Essay

Not peer-reviewed version

A New Proof Of Eckart-Young-Mirsky Theorem

Haoyuan Wang *

Posted Date: 17 February 2025

doi: 10.20944/preprints202502.1203.v1

Keywords: Frobenius Norm; SVD; Low Rank Approximation

Essay

A New Proof of Eckart-Young-Mirsky Theorem

Haoyuan Wang

The High School Affiliated to Beijing Normal University; 17319269550@163.com

Abstract: It is apologize to upload this incomplete draft since the time limitation, yet in this paper, we are going to give a new proof of the Eckart-Young-Mirsky Theorem which is crucial in machine learning, image and date processing etc.

Keywords: Frobenius norm; SVD; low rank approximation

MSC: 15A18; 15A60

---

1. Introduction

The Eckart–Young–Mirsky theorem is a fundamental result in matrix approximation, stating that for a given matrix A and rank k, the best rank-k approximation in the Frobenius norm (or any unitarily invariant norm) is obtained by truncating the Singular Value Decomposition (SVD) of A.[1,3] Formally, if

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

is the SVD of A and Σ_k is obtained from Σ by keeping only the k largest singular values (and setting the rest to zero), then

$$A_k = U\Sigma_k V^T$$

is the unique minimizer of $\|A - X\|$ over all rank-k matrices X [2,4].

This theorem underpins numerous applications, from image compression to principal component analysis, yet standard proofs often rely on variational arguments or operator norm inequalities that can obscure geometric intuition [5]. In this paper, we present a more elementary proof(only using basic linear algebraic).

2. Proof

In this section, we will give an elementary and short proof of the Eckart-Young-Mirsky Theorem.

Let A be a real matrix with $\text{rank}(A) = r$ and $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ in a descending order be all the non-zero singular values of A. The SVD factors A into

$$A = U\Sigma V^t = U\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} V^t$$

where U and V are orthogonal matrices and $\Lambda = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is a $r \times r$ diagonal matrix.

Let $0 < k < r$ be an integer. Define

$$A_k = U\Sigma_k V^t = U\begin{pmatrix} \Lambda_k & 0 \\ 0 & 0 \end{pmatrix} V^t$$

with $\Lambda_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0)$ being a $r \times r$ diagonal matrix.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2025 doi:10.20944/preprints202502.1203.v1

The **Eckart-Young-Mirsky Theorem** states that

$$\|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2} = \min_{\text{rank}(X)=k} \|A - X\|_F \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm defined by

$$\|A\|_F = \sqrt{\text{trace}(A^t A)} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}. \tag{2}$$

for any real matrix $A = (a_{ij})_{m \times n}$.

In the following, we will relax the condition $\text{rank}(X) = k$ to $\text{rank}(X) \le k$ and prove that

$$\min_{\text{rank}(X)\le k} \|A - X\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}. \tag{3}$$

Let

$$X = UYV^t = U \begin{pmatrix} M & * \\ * & * \end{pmatrix} V^t$$

with M being a $r \times r$ matrix of $\text{rank}(M) \le k$. By (2),

$$\|A - X\|_F = \|U(\Sigma - Y)V^t\|_F = \|\Sigma - Y\|_F \ge \|\Lambda - M\|_F.$$

Therefore, to show (3), it suffices to prove

$$\|\Lambda - M\|_F \ge \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}. \tag{4}$$

as $X = A_k$ achieves the minimum.

Fix a k-dimensional subspace $W \subset \mathbb{R}^r$ such that the column vectors of M lie in W. Choose an orthonormal basis $v_1, v_2, ..., v_p, v_{p+1}, ..., v_r$ of $\mathbb{R}^r$ such that $v_{p+1}, ..., v_r$ span W, where $p + k = r$. Let

$$\Lambda = (\sigma_1 e_1, \sigma_2 e_2, \ldots, \sigma_r e_r) \text{ and } M = (w_1, w_2, \ldots, w_r)$$

where $\sigma_i e_i$-s and w_i-s are column vectors of Λ and M, respectively. We have

$$\|\Lambda - M\|_F^2 = \sum_{i=1}^{r} \|\sigma_i e_i - w_i\|^2. \tag{5}$$

To minimize (5), w_i should be the projection of $\sigma_i e_i$ onto W, i.e.,

$$\sigma_i e_i - w_i = \sum_{j=1}^{p} \langle \sigma_i e_i, v_j \rangle v_j$$

where $\langle , \rangle$ is the standard inner product. Then for any M whose column vectors are in W,

$$\min \|\Lambda - M\|_F^2 = \sum_{i=1}^{r} \sum_{j=1}^{p} \langle e_i, v_j \rangle^2 \sigma_i^2. \tag{6}$$

 doi:10.20944/preprints202502.1203.v1

The coefficients of σ_i^2-s of (6) satisfy:

$$0 \leq \sum_{j=1}^{p} \langle e_i, v_j \rangle^2 \leq \langle e_i, e_i \rangle \leq 1;$$
$$\sum_{i=1}^{r} \sum_{j=1}^{p} \langle e_i, v_j \rangle^2 = \sum_{j=1}^{p} \sum_{i=1}^{r} \langle e_i, v_j \rangle^2 = p.$$

Since $\sigma_1^2 \geq \sigma_2^2 \geq ... \geq \sigma_r^2 > 0$ are in descending order and their coefficients all belong to $[0, 1]$ with the sum being p, to minimize the right hand side of (6), the coefficients should concentrate to the lowest p singular values. Therefore,

$$\sum_{i=1}^{r} \sum_{j=1}^{p} \langle e_i, v_j \rangle^2 \sigma_i^2 \geq \sum_{i=k+1}^{r} \sigma_i^2. \tag{7}$$

3. Conclusion

This paper offers an elementary yet powerful proof of the Eckart-Young-Mirsky theorem, which is essential for many fields, such as machine learning, image processing, and data science. By demonstrating the best rank-k approximation through a clear application of basic linear algebra techniques, the paper contributes to a deeper understanding of low-rank matrix approximation. This work simplifies the theorem's proof, making it more accessible for those familiar with basic matrix theory and reinforcing its crucial role in real-world applications like dimensionality reduction, data compression, and statistical analysis.

Combining (6) and (7), we get (4), which concludes the proof.

References

1. N. Kishore Kumar and J. Schneider, *Literature survey on low rank approximation of matrices*, Linear and Multi-linear Algebra, vol. 65, no. 11, pp. 2212-2244, 2017.
2. L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, 1997.
3. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 2013.
4. G. W. Stewart, *Matrix Algorithms: Volume I: Basic Decompositions*, SIAM, 1998.
5. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.