Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(5): 7595–7605.
DOI: 10.3934/math.2022426
Received: 21 September 2021
Revised: 10 February 2022
Accepted: 11 February 2022
Published: 15 February 2022

Research article

Minimax perturbation bounds of the low-rank matrix under Ky Fan norm

Xinyu Qi, Jinru Wang* **and Jiating Shao**

College of Mathematics, Faculty of Science, Beijing University of Technology, Beijing, 100124, China

* **Correspondence:** Email: wangjinru@bjut.edu.cn.

Abstract: This paper considers the minimax perturbation bounds of the low-rank matrix under Ky Fan norm. We first explore the upper bounds via the best rank-r approximation $\hat{A}_r$ of the observation matrix $\hat{A}$. Next, the lower bounds are established by constructing special matrix groups to show the upper bounds are tight on the low-rank matrix estimation error. In addition, we derive the rate-optimal perturbation bounds for the left and right singular subspaces under Ky Fan norm $\sin\Theta$ distance. Finally, some simulations have been carried out to support our theories.

Keywords: singular value decomposition; Ky Fan norm; perturbation theory; $\sin\Theta$ distance; low-rank matrix estimation
Mathematics Subject Classification: 15A42, 65F55

1. Introduction

Singular value decomposition (SVD) has been widely used in statistics, machine learning, and applied mathematics. Perturbation bounds often play a critical role in the analysis of the SVD. To be more specific, let

$$\hat{A} = A + E,$$

where both A and E have the same size $d_1 \times d_2$, and A is a signal matrix which we are interested in, while E stands for a perturbation matrix. In this paper, suppose that $\hat{A}$ and A have the following singular value decompositions,

$$A = U\Sigma_r V^T + U_\perp \Sigma_{r\perp} V_\perp^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T + \sum_{i=r+1}^{d_1\wedge d_2} \sigma_i u_i v_i^T, \tag{1.1}$$

$$\hat{A} = \hat{U}\hat{\Sigma}_r \hat{V}^T + \hat{U}_\perp \hat{\Sigma}_{r\perp} \hat{V}_\perp^T = \sum_{i=1}^{r} \hat{\sigma}_i \hat{u}_i \hat{v}_i^T + \sum_{i=r+1}^{d_1\wedge d_2} \hat{\sigma}_i \hat{u}_i \hat{v}_i^T, \tag{1.2}$$

where $r \leq \text{rank}(A)$, $d_1 \wedge d_2$ stands for $\min\{d_1, d_2\}$. The singular values σ_i and $\hat{\sigma}_i$ are in the decreasing order. $U = [u_1, \ldots, u_r]$, $\hat{U} = [\hat{u}_1, \ldots, \hat{u}_r] \in \mathbb{O}_{d_1,r}$ (the set of all $d_1 \times r$ orthonormal columns and $\mathbb{O}_{d_1}:=\mathbb{O}_{d_1,d_1}$), and $V = [v_1, \ldots, v_r]$, $\hat{V} = [\hat{v}_1, \ldots, \hat{v}_r] \in \mathbb{O}_{d_2,r}$. Unlike compressed sensing [5] to reconstruct the original signal, our goal is to estimate the underlying low-rank matrix A and its leading left and right singular matrices U, V.

The problems to estimate U, V have been widely studied in the literature [1, 3, 4, 10, 12]. Among these results, Davis and Kahan [3], Wedin [12] established the fundamental methods for matrix perturbation theory; Vu [10], Wang [11] discussed the rotations of singular vectors after random perturbation; Cai and Zhang [1] studied the rate-optimal perturbation bounds for singular subspaces; Fan et al. [4] gave an eigenvector perturbation bound and the robust covariance estimation. In addition, Luo et al. [6] considered the perturbation bound under Schatten-q norm. Till now, a few of the existing works focused on the perturbation analysis of the matrix A itself. This paper will consider the estimation of rank-r matrix A under Ky Fan norm which extends the results of Luo et al.

For a given $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$, the Ky Fan norm $\|M\|_{(k)}$ of the matrix $M \in \mathbb{R}^{d_1 \times d_2}$ is given by $\|M\|_{(k)} = \sum_{i=1}^{k} \sigma_i(M)$. Clearly, $\|\cdot\|_{(k)}$ is a unitarily invariant norm.

In this paper, we consider the estimation of rank-r matrix A (i.e., $\Sigma_{r\perp} = 0$) via rank-r truncated SVD $\hat{A}_r := \hat{U}\hat{\Sigma}_r\hat{V}^T$ of $\hat{A}$. It is widely known that $\hat{A}_r$ is the best rank-r approximation of $\hat{A}$. Here and throughout, A_l or $(A)_l$ denotes the best rank-l approximation of the matrix A.

Firstly, we establish the following upper bound.

Theorem 1.1. *Let the observation matrix $\hat{A} = A + E \in \mathbb{R}^{d_1 \times d_2}$, where A is an unknown rank-r matrix and E is the perturbation matrix. Then*

$$\|\hat{A}_r - A\|_{(k)} \leq 3\|E_r\|_{(k)}, \quad k = 1, 2, \cdots, d_1 \wedge d_2,$$

where E_r denotes the best rank-r approximation of the matrix E.

Remark 1.1. *According to Eckart-Young-Mirsky Theorem and rank$(A) = r$, we have $\|\hat{A}_r - \hat{A}\|_{(k)} \leq \|A - \hat{A}\|_{(k)}$. Therefore,*

$$\|\hat{A}_r - A\|_{(k)} \leq \|\hat{A}_r - \hat{A}\|_{(k)} + \|\hat{A} - A\|_{(k)} \leq 2\|\hat{A} - A\|_{(k)} = 2\|E\|_{(k)}. \tag{1.3}$$

If $r \ll d_1 \wedge d_2$, then $\|E_r\|_{(k)}$ can be much smaller than $\|E\|_{(k)}$ for any $k \gg r$.

Remark 1.2. *If $k = d_1 \wedge d_2$, both the Ky Fan norm and the Schatten-1 norm are equal to the nuclear norm; If $k = 1$, both the Ky Fan norm and the Schatten-∞ norm are equal to the spectral norm. Otherwise, the two norms are not included each other. Therefore, our results can be regarded as a supplement to the existing results.*

Before stating the lower bound, for any $t > 0$, we define the class of (A, E) as

$$\mathcal{F}_r(t) = \{(A, E) : \ \text{rank}(A) = r, \|E\|_{(k)} \leq t\}. \tag{1.4}$$

Here $A, E \in \mathbb{R}^{d_1 \times d_2}$ and $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$.

Theorem 1.2. *For the low-rank perturbation model $\hat{A} = A + E \in \mathbb{R}^{d_1 \times d_2}$, if $r \leq \frac{1}{2}(d_1 \wedge d_2)$, then for any estimator $\tilde{A}$ based on the observation matrix $A + E$,*

$$\inf_{\tilde{A}} \sup_{(A,E)\in\mathcal{F}_r(t)} \|\tilde{A} - A\|_{(k)} \geq \frac{t}{2},$$

where $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$.

Theorem 1.2 shows that the upper bound given in Theorem 1.1 is sharp for the rank-r truncated singular value decomposition estimator $\hat{A}_r$.

The principle angle $\Theta(V_1, V_2)$ of the matrices $V_1, V_2 \in \mathbb{O}_{d,r}$ means the diagonal matrix

$$\Theta(V_1, V_2) = \mathrm{diag}\{\cos^{-1}(\sigma_1), \cos^{-1}(\sigma_2), \cdots, \cos^{-1}(\sigma_r)\}$$

with the singular values $\sigma_i := \sigma_i(V_1^T V_2)$ of $V_1^T V_2$ satisfying $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. When $r = 1$, $\Theta(V_1, V_2)$ coincides with the angle of two d dimensional unit vectors. In this paper, the $\sin\Theta$ distance is used to measure the difference between V_1 and V_2. i.e.,

$$\|\sin\Theta(V_1, V_2)\|_{(k)} = \|\mathrm{diag}\{\sin\cos^{-1}\sigma_1, \ldots, \sin\cos^{-1}\sigma_r\}\|_{(k)} = \sum_{i=1}^{k}(1 - \sigma_i^2)^{1/2}.$$

Indeed, although $\|\sin\Theta(V_1, V_2)\|$ defines a semi-metric on $\mathbb{O}_{d,r}$, it is also satisfied

$$\|\sin\Theta(V_1, V_2)\|_{(k)} \leq \|\sin\Theta(V_1, V_3)\|_{(k)} + \|\sin\Theta(V_3, V_2)\|_{(k)} \tag{1.5}$$

and

$$\|\sin\Theta(V_1, V_2)\|_{(k)} = \|V_{2\perp}^T V_1\|_{(k)} \tag{1.6}$$

following from [7].

As a byproduct of Theorem 1.1, we can derive the perturbation bounds for the leading singular subspaces U and V under Ky Fan norm $\sin\Theta$ distance. i.e.,

$$\|\sin\Theta(\hat{U}, U)\|_{(k)} \leq \frac{2\|E_r\|_{(k)}}{\sigma_r(A)},$$
$$\|\sin\Theta(\hat{V}, V)\|_{(k)} \leq \frac{2\|E_r\|_{(k)}}{\sigma_r(A)}.$$

Furthermore, we also give the corresponding lower bounds to show the above upper bounds are sharp.

2. Proofs

Firstly, let us introduce some lemmas in order to prove Theorem 1.1.

2.1. Auxiliary lemmas

A function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called a symmetric gauge function ([9]) if (1) $\mathbf{x} \neq 0 \Longrightarrow \Phi(\mathbf{x}) > 0$; (2) $\Phi(\alpha\mathbf{x}) = |\alpha|\Phi(\mathbf{x})$ for $\alpha \in \mathbb{R}$; (3) $\Phi(\mathbf{x}+\mathbf{y}) \leq \Phi(\mathbf{x}) + \Phi(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, and (4) $\Phi(J\mathbf{x}_\pi) = \Phi(\mathbf{x})$, where J is any diagonal matrix whose diagonal elements are 1 or -1, and π is any permutation with $1, \ldots, d$.

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, define the function

$$\Psi(\mathbf{y}) := \sup_{\Phi(\mathbf{x})=1} \langle \mathbf{y}, \mathbf{x} \rangle.$$

It is easy to check $\Psi(\cdot)$ is also a symmetric gauge function. In general, $\Psi(\mathbf{y})$ is usually called the dual symmetric gauge function of $\Phi(\mathbf{x})$. In particular, for a matrix $A \in \mathbb{R}^{d_1 \times d_2}$, we can define

$$\Phi(A) := \Phi(\sigma_1, \ldots, \sigma_{d_1 \wedge d_2}),$$

where $\sigma_1, \ldots, \sigma_{d_1 \wedge d_2}$ are the singular values of A, then the following lemma is Lemma 3.4 in [9].

Lemma 2.1. *Let $A, B \in \mathbb{R}^{d_1 \times d_2}$ and their singular values are $\sigma_1 \geq \cdots \geq \sigma_{d_1 \wedge d_2} \geq 0$, $\xi_1 \geq \cdots \geq \xi_{d_1 \wedge d_2} \geq 0$ respectively. Then*

$$\max_{U \in \mathbb{O}_{d_1}, V \in \mathbb{O}_{d_2}} \operatorname{tr}(UAVB^T) = \sum_{i=1}^{d_1 \wedge d_2} \sigma_i \xi_i. \tag{2.1}$$

According to Lemma 2.1, we introduce a dual characterization lemma.

Lemma 2.2. *Let $A \in \mathbb{R}^{d_1 \times d_2}$, there exists a symmetric gauge function $\Psi_k(\cdot)$ such that*

$$\|A_r\|_{(k)} = \sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \operatorname{tr}(X^T A) \tag{2.2}$$

for $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$. In special case, if $\operatorname{rank}(A) \leq r$, *then*

$$\|A\|_{(k)} = \sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \operatorname{tr}(X^T A). \tag{2.3}$$

Proof. For any $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$, define

$$\Phi_k(A) = \Phi_k(\sigma_1, \sigma_2, \cdots, \sigma_{d_1 \wedge d_2}) := \sum_{i=1}^{k} \sigma_i,$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{d_1 \wedge d_2} \geq 0$ are the singular values of A. Clearly, $\Phi_k(A)$ is a symmetric gauge function and $\Phi_k(A) = \|A\|_{(k)}$. Furthermore, denote Ψ_k the dual symmetric gauge function of Φ_k, then for any $U \in \mathbb{O}_{d_1}, V \in \mathbb{O}_{d_2}$, we have $\Psi_k(U^T X V^T) = \Psi_k(X)$ and

$$\begin{aligned}
&\sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \operatorname{tr}(X^T A) = \sup_{\Psi_k(U^T X V^T)=1, \operatorname{rank}(\mathrm{U^T X V^T}) \leq \mathrm{r}} \operatorname{tr}(V X^T U A) \\
= &\sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \operatorname{tr}(V X^T U A) = \sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \max_{U \in \mathbb{O}_{d_1}, V \in \mathbb{O}_{d_2}} \operatorname{tr}(V X^T U A).
\end{aligned}$$

This along with Lemma 2.1 shows that

$$\begin{aligned}
&\sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \operatorname{tr}(X^T A) = \sup_{\Psi_k(X)=1, \operatorname{rank(X)} \leq \mathrm{r}} \sum_{i=1}^{d_1 \wedge d_2} \sigma_i \xi_i = \sup_{\Psi_k(X)=1} \sum_{i=1}^{r} \sigma_i \xi_i \\
&= \Phi_k(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) = \Phi_k(A_r) = \|A_r\|_{(k)},
\end{aligned}$$

where $\xi_1 \geq \cdots \geq \xi_{d_1 \wedge d_2} \geq 0$ are the singular values of X. □

For any $U \in \mathbb{O}_{d,r}$, $P_U = UU^T$ is the projection matrix onto the column span of U. The next technical lemma is useful in the proof of Theorem 1.1.

Lemma 2.3. *Let* $\hat{A} = A + E \in \mathbb{R}^{d_1 \times d_2}$, rank$(A) = r$, *and (1.2) holds. Then for any* $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$,

$$\max\left\{\|P_{\hat{U}_\perp} A\|_{(k)},\ \|A P_{\hat{V}_\perp}\|_{(k)}\right\} \le 2\|E_r\|_{(k)}.$$

Proof. Since rank$(P_{\hat{U}_\perp} A) \le$ rank$(A) = r$, and (2.3) of Lemma 2.2 are satisfied, we have

$$\begin{aligned}
\|P_{\hat{U}_\perp} A\|_{(k)} &= \sup_{\Psi_k(X)=1,\mathrm{rank(X)}\le \mathrm{r}} \mathrm{tr}\Big[X^T (P_{\hat{U}_\perp} A)\Big] \\
&= \sup_{\Psi_k(X)=1,\mathrm{rank(X)}\le \mathrm{r}} \mathrm{tr}\Big[X^T (P_{\hat{U}_\perp} \hat{A} - P_{\hat{U}_\perp} E)\Big] \\
&\le \sup_{\Psi_k(X)=1,\mathrm{rank(X)}\le \mathrm{r}} \mathrm{tr}\Big[X^T (P_{\hat{U}_\perp} \hat{A})\Big] + \sup_{\Psi_k(X)=1,\mathrm{rank(X)}\le \mathrm{r}} \mathrm{tr}\Big[X^T (P_{\hat{U}_\perp} E)\Big].
\end{aligned}$$

According to Lemma 2.2 and (2.2),

$$\|P_{\hat{U}_\perp} A\|_{(k)} \le \left\|\left(P_{\hat{U}_\perp} \hat{A}\right)_r\right\|_{(k)} + \left\|\left(P_{\hat{U}_\perp} E\right)_r\right\|_{(k)}. \tag{2.4}$$

In addition, $\left\|\left(P_{\hat{U}_\perp} \hat{A}\right)_r\right\|_{(k)} = \left\|\left(\hat{A} - \hat{A}_r\right)_r\right\|_{(k)}$ due to $P_{\hat{U}} \hat{A} = \hat{A}_r$. On the other hand, based on Theorem 2 in [8] and the fact that the norm $\|(\cdot)_r\|_{(k)}$ is unitarily invariant, we have

$$\|(A - A_r)_l\|_{(k)} = \inf_{M \in \mathbb{R}^{d_1 \times d_2}, \mathrm{rank}(M) \le r} \|(A - M)_l\|_{(k)}.$$

Therefore,

$$\begin{aligned}
\left\|\left(P_{\hat{U}_\perp} \hat{A}\right)_r\right\|_{(k)} &= \inf_{\mathrm{rank}(M) \le r} \|(\hat{A} - M)_r\|_{(k)} \\
&\le \|(\hat{A} - P_U \hat{A})_r\|_{(k)} = \left\|\left(P_{U_\perp} E\right)_r\right\|_{(k)}.
\end{aligned}$$

For two matrices $B, C \in \mathbb{R}^{d_1 \times d_2}$, it is known that

$$\sigma_{i+j-1}(BC^T) \le \sigma_i(B) \cdot \sigma_j(C). \tag{2.5}$$

Thus, $\sigma_i(P_{U_\perp} E) \le \sigma_1(P_{U_\perp}) \sigma_i(E) = \sigma_i(E)$ and $\sigma_i(P_{\hat{U}_\perp} E) \le \sigma_i(E)$. Hence, by (2.4),

$$\|P_{\hat{U}_\perp} A\|_{(k)} \le 2\|E_r\|_{(k)}.$$

Similarly, $\|A P_{\hat{V}_\perp}\|_{(k)} \le 2\|E_r\|_{(k)}$. This completes the proof of Lemma 2.3. □

2.2. *Proof of Theorem 1.1*

Now, we are in the position to prove Theorem 1.1.

Proof. By (1.2), we know that $\hat{U}$ is composed of the first r left singular vectors of $\hat{A}$. Thus, $\hat{A}_r = P_{\hat{U}} \hat{A}$. For any $k \in \{1, 2, \ldots, d_1 \wedge d_2\}$,

$$\begin{aligned}
\|\hat{A}_r - A\|_{(k)} &= \|P_{\hat{U}} \hat{A} - (P_{\hat{U}} + P_{\hat{U}_\perp}) A\|_{(k)} \\
&= \|P_{\hat{U}} E - P_{\hat{U}_\perp} A\|_{(k)} \le \|P_{\hat{U}} E\|_{(k)} + \|P_{\hat{U}_\perp} A\|_{(k)}.
\end{aligned}$$

This with (2.5) and Lemma 2.3 derives

$$\|\hat{A}_r - A\|_{(k)} \le 3\|E_r\|_{(k)}.$$

The proof of Theorem 1.1 is complete. □

2.3. Proof of Theorem 1.2

Proof. First, for any $k \le r$, define $A_i, E_i \in \mathbb{R}^{d_1 \times d_2}$ $(i = 1, 2)$ with

$$A_1 = \begin{pmatrix} \frac{t}{k}\mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}, \quad E_1 = \begin{pmatrix} \mathbf{0}_{r\times r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{t}{k}\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r}, \end{pmatrix}$$
$$A_2 = \begin{pmatrix} \mathbf{0}_{r\times r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{t}{k}\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r}, \end{pmatrix}, \quad E_2 = \begin{pmatrix} \frac{t}{k}\mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix},$$

then we have $A_1 + E_1 = A_2 + E_2 = \hat{A}$ and $\text{rank}(A_1) = \text{rank}(A_2) = r$. Where $\text{rank}(A_1) = \text{rank}(A_2) = r$ and $\left\|\left(E_1\right)_r\right\|_{(k)} = \left\|\left(E_2\right)_r\right\|_{(k)} = \frac{k}{k}t = t$. Therefore, $(A_1, E_1), (A_2, E_2) \in \mathcal{F}_r(t)$.

For any estimator $\tilde{A}$ of A, one derives

$$\begin{aligned} \inf_{\tilde{A}} \sup_{(A,E)\in\mathcal{F}_r(t)} \|\tilde{A} - A\|_{(k)} &\ge \inf_{\tilde{A}} \left(\max\left\{\|\tilde{A} - A_1\|_{(k)}, \|\tilde{A} - A_2\|_{(k)}\right\}\right) \\ &\ge \inf_{\tilde{A}} \frac{1}{2}\left(\|\tilde{A} - A_1\|_{(k)} + \|\tilde{A} - A_2\|_{(k)}\right) \\ &\ge \inf_{\tilde{A}} \frac{1}{2}\|A_1 - A_2\|_{(k)} = \frac{t}{2}. \end{aligned} \tag{2.6}$$

Next to show Theorem 1.2 is established for $k > r$. One takes

$$A_1 = \begin{pmatrix} \frac{t}{r}\mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}, \quad E_1 = \begin{pmatrix} \mathbf{0}_{r\times r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{t}{r}\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix};$$
$$A_2 = \begin{pmatrix} \mathbf{0}_{r\times r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{t}{r}\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r}, \end{pmatrix}, \quad E_2 = \begin{pmatrix} \frac{t}{r}\mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}.$$

Then $A_1 + E_1 = A_2 + E_2 = \hat{A}$, $\text{rank}(A_1) = \text{rank}(A_2) = r$ and $\left\|\left(E_1\right)_r\right\|_{(k)} = \left\|\left(E_2\right)_r\right\|_{(k)} = t$. Therefore, $(A_1, E_1), (A_2, E_2) \in \mathcal{F}_r(t)$. We can use similar processes to prove (2.6). i.e.,

$$\inf_{\tilde{A}} \sup_{(A,E)\in\mathcal{F}_r(t)} \|\tilde{A} - A\|_{(k)} \ge \frac{t}{2}.$$

Theorem 1.2 is finished.

□

3. Perturbation bounds of singular subspaces

As a byproduct of the perturbation theory, this paper derives $\sin\Theta$ perturbation bounds of the left and right subspaces U, V under Ky Fan norm.

3.1. Upper bounds

Theorem 3.1. *Let* $\hat{A} = A + E \in \mathbb{R}^{d_1 \times d_2}$, $\operatorname{rank}(A) = r$. *If the singular value decompositions (1.1) and (1.2) hold, then*

$$\|\sin\Theta(\hat{U}, U)\|_{(k)} \le \frac{2\|E_r\|_{(k)}}{\sigma_r(A)},$$
$$\|\sin\Theta(\hat{V}, V)\|_{(k)} \le \frac{2\|E_r\|_{(k)}}{\sigma_r(A)}$$

for any $k \in \{1, 2, \dots, d_1 \wedge d_2\}$.

Proof. By Theorem 3.9 (II) in [9], one knows $\|BC^T\|_{(k)} \ge \|B\|_{(k)}\sigma_{d_1\wedge d_2}(C)$ for any two matrices $B, C \in \mathbb{R}^{d_1\times d_2}$. This with (1.6) shows

$$\|\sin\Theta(\hat{U}, U)\|_{(k)} = \|\hat{U}_\perp^T U\|_{(k)} \le \frac{\|\hat{U}_\perp^T U U^T A\|_{(k)}}{\sigma_r(U^T A)}.$$

According to (1.1) and $\operatorname{rank}(A) = r$, one has $UU^TA = A$ and $\sigma_r(U^TA) = \sigma_r(A)$. Thus

$$\|\sin\Theta(\hat{U}, U)\|_{(k)} \le \frac{\|\hat{U}_\perp^T A\|_{(k)}}{\sigma_r(A)} \le \frac{2\|E_r\|_{(k)}}{\sigma_r(A)}$$

thanks to Lemma 2.3. Similarly, one also can get $\|\sin\Theta(\hat{V}, V)\|_{(k)} \le \frac{2\|E_r\|_{(k)}}{\sigma_r(A)}$. We have concluded the proof of Theorem 3.1. □

3.2. Lower bounds

Theorem 3.2. *For* $k \in \{1, 2, \dots, d_1 \wedge d_2\}$, *define the following class*

$$\mathcal{F}_r(\alpha, \beta) = \left\{(A, E) : \operatorname{rank}(A) = r, \sigma_r(A) \ge \alpha, \|E\|_{(k)} \le \beta\right\}.$$

If $r \le \frac{1}{2}(d_1 \wedge d_2)$ *and* $\alpha(k \wedge r) \ge \beta$, *then for any estimators* $\tilde{U}$ *and* $\tilde{V}$ *based on the observation matrix* $A + E$, *we have*

$$\inf_{\tilde{U}} \sup_{(A,E)\in\mathcal{F}_r(\alpha,\beta)} \|\sin\Theta(\tilde{U}, U)\|_{(k)} \ge \frac{1}{2\sqrt{10}}\frac{\beta}{\alpha}, \tag{3.1}$$
$$\inf_{\tilde{V}} \sup_{(A,E)\in\mathcal{F}_r(\alpha,\beta)} \|\sin\Theta(\tilde{V}, V)\|_{(k)} \ge \frac{1}{2\sqrt{10}}\frac{\beta}{\alpha}. \tag{3.2}$$

Proof. We only need to show (3.2) since the statement (3.1) can be gotten by similar process. First, we introduce the following singular value decomposition,

$$\begin{pmatrix} \alpha & \frac{\beta}{k\wedge r} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}^T$$
$$= \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix} \sigma_1 \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix}^T,$$

then by Lemma 3 in [2] and $\alpha(k \wedge r) \geq \beta$, we know

$$|v_{21}| \geq \frac{1}{\sqrt{10}(k \wedge r)} \frac{\beta}{\alpha}. \tag{3.3}$$

Second, based on the above matrix, the following matrices are constructed.

$$A_1 = \begin{pmatrix} \sigma_1 u_{11} v_{11} \mathbf{I}_r & \sigma_1 u_{11} v_{21} \mathbf{I}_r & \mathbf{0} \\ \sigma_1 u_{21} v_{11} \mathbf{I}_r & \sigma_1 u_{21} v_{21} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}, \quad E_1 = \mathbf{0}_{d_1,d_2};$$

$$A_2 = \begin{pmatrix} \alpha \mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r,} \end{pmatrix}, \quad E_2 = \begin{pmatrix} \mathbf{0} & \frac{\beta}{k\wedge r} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}.$$

Obviously, $\text{rank}(A_1) = \text{rank}(A_2) = r$ and

$$\hat{A} = A_1 + E_1 = A_2 + E_2 = \begin{pmatrix} \alpha \mathbf{I}_r & \frac{\beta}{k\wedge r} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{d_1-2r, d_2-2r} \end{pmatrix}.$$

On the other hand, It is easy to check $\sigma_r(A_1) = \sigma_1(A_1) \geq \alpha, \left\|\left(E_1\right)_r\right\|_{(k)} = 0 \leq \beta$ and $\sigma_r(A_2) = \alpha, \left\|\left(E_1\right)_r\right\|_{(k)} = \frac{k\wedge r}{k\wedge r}\beta \leq \beta$. Hence, $(A_1, E_1), (A_2, E_2) \in \mathcal{F}_r(\alpha, \beta)$. Let V_1, V_2 are the leading r singular vector of A_1, A_2 respectively, then

$$V_1 = \begin{pmatrix} v_{11} \mathbf{I}_r \\ v_{21} \mathbf{I}_r \\ \mathbf{0}_{d_2-2r} \end{pmatrix}, \quad V_2 = \begin{pmatrix} \mathbf{I}_r \\ \mathbf{0}_r \\ \mathbf{0}_{d_2-2r} \end{pmatrix}$$

follow from the structure of A_1, A_2, Therefore, for any estimator $\tilde{V}$ of the leading r right singular space, we have

$$\begin{aligned} &\inf_{\tilde{V}} \sup_{(A,E)\in\mathcal{F}_r(\alpha,\beta)} \|\sin\Theta(\tilde{V}, V)\|_{(k)} \\ \geq &\inf_{\tilde{V}} \max\left\{\|\sin\Theta(\tilde{V}, V_1)\|_{(k)}, \|\sin\Theta(\tilde{V}, V_2)\|_{(k)}\right\} \\ \geq &\frac{1}{2}\Big(\|\sin\Theta(\tilde{V}, V_1)\|_{(k)} + \|\sin\Theta(\tilde{V}, V_2)\|_{(k)}\Big) \\ \overset{(1.5)}{\geq} &\frac{1}{2}\|\sin\Theta(V_1, V_2)\|_{(k)} \overset{(1.6)}{=} \frac{1}{2}\|v_{21}\mathbf{I}_r\|_{(k)} \\ = &\frac{1}{2}(k \wedge r)|v_{21}| \overset{(3.3)}{\geq} \frac{1}{2\sqrt{10}} \frac{\beta}{\alpha}. \end{aligned}$$

The proof of Theorem 3.2 is finished. □

Remark 3.1. *In Theorem 3.2, the assumption $\alpha(k \wedge r) \geq \beta$ is necessary to obtain a consistent estimator. In fact, if $\alpha(k \wedge r) < \beta$, there is no stable algorithm to recover either U or V in the sense that there exists uniform constant $\frac{1}{2\sqrt{2}}$ such that*

$$\inf_{\tilde{U}} \sup_{(A,E)\in\mathcal{F}_r(\alpha,\beta)} \|\sin\Theta(\tilde{U}, U)\|_{(k)} \geq \frac{1}{2\sqrt{2}},$$

$$\inf_{\tilde{V}} \sup_{(A,E)\in\mathcal{F}_r(\alpha,\beta)} \|\sin\Theta(\tilde{V},V)\|_{(k)} \geq \frac{1}{2\sqrt{2}}.$$

Proof. Let

$$\begin{pmatrix} \alpha & \frac{\beta}{k\wedge r} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix} \sigma_1 \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix}^T,$$

then by Lemma 3 in [2] and $\alpha(k\wedge r) < \beta$, we know $|v_{21}| \geq \frac{1}{\sqrt{2}}$. Therefore, based on the similar discussion of the proof of Theorem 3.2, Remark 3.1 is established. □

Remark 3.2. *By Theorem 3.2, we can know that the rates given in Theorem 3.1 are optimal, but the corresponding lower bounds for the singular subspaces are not given in Luo et al. [6].*

4. Numerical simulations

In this section, we provide some numerical studies to support our theoretical results. Throughout the simulation studies, we consider the nuclear norm $\|\cdot\|_*$ (the sum of all singular values) as the error metric. i.e., $k = d_1 \wedge d_2$. Without loss of generality, we assume $d_1 = d_2 := d$. In each setting, we randomly generate the perturbation $E = uv^T + Z \in \mathbb{R}^{d\times d}$, where $u, v \in \mathbb{R}^d$ are randomly generated unit vectors and Z has independent identically distributed $N(0,\sigma)$ entries. On the other hand, we generate low-rank matrix $A = U\Sigma_r V^T$ by a special structure. Here $U, V \in \mathbb{R}^{d\times r}$ are independently drawn from $\mathbb{O}_{d,r}$ uniformly at random; Σ_r is a $r\times r$ diagonal matrix with singular values decaying polynomially as $(\Sigma_r)_{ii} = \frac{10}{i}$, $1 \leq i \leq r$. Each simulation setting is repeated for 100 times and the average values are reported. The Figure 1 is the result of numerical studies.

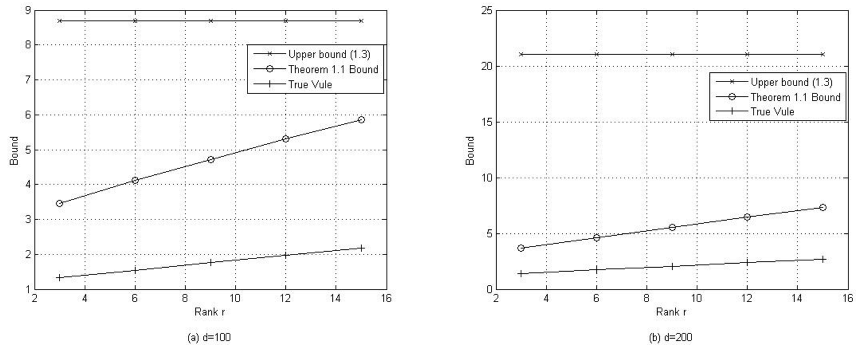

Figure 1. Theorem 1.1 bound, upper bound (1.3) and the true value of $\|\hat{A}_r - A\|_*$

We set $d \in \{100, 200\}$, $r \in \{3, 6, 9, 12, 15\}$, $\sigma = 0.004$. The results of the upper bounds in Theorem 1.1, (1.3) and the true value of $\|\hat{A}_r - A\|_*$ are given in Figure 1. It shows that the upper bound in Theorem 1.1 is tighter than the upper bound in (1.3) in all setting. Furthermore, the upper bound of Theorem 1.1 remains steady while the upper bound of (1.3) significantly increases when d increases form 100 to 200.

5. Conclusions

In this paper, we give a sharp upper bound for rank-r matrix A under Ky Fan norm, and show that it is optimal by establishing the corresponding lower bound. As a byproduct, we provide the perturbation bounds for the singular subspaces under Ky Fan norm $\sin\Theta$ distance. Furthermore, we give the corresponding lower bound to show its optimality. Finally, we provide numerical studies to support our theoretical results.

As a unitarily invariant norm, Ky Fan norm which is different from Schatten-q norm is also an important matrix norm. So it makes sense to study the perturbation bound for the low-rank matrix. It is worth mentioning that the approach of proving Lemma 2.3 can be generalized any unitarily invariant norm. Therefore, it can be used to study other perturbation theory in future.

Acknowledgments

The authors would like to express their gratitude to the editor and anonymous referees for their constructive and valuable suggestions which improved the paper. This work is supported by the National Natural Science Foundation of China (No. 11771030 and 12171016).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. T. T. Cai, A. R. Zhang, Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics, *Ann. Statist.*, **46** (2018), 60–89. https://doi.org/10.1214/17-AOS1541
2. T. T. Cai, A. R. Zhang, Supplement to "Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics", 2018. https://doi.org/10.1214/17-AOS1541SUPP
3. C. Davis, W. M. Kahan, The rotation of eigenvectors by a perturbation, *SIAM J. Numer. Anal.*, **7** (1970), 1–46. https://doi.org/10.1137/0707001
4. J. Fan, W. Wang, Y. Zhong, An l_∞ eigenvector perturbation bound and its applicabtion to robust covariance estimation, *J. Mach. Learn. Res.*, **18** (2018), 1–42.
5. J. W. Huang, J. J. Wang, F. Zhang, H. L. Wang, W. D. Wang, Perturbation analysis of low-rank matrix stable recovery, *Int. J. Wavelets Multi.*, **19** (2021), 2050091. https://doi.org/10.1142/S0219691320500915
6. Y. T. Luo, R. G. Han, A. R. Zhang, A Schatten-q low-rank matrix perturbation analysis via perturbation projection error bound, *Linear Algebra Appl.*, **630** (2021), 225–240. https://doi.org/10.1016/j.laa.2021.08.005
7. Y. M. Liu, C. G. Ren, An optimal perturbation bound, *Math. Method. Appl. Sci.*, **42** (2019), 3791–3798. https://doi.org/10.1002/mma.5612

8. L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Q. J. Math.*, **11** (1960), 50–59. https://doi.org/10.1093/qmath/11.1.50
9. G. W. Stewart, J. G. Sun, *Matrix perturbation theory*, New York: Academic Press, 1990.
10. V. Vu, Singular vectors under random perturbation, *Random Struct. Algor.*, **39** (2011), 526–538. https://doi.org/10.1002/rsa.20367
11. R. R. Wang, Singular vector perturbation under Gaussian noise, *SIAM J. Matrix Anal. Appl.*, **36** (2015), 158–177. https://doi.org/10.1137/130938177
12. P. A. Wedin, Perturbation bounds in connection with singular value decomposition, *BIT*, **12** (1972), 99–111. https://doi.org/10.1007/BF01932678