigma_i(A - B)$$

Step 1: Expressing the Error of A_k

First, let's explicitly write the squared Frobenius norm of the error for our optimal matrix A_k.
The Frobenius norm of a matrix is the square root of the sum of squares of its singular values.
The matrix $A - A_k$ is:

$$A - A_k = \left(\sum_{i=1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T\right) - \left(\sum_{i=1}^{k} \sigma_i \mathrm{u}_i \mathrm{v}_i^T\right) = \sum_{i=k+1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$$

The singular values of the matrix $(A - A_k)$ are precisely $\{\sigma_{k+1}(A), \sigma_{k+2}(A), \ldots, \sigma_r(A)\}$.
Therefore, its squared Frobenius norm is:

$$\|A - A_k\|^2 = \sum_{i=k+1}^{r} \sigma_i(A)^2$$

Step 2: Connecting the Errors with Weyl's Inequality

Now we use the result from Weyl's inequality, $\sigma_{i+k}(A) \leq \sigma_i(A - B)$, to relate the error of A_k to the error of any other rank-k matrix B.
Let's re-index the sum for $\|A - A_k\|^2$ by letting $j = i - k$ (so $i = j + k$):

$$\|A - A_k\|^2 = \sum_{i=k+1}^{r} \sigma_i(A)^2 = \sum_{j=1}^{r-k} \sigma_{j+k}(A)^2$$

Now, apply Weyl's inequality to each term in the sum:

$$\sum_{j=1}^{r-k} \sigma_{j+k}(A)^2 \leq \sum_{j=1}^{r-k} \sigma_j(A - B)^2$$

The sum on the right is a sum of some of the squared singular values of the matrix $(A - B)$.

Step 3: Finalizing the Proof

We have the inequality: $\|A - A_k\|^2 \leq \sum_{j=1}^{r-k} \sigma_j(A-B)^2$

The full squared Frobenius norm of $(A - B)$ is the sum of *all* its squared singular values:

$$\|A - B\|^2 = \sum_{j=1}^{\min(m,n)} \sigma_j(A-B)^2$$

Since the terms in the sum are non-negative, we have:

$$\sum_{j=1}^{r-k} \sigma_j(A-B)^2 \leq \sum_{j=1}^{\min(m,n)} \sigma_j(A-B)^2 = \|A - B\|^2$$

Conclusion

Combining the steps, we get: $\|A - A_k\|^2 \leq \|A - B\|^2$
Taking the square root of both sides gives the final result: $\|A - A_k\| \leq \|A - B\|$
This proves that A_k is the best rank-k approximation to A in the Frobenius norm. $\square$