Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 37, pp. 544-548, July 2021.

ERROR ANALYSIS OF THE GENERALIZED LOW-RANK MATRIX APPROXIMATION*

PABLO SOTO-QUIROS†

Abstract. In this paper, we propose an error analysis of the generalized low-rank approximation, which is a generalization of the classical approximation of a matrix $A \in \mathbb{R}^{m\times n}$ by a matrix of a rank at most r, where $r \leq \min\{m,n\}$.

Key words. Generalized low-rank approximation, Frobenius norm, SVD, Pseudoinverse, Error analysis.

AMS subject classifications. 15A18, 15A29.

1. Introduction. Throughout this paper, we adopt the following notation: $\mathbb{R}_r^{m\times n}$ denotes the set of all real $m\times n$ matrices of rank at most r, where $r \leq \min\{m,n\}$, i.e., $A \in \mathbb{R}_r^{m\times n}$ if and only if $\operatorname{rank}(A) \leq r$. $I_m \in \mathbb{R}^{m\times m}$ and $\boldsymbol{0}_{m\times n} \in \mathbb{R}^{m\times n}$ are the identity matrix of order m and the null $m\times n$ matrix, respectively. $M^\dagger$, $\operatorname{tr}\{M\}$, $\|M\|$ and $\mathcal{N}(M)$ denote the pseudoinvserse, the trace, the Frobenius norm and the null space of M, respectively. Furthermore, $N^{1/2}$ is the square root of $N \in \mathbb{R}^{m\times m}$, i.e., $N = N^{1/2}N^{1/2}$.

Let $D = U_D\Sigma_D V_D^T$ be the singular value decomposition (SVD) of $D \in \mathbb{R}^{m\times n}$, where $U_D \in \mathbb{R}^{m\times m}$ and $V_D \in \mathbb{R}^{n\times n}$ are two orthogonal matrices, and $\Sigma_D = \operatorname{diag}(\sigma_1(D), ..., \sigma_{\min(m,n)}(D)) \in \mathbb{R}^{m\times n}$ is a generalized diagonal matrix with singular values $\sigma_1(D) \geq \sigma_2(D) \geq ... \geq \sigma_{\min(m,n)}(D) \geq 0$ on the main diagonal. The r-truncated SVD is defined by

$$\lfloor D\rfloor_r = \sum_{i=1}^{r} \sigma_i(D) u_i v_i^T = U_{D,r}\Sigma_{D,r}V_{D,r}^T \in \mathbb{R}^{m\times n},$$

where $U_{D,r} \in \mathbb{R}^{m\times r}$ and $V_{D,r} \in \mathbb{R}^{n\times r}$ are formed with the first r columns of U_D and V_D, respectively, and $\Sigma_{D,r} = \operatorname{diag}(\sigma_1(D), \ldots, \sigma_r(D)) \in \mathbb{R}^{r\times r}$. If $k = \operatorname{rank}(D)$, then $P_{D,L} \in \mathbb{R}^{m\times m}$ and $P_{D,R} \in \mathbb{R}^{n\times n}$ are the orthogonal projections of D on the range of D and D^T, respectively, where $P_{D,L} = DD^\dagger = U_{D,k}U_{D,k}^T$ and $P_{D,R} = D^\dagger D = V_{D,k}V_{D,k}^T$.

A generalization of the low-rank approximation was proposed in [2, 3, 4]. Given matrices $A \in \mathbb{R}^{p\times q}$, $B \in \mathbb{R}^{p\times m}$ and $C \in \mathbb{R}^{n\times q}$, and rank $r \leq \min\{m,n\}$, the generalized low-rank approximation finds a matrix $\widehat{X}_r \in \mathbb{R}_r^{m\times n}$ such that

$$\|A - B\widehat{X}_rC\|^2 = \min_{X\in\mathbb{R}_r^{m\times n}} \|A - BXC\|^2 . \tag{1.1}$$

Note that if B and C are identity matrices, the problem (1.1) is the well-known low-rank approximation problem proposed by Eckart and Young [1]. The problem in (1.1) was studied in [2, 3, 4] by Sonderman, Friedland and Torokthi, respectively. The following theorem presents the solution of the generalized low-rank approximation given in [3].

*Received by the editors on March 4, 2021. Accepted for publication on June 27, 2021. Handling Editor: Shmuel Friedland.
†Escuela de Matemática, Instituto Tecnológico de Costa Rica, Cartago 30101, Costa Rica (jusoto@tec.ac.cr).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 37, pp. 544-548, July 2021.

THEOREM 1.1. *Let* $A \in \mathbb{R}^{p\times q}$, $B \in \mathbb{R}^{p\times m}$ *and* $C \in \mathbb{R}^{n\times q}$, *and let* $M = U_M \Sigma_M V_M^T$ *be the SVD of* $M = P_{B,L} A P_{C,R} = BB^\dagger AC^\dagger C$. *Then matrix*

$$\widehat{X}_r = B^\dagger \lfloor M \rfloor_r C^\dagger = B^\dagger U_{M,r} \Sigma_{M,r} V_{M,r}^T C^\dagger, \tag{1.2}$$

minimizes the problem (1.1). *This solution is unique if and only if either*

$$r \geq \operatorname{rank}(M),$$

or

$$1 \leq r < \operatorname{rank}(M) \quad \text{and} \quad \sigma_r(M) \geq \sigma_{r+1}(M).$$

In this paper, we present an error analysis for the solution of the problem (1.1). The main result of this paper is presented below, in Theorem 3.1. In the related work, for the same problem, in [5, Theorem 3.2] Wang presented an error analysis for the specific case of $\operatorname{rank}(\widehat{X}_r) = r$. In this paper, we extend the analysis to the case $\operatorname{rank}(\widehat{X}_r) \leq r$. The formula for the error in [5, Theorem 3.2] is different from that in Theorem 3.1.

2. Preliminaries.

In this section, we present some preliminary results that will be used in the next section to study the error associated with the solution of the problem (1.1).

LEMMA 2.1 (Theorem 2.8 in [6]). *If* $A \in \mathbb{R}^{m\times n}$ *and* $B \in \mathbb{R}^{n\times m}$, *then* AB *and* BA *have the same nonzero eigenvalues, counting multiplicity.*

LEMMA 2.2 (Propositions 3.1 and 3.2 in [7]). $A^\dagger = A^T(A^\dagger)^T A^\dagger = A^\dagger (A^\dagger)^T A^T = (A^T A)^\dagger A^T$, *for all* $A \in \mathbb{R}^{m\times n}$.

LEMMA 2.3 (Lemma 2.4.1 in [8]). *Let* $A \in \mathbb{R}^{m\times n}$ *and* $B \in \mathbb{R}^{p\times m}$. *Then* $\mathcal{N}(A) \subseteq \mathcal{N}(BA)$.

LEMMA 2.4 (Lemma 23 in [9] - Fact 2 in [10]). *For any* $M \in \mathbb{R}^{m\times n}$, $N \in \mathbb{R}^{p\times n}$ *and* $S \in \mathbb{R}^{m\times s}$, *the following statements hold.*

(a) *If* $\mathcal{N}(M) \subseteq \mathcal{N}(N)$, *then* $NM^\dagger M = N$.
(b) *If* $\mathcal{N}(M^T) \subseteq \mathcal{N}(S^T)$, *then* $MM^\dagger S = S$.

LEMMA 2.5. *Let* $M \in \mathbb{R}^{m\times n}$ *and* $r \leq \min\{m, n\}$. *Then* $\mathcal{N}(M) \subseteq \mathcal{N}(\lfloor M \rfloor_r)$.

Proof. Without loss of generality, we assume $m \leq n$. Let $M = U_M \Sigma_M V_M^T$ be the SVD of M. If $x \in \mathcal{N}(M)$, then $Mx = \boldsymbol{0}_{n\times 1}$, and therefore, $\Sigma_M V_M^T x = \boldsymbol{0}_{n\times 1}$, because U_M is an orthogonal matrix. If $\overline{\Sigma}_{M,r} = \operatorname{diag}(\sigma_1(M), ..., \sigma_r(M), 0, ..., 0) \in \mathbb{R}^{m\times n}$, then

$$\begin{aligned}\Sigma_M V_M^T x = \boldsymbol{0}_{n\times 1} &\Rightarrow \begin{bmatrix} I_r & \boldsymbol{0}_{k\times m-r} \\ \boldsymbol{0}_{m-r\times r} & \boldsymbol{0}_{m-r\times m-r} \end{bmatrix} \Sigma_M V_M^T x = \boldsymbol{0}_{n\times 1} \\ &\Rightarrow \overline{\Sigma}_{M,r} V_M^T x = \boldsymbol{0}_{n\times 1} \\ &\Rightarrow U_M \overline{\Sigma}_{M,r} V_M^T x = \boldsymbol{0}_{n\times 1}.\end{aligned}$$

Note that

$$U_M \overline{\Sigma}_{M,r} V_M^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i(M) u_i v_i^T = \sum_{i=1}^{r} \sigma_i(M) u_i v_i^T = \lfloor M \rfloor_r.$$

Finally, $\lfloor M \rfloor_r x = \boldsymbol{0}_{n\times 1}$. Thus, $x \in \mathcal{N}(\lfloor M \rfloor_r)$. ∎

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 37, pp. 544-548, July 2021.

LEMMA 2.6. *If $M \in \mathbb{R}^{m\times n}$, $S \in \mathbb{R}^{n\times s}$ and $R \in \mathbb{R}^{s\times m}$, then the following statements hold.*

(a) $\lfloor RM \rfloor_r M^\dagger M = \lfloor RM \rfloor_r$.
(b) $SS^\dagger \lfloor SR \rfloor_r = \lfloor SR \rfloor_r$.

Proof. We consider (a). It follows from Lemma 2.3 that $\mathcal{N}(M) \subseteq \mathcal{N}(RM)$. By Lemma 2.5, we obtain $\mathcal{N}(RM) \subseteq \mathcal{N}(\lfloor RM \rfloor_r)$, and therefore, $\mathcal{N}(M) \subseteq \mathcal{N}(\lfloor RM \rfloor_r)$. As a result, (a) follows from Lemma 2.4. The proof of (b) is similar to the proof of (a). ∎

3. Main Result.

THEOREM 3.1. *Let $A \in \mathbb{R}^{p\times q}$, $B \in \mathbb{R}^{p\times m}$, $C \in \mathbb{R}^{n\times q}$ and $r \le \min\{m, n\}$. The error of the solution of the problem (1.1) is given by*

$$\min_{X\in\mathbb{R}_r^{m\times n}} \|A - BXC\|^2 = \|A\|^2 - \sum_{i=1}^{r} \lambda_i(T), \tag{3.3}$$

where $T = B^\dagger AC^\dagger CA^T B \in \mathbb{R}^{m\times m}$ and $\lambda_i(T)$ is the $i-th$ eigenvalue of T with $\lambda_1(T) \ge \lambda_2(T) \ge ... \ge \lambda_m(T)$.

Proof. It follows from the identity $||D||^2 = \mathrm{tr}\{DD^T\}$ and the linearity of the trace operator that

$$\|A - BXC\|^2 = \|A\|^2 - \|M\|^2 + \|M - (B^T B)^{1/2} X (CC^T)^{1/2}\|^2, \tag{3.4}$$

where $M = (B^T B)^{1/2\dagger} B^T AC^T (CC^T)^{1/2\dagger}$. Note that $\|M - (B^T B)^{1/2} X (CC^T)^{1/2}\|^2$ is the only term in (3.4) that depends on X. Therefore, problem (1.1) is equivalent to

$$\min_{X\in\mathbb{R}_r^{m\times n}} \|M - (B^T B)^{1/2} X (CC^T)^{1/2}\|^2. \tag{3.5}$$

Based on Theorem 1.1, the solution of the problem (3.5) is given by

$$\widehat{X}_r = (B^T B)^{1/2\dagger} \left\lfloor P_{(B^T B)^{1/2},L} M P_{(CC^T)^{1/2},R} \right\rfloor_r (CC^T)^{1/2\dagger}. \tag{3.6}$$

Note that $(B^T B)^{1/2\dagger}$ and $(CC^T)^{1/2\dagger}$ are symmetric matrices. Therefore, it follows from Lemma 2.2 that

$$P_{(B^T B)^{1/2},L} (B^T B)^{1/2\dagger} = (B^T B)^{1/2} (B^T B)^{1/2\dagger} (B^T B)^{1/2\dagger} = (B^T B)^{1/2\dagger}, \tag{3.7}$$

and

$$(CC^T)^{1/2\dagger} P_{(CC^T)^{1/2},R} = (CC^T)^{1/2\dagger} (CC^T)^{1/2\dagger} (CC^T)^{1/2} = (CC^T)^{1/2\dagger}. \tag{3.8}$$

Further, (3.7) and (3.8) imply

$$\begin{aligned} \lfloor P_{(B^T B)^{1/2},L} M P_{(CC^T)^{1/2},R} \rfloor_r &= \lfloor P_{(B^T B)^{1/2},L} (B^T B)^{1/2\dagger} B^T AC^T (CC^T)^{1/2\dagger} P_{(CC^T)^{1/2},R} \rfloor_r \\ &= \lfloor (B^T B)^{1/2\dagger} B^T AC^T (CC^T)^{1/2\dagger} \rfloor_r \\ &= \lfloor M \rfloor_r. \end{aligned} \tag{3.9}$$

It follows from (3.6) and (3.9) that $\widehat{X}_r = (B^T B)^{1/2\dagger} \lfloor M \rfloor_r (CC^T)^{1/2\dagger}$. On the basis of Lemma 2.6, we obtain that

$$(B^T B)^{1/2} \widehat{X}_r (CC^T)^{1/2} = (B^T B)^{1/2} (B^T B)^{1/2\dagger} \lfloor M \rfloor_r (CC^T)^{1/2\dagger} (CC^T)^{1/2} = \lfloor M \rfloor_r. \tag{3.10}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 37, pp. 544-548, July 2021.

By equations (3.4) and (3.10) and the facts that $\|M\|^2 = \mathrm{tr}\{MM^T\} = \sum_{i=1}^m \lambda_i(MM^T)$ and $\|\cdot\|$ is unitary invariant [6], we obtain the following identity

$$\begin{aligned}\min_{X\in\mathbb{R}_r^{p\times q}} \|A - BXC\|^2 &= \|A\|^2 - \|M\|^2 + \|M - (B^TB)^{1/2}\widehat{X}_r(CC^T)^{1/2}\|^2 \\ &= \|A\|^2 - \|M\|^2 + \|M - \lfloor M\rfloor_r\|^2 \\ &= \|A\|^2 - \sum_{i=1}^m \lambda_i(MM^T) + \sum_{i=r+1}^m \lambda_i(MM^T) \\ &= \|A\|^2 - \sum_{i=1}^r \lambda_i(MM^T). \qquad (3.11)\end{aligned}$$

Note that $MM^T = (B^TB)^{1/2\dagger}B^TAC^\dagger CA^TB(B^TB)^{1/2\dagger}$. From Lemmas 2.1 and 2.2, we obtain that

$$\lambda_i(MM^T) = \lambda_i(B^\dagger AC^\dagger CA^TB), \qquad (3.12)$$

for all $i = 1, ..., m$. Finally, (3.3) follows from (3.11) and (3.12). □

4. Advantage and Numerical Example. The error formula given by (3.3) is useful for choosing an optimal value for the rank r, before computing the matrix $\widehat{X}_r$ in (1.2). For example, we consider matrices $A \in \mathbb{R}^{20\times 35}$, $B \in \mathbb{R}^{20\times 30}$ and $C \in \mathbb{R}^{40\times 35}$ generated from a uniform distribution with zero-mean and standard deviation 1. Figure 1 shows the relationship between the error associated with the solution of problem (1.1) and the rank r. It follows from Figure 1 that the associated error is 0 when $r \geq 20$. Therefore, the smallest value of r that implies the minimal error of the solution of problem (1.1) is given by $r = 20$. Note that it was not necessary to compute each optimal matrix $\widehat{X}_r$, for $r = 1, ..., 30$, to obtain the associated error. In this example, we only compute the eigenvalues of $T = B^\dagger AC^\dagger CA^TB \in \mathbb{R}^{30\times 30}$ and use the formula (3.3). Furthermore, in this numerical simulation, we obtain that $\mathrm{rank}(T) = 20$ and $\sigma_i^2(A) = \lambda_i(T)$, for all $i = 1, ..., 20$. Thus,

$$\min_{X\in\mathbb{R}_r^{30\times 40}} \|A - BXC\|^2 = \|A\|^2 - \sum_{i=1}^r \lambda_i(T) = 0,$$

for all $r \geq 20$.

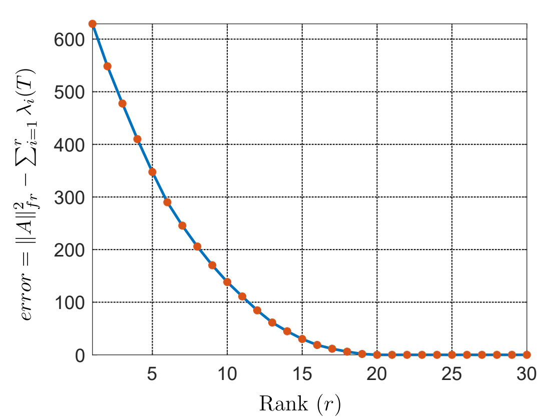

FIGURE 1. *Diagram of the error associated with the solution of problem* (1.1) *versus rank r.*

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 37, pp. 544-548, July 2021.

Acknowledgments. The author thanks Juan José Fallas Monge for his comments on this manuscript and the reviewer of the paper for their insightful comments. This work was financially supported by *Vicerrectoría de Investigación y Extensión* from *Instituto Tecnológico de Costa Rica* (Research #1440037).

REFERENCES

[1] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.
[2] D. Sondermann. Best approximate solutions to matrix equations under rank restrictions. *Statistische Hefte*, 27(1):57–66, 1986.
[3] S. Friedland and A. Torokhti. Generalized rank-constrained matrix approximations. *SIAM J. Matrix Anal. Appl.*, 29(2):656–659, 2007.
[4] A. Torokhti and S. Friedland. Towards theory of generic principal component analysis. *J. Multivar. Anal.*, 100(4):661–669, 2009.
[5] H. Wang. Rank constrained matrix best approximation problem. *Appl. Math. Lett.*, 50:98–104, 2015.
[6] F. Zhang. *Matrix Theory: Basic Results and Techniques*. Springer Science & Business Media, 2011.
[7] J.C.A. Barata and M.S. Hussein. The MoorePenrose pseudoinverse: A tutorial review of the theory. *Braz. J. Phys.*, 42(1–2):146–465, 2012.
[8] D.S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd edition. Princeton University Press, 2009.
[9] A. Torokhti and P. Howlett. *Computational Methods for Modeling of Nonlinear Systems* (Mathematics in Science and Engineering). Elsevier Science, 2007.
[10] A. Torokhti and P. Soto-Quiros. Generalized Brillinger-like transforms. *IEEE Sig. Process. Lett.*, 23(6):843–847, 2016.