

RESEARCH

Open Access



# Lower bounds for the low-rank matrix approximation

Jicheng Li<sup>1</sup>, Zisheng Liu<sup>1,2\*</sup> and Guo Li<sup>3</sup>

\*Correspondence:

liuzisheng0710@163.com

<sup>1</sup>School of Mathematics and Statistics, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, 710049, China

<sup>2</sup>School of Statistics, Henan University of Economics and Law, No. 180, Jinshui East Road, Zhengzhou, 450046, China

Full list of author information is available at the end of the article

## Abstract

Low-rank matrix recovery is an active topic drawing the attention of many researchers. It addresses the problem of approximating the observed data matrix by an unknown low-rank matrix. Suppose that  $A$  is a low-rank matrix approximation of  $D$ , where  $D$  and  $A$  are  $m \times n$  matrices. Based on a useful decomposition of  $D^\dagger - A^\dagger$ , for the unitarily invariant norm  $\|\cdot\|$ , when  $\|D\| \geq \|A\|$  and  $\|D\| \leq \|A\|$ , two sharp lower bounds of  $D - A$  are derived respectively. The presented simulations and applications demonstrate our results when the approximation matrix  $A$  is low-rank and the perturbation matrix is sparse.

**MSC:** 15A23; 34D10; 68W25; 90C25; 90C59

**Keywords:** low-rank matrix; approximation; error estimation; pseudo-inverse; matrix norms

## 1 Introduction

In mathematics, low-rank approximation is a minimization problem, in which the cost function measures the fit between a given matrix (the data) and an approximating matrix (the optimization variable), subject to a constraint that the approximating matrix has reduced rank. The problem is used for mathematical modeling and data compression. The rank constraint is related to a constraint on the complexity of a model that fits the data.

Low-rank approximation of a linear operator is ubiquitous in applied mathematics, scientific computing, numerical analysis, and a number of other areas. For example, a low-rank matrix could correspond to a low-degree statistical model for a random process (e.g., factor analysis), a low-order realization of a linear system [1], or a low-dimensional embedding of data in the Euclidean space [2], the image and computer vision [3–5], bioinformatics, background modeling and face recognition [6], latent semantic indexing [7, 8], machine learning [9–12] and control [13] *etc.* These data may have thousands or even billions of dimensions, and a large number of samples may have the same or similar structure. As we know, the important information lies in some low-dimensional subspace or low-dimensional manifold, but interfered with some perturbative components (sometimes interfered by the sparse component).

Let  $D \in \mathbb{R}^{m \times n}$  be an observed data matrix which is combined as

$$D = A + E, \tag{1}$$

where  $A \in \mathbb{R}^{m \times n}$  is the low-rank component and  $E \in \mathbb{R}^{m \times n}$  is the perturbation component of  $D$ . The singular value decomposition (SVD [14]) is a method for dealing with such high-dimensional data. If the matrix  $E$  is small, the classical principal components analysis (PCA [15–17]) can seek the best rank- $r$  estimation of  $A$  by solving the following constrained optimization via SVD of  $D$  and then projecting the columns of  $D$  onto the subspace spanned by the  $r$  principal left singular vectors of  $D$ :

$$\begin{aligned} \min_E \|E\|_F \\ \text{s.t. } \text{Rank}(A) \leq r, \\ \|D - A\|_F \leq \epsilon, \end{aligned} \quad (2)$$

where  $r \ll \min(m, n)$  is the target dimension of the subspace,  $\epsilon$  is an upper bound on the perturbative component  $\|E\|_F$  and  $\|\cdot\|_F$  is the Frobenius norm.

Despite its many advantages, the traditional PCA suffers from the fact that the estimation  $\hat{A}$  obtained by classical PCA can be arbitrarily far from the true  $A$ , when  $E$  is sufficiently sparse (relative to the rank of  $A$ ). The reason for this poor performance is precisely that the traditional PCA makes sense for Gaussian noise and not for sparse noise. Recently, robust PCA (RPCA [6]) is a family of methods that aims to make PCA robust to large errors and outliers. That is, RPCA is an upgrade of PCA.

There are some reasons for the study of lower bound of a low-rank matrix approximation problem. Firstly, as far as we know, there is no literature to consider the lower bound of the low-rank matrix approximation problem. In our paper, we first put forward the lower bound. Secondly, for the low-rank approximation, when a perturbation  $E$  exists, there is an approximation error which cannot be avoided, that is, the approximation error cannot equal 0, but tends to 0. Thirdly, from our main results, we can clearly find the influence of the spectral norm ( $\|\cdot\|_2$ ) on the low-rank matrix approximation. For example, for our main result of Case II, when the maximum eigenvalue of the matrix  $D$  is larger, the approximation error of  $(D - A)$  is smaller. In addition, the lower bound can verify whether the solution obtained by algorithms is optimal. For details, please refer to the experiments Section 4 of our paper. Therefore, it is necessary and significant to study the lower bound of the low-rank matrix approximation problem.

**Remark 1.1** PCA and RPCA are methods for the low-rank approximation problem when perturbation item exists. Our aim is to prove that no matter what method is used, the lower bound of error always exists and it cannot be avoided with the perturbation item  $E$ . Considering the existence of error, this paper focuses on the specific situation of this lower bound.

### 1.1 Notations

For a matrix  $A \in \mathbb{R}^{m \times n}$ , let  $\|A\|_2$  and  $\|A\|_*$  denote the spectral norm and the nuclear norm (i.e., the sum of its singular values), respectively. Let  $\|\cdot\|$  be a unitarily invariant norm. The pseudo-inverse and the conjugate transpose of  $A$  are denoted by  $A^\dagger$  and  $A^H$ , respectively. We consider the singular value decomposition (SVD) of a matrix  $A$  of rank  $r$

$$A = USV^H, \quad S = \text{diag}(\{\sigma_i\}), 1 \leq i \leq r,$$

where  $U$  and  $V$  are  $m \times r$  and  $n \times r$  matrices with orthonormal columns, respectively, and  $\sigma_i$  is the positive singular values. We always assume that the SVD of a matrix is given in the reduced form above. Furthermore,  $\langle A, B \rangle = \text{trace}(A^H B)$  denotes the standard inner product, then the Frobenius norm is

$$\|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\text{tr}(A^H A)} = \left( \sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \right)^{\frac{1}{2}} = \left( \sum_{i=1}^r \sigma_i^2 \right)^{\frac{1}{2}}.$$

## 1.2 Organization

In this paper, we study a perturbation theory for low-rank matrix approximation. When  $\|D\| \geq \|A\|$  or  $\|D\| \leq \|A\|$ , two sharp lower bounds of  $D - A$  are derived for a unitarily invariant norm respectively. This work is organized as follows. In Section 2, we provide a review of relevant linear algebra and some preliminary results. In Section 3, under different norms, two sharp lower bounds of  $D - A$  are given for the low-rank approximation problem and some proofs of Theorem 3.5 are presented. In Section 4, example and applications are given to verify the provided lower bounds. Finally, we conclude the paper with a short discussion.

## 2 Preliminaries

In order to prove our main results, we mention the following results for our further discussions.

### 2.1 Unitarily invariant norm

An important property of a Euclidean space is that shapes and distance do not change under rotation. In particular, for any vector  $x$  and for any unitary matrices  $U$ , we have

$$\|Ux\|_2 = \|x\|_2.$$

An analogous property is shared by the spectral and Frobenius norms: namely, for any unitary matrices  $U$  and  $V$ , the product  $UAV^H$  is defined by

$$\|UAV^H\|_p = \|A\|_p, \quad p = 2, F.$$

These examples suggest the following definition.

**Definition 2.1** ([18]) A norm  $\|\cdot\|$  on  $\mathbb{C}^{m \times n}$  is unitarily invariant if it satisfies

$$\|UAV^H\| = \|A\|$$

for any unitary matrices  $U$  and  $V$ . It is normalized if

$$\|A\| = \|A\|_2$$

whenever  $A$  is of rank one.

**Remark 2.2** Let  $\Sigma = UAV^H$  be the singular value decomposition of the matrix  $A$  with order  $n$ . Let  $\|\cdot\|$  be a unitarily invariant norm. Since  $U$  and  $V$  are unitary,

$$\|A\| = \|\Sigma\|.$$

Thus  $\|A\|$  is a function of the singular values of  $A$ .

The 2-norm plays a special role in the theory of unitarily invariant norms as the following theorem shows.

**Theorem 2.3** ([18]) *Let  $\|\cdot\|$  be a family of unitarily invariant norm. Then*

$$\|AB\| \leq \|A\| \|B\|_2 \quad (3)$$

and

$$\|AB\| \leq \|A\|_2 \|B\|. \quad (4)$$

Moreover, if  $\text{Rank}(A) = 1$ , then

$$\|A\| = \|A\|_2.$$

We have observed that the spectral and Frobenius norms are unitarily invariant. However, not all norms are unitarily invariant as the following example shows.

**Example 2.4** Let

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

obviously,  $\|A\|_\infty = 2$ , but for a unitary matrix

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

we have

$$\|UA\|_\infty = \left\| \begin{pmatrix} \frac{2}{\sqrt{2}} & \frac{2}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \right\|_\infty = \frac{4}{\sqrt{2}}.$$

**Remark 2.5** It is easy to verify that the nuclear norm  $\|\cdot\|_*$  is a unitarily invariant norm.

## 2.2 Projection

Let  $\mathbb{C}^m$  and  $\mathbb{C}^n$  be  $m$  and  $n$ -dimensional inner product spaces over the complex field, respectively, and  $A \in \mathbb{C}^{m \times n}$  be a linear transformation from  $\mathbb{C}^n$  into  $\mathbb{C}^m$ .

**Definition 2.6** ([18]) The column space (range) of  $A$  is denoted by

$$\mathcal{R}(A) = \{x \in \mathbb{C}^m | x = Ay, y \in \mathbb{C}^n\} \quad (5)$$

and the null space of  $A$  by

$$\mathcal{N}(A) = \{y \in \mathbb{C}^n | Ay = 0\}. \quad (6)$$

Further, we let  $\perp$  denote the orthogonal complement and get  $\mathcal{R} = \mathcal{N}(A^H)^\perp$  and  $\mathcal{N}(A) = \mathcal{R}(A^H)^\perp$ .

The following properties [18] of the pseudo-inverse are easily established.

**Theorem 2.7** ([18]) *For any matrix  $A$ , the following hold.*

1. If  $A \in \mathbb{C}^{m \times n}$  has rank  $n$ , then  $A^\dagger = (A^H A)^{-1} A^H$  and  $A^\dagger A = I^{(n)}$ .
2. If  $A \in \mathbb{C}^{m \times n}$  has rank  $m$ , then  $A^\dagger = A^H (A A^H)^{-1}$  and  $A A^\dagger = I^{(m)}$ .

Here  $I^{(n)} \in \mathbb{R}^{n \times n}$  is the identity matrix.

**Theorem 2.8** ([18]) *For any matrix  $A$ ,  $P_A = A A^\dagger$  is the orthogonal projector onto  $\mathcal{R}(A)$ ,  $P_{A^H} = A^\dagger A$  is the orthogonal projector onto  $\mathcal{R}(A^H)$ ,  $I - P_{A^H}$  is the orthogonal projector onto  $\mathcal{N}(A)$ .*

### 2.3 The decomposition of $D^\dagger - A^\dagger$

In this section, we focus on the decomposition of  $D^\dagger - A^\dagger$  and a general bound of the perturbation theory for pseudo-inverses. Firstly, according to the orthogonal projection, we can deduce the following lemma.

**Lemma 2.9** *For any matrix  $A$ ,  $P_A = A A^\dagger$  and  $P_{A^H} = A^\dagger A$ , then we have*

$$P_A^\perp A = 0, \quad A P_{A^H}^\perp = 0, \quad P_{A^H}^\perp A^H = 0, \quad A^\dagger P_A^\perp = 0. \quad (7)$$

*Proof* Since  $P_A^\perp = I - P_A$  and  $P_{A^H}^\perp = I - P_{A^H}$ , then we have that

$$\begin{aligned} P_A^\perp A &= (I - P_A)A = A - A A^\dagger A = A - A A^H (A A^H)^{-1} A = 0, \\ A P_{A^H}^\perp &= A(I - P_{A^H}) = A - A A^\dagger A = A - A (A^H A)^{-1} A^H A = 0, \\ P_{A^H}^\perp A^H &= A^H - P_{A^H} A^H = A^H - A^\dagger A A^H = A^H - (A^H A)^{-1} A^H A A^H = 0, \\ A^\dagger P_A^\perp &= A^\dagger (I - P_A) = A^\dagger - A^\dagger A A^\dagger = A^\dagger - (A^H A)^{-1} A^H A A^\dagger = 0. \end{aligned}$$

The proof is completed.  $\square$

Using Lemma 2.9, the decompositions of  $D^\dagger - A^\dagger$  are developed by Wedin [19].

**Theorem 2.10** ([19]) *Let  $D = A + E$ , then the difference  $D^\dagger - A^\dagger$  is given by the expressions*

$$D^\dagger - A^\dagger = -A^\dagger E D^\dagger - A^\dagger P_D^\perp + P_{A^H}^\perp D^\dagger, \quad (8)$$

$$D^\dagger - A^\dagger = -A^\dagger P_A E P_{D^H} D^\dagger - A^\dagger P_A P_D^\perp + P_{A^H}^\perp P_{D^H} D^\dagger, \quad (9)$$

**Table 1** Value options for  $\gamma$ 

$\ \cdot\ $	Arbitrary	Spectral	Frobenius
$\gamma$	3	$\frac{1+\sqrt{5}}{2}$	$\sqrt{2}$

$$D^\dagger - A^\dagger = -D^\dagger P_D E P_{A^H} A^\dagger + (D^H D)^\dagger P_{D^H} E^H P_A^\perp - P_{D^H}^\perp E P_A (A A^H)^\dagger. \quad (10)$$

By Lemma 2.9, using  $P_A = A A^\dagger$ ,  $P_{A^H} = A^\dagger A$ ,  $P_A^\perp = I - P_A$ ,  $P_{A^H}^\perp = I - P_{A^H}$ , these expressions can be verified.

In previous work [19], Wedin developed a general bound of the perturbation theory for pseudo-inverses. Theorem 2.11 is based on a useful decomposition of  $D^\dagger - A^\dagger$ , where  $D$  and  $A$  are  $m \times n$  matrices. Sharp estimates of  $\|D^\dagger - A^\dagger\|$  are derived for a unitarily invariant norm. In [20], Chen et al. presented some new perturbation bounds for the orthogonal projections  $\|P_D - P_A\|$ .

**Theorem 2.11** ([19]) *Suppose  $D = A + E$ , then the error of  $D^\dagger - A^\dagger$  has the following bound:*

$$\|D^\dagger - A^\dagger\| \leq \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|E\|, \quad (11)$$

where  $\gamma$  is given in Table 1.

**Remark 2.12** For the spectral norm, by formula (11) we can achieve  $\gamma = \frac{1+\sqrt{5}}{2}$ . When  $\|\cdot\|$  is the Frobenius norm, by formula (12), we have  $\gamma = \sqrt{2}$ . Similarly, for an arbitrary unitarily invariant norm, according to formula (13), we can deduce  $\gamma = 3$ .

**Remark 2.13** From Theorem 2.11, since  $E = D - A$ , in fact, if  $\text{Rank}(A) \leq \text{Rank}(D)$ , then (11) gives the lower bound of the low-rank matrix approximation:

$$\|D - A\| \geq \frac{\|D^\dagger - A^\dagger\|}{\gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\}}. \quad (12)$$

In the following section, based on Theorem 2.11, we provide two lower error bounds of  $D - A$  for a unitarily invariant norm.

### 3 Our main results

In this section, we consider the lower bound theory for the low-rank matrix approximation based on a useful decomposition of  $D^\dagger - A^\dagger$ . When  $\text{Rank}(A) \leq \text{Rank}(D)$ , some sharp lower bounds of  $D - A$  are derived in terms of a unitarily invariant norm. In order to prove our result, some lemmas are listed below.

**Lemma 3.1** ([18]) *Let  $D = A + E$ , the projections  $P_D$  and  $P_A$  satisfy*

$$P_D P_A^\perp = (D^\dagger)^H P_{D^H} E^H P_A^\perp = (P_A^\perp P_D)^H, \quad (13)$$

therefore

$$\|P_D P_A^\perp\| \leq \|D^\dagger\|_2 \|E\|. \quad (14)$$

If  $\text{Rank}(A) \leq \text{Rank}(D)$ , then

$$\|P_D^\perp P_A\| \leq \|P_D P_A^\perp\|. \quad (15)$$

**Lemma 3.2** ([21]) *Let  $A, D \in \mathbb{C}^{m \times n}$ ,  $\text{Rank}(A) = r$ ,  $\text{Rank}(D) = s$ ,  $r \leq s$ , then there exists a unitary matrix  $Q \in \mathbb{C}^{m \times m}$  such that*

$$QP_A Q^H = \begin{pmatrix} I^{(r)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad QP_D Q^H = \begin{pmatrix} \Gamma_r^2 & 0 & \Gamma_r \Sigma_r \\ 0 & I^{(s-r)} & 0 \\ \Sigma_r \Gamma_r & 0 & \Sigma_r^2 \end{pmatrix}, \quad (16)$$

where

$$\Gamma_r = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & I \end{pmatrix} \quad \text{and} \quad \Sigma_r = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix},$$

$\Gamma_1 = \text{diag}(\gamma_1, \dots, \gamma_{r_1})$ ,  $0 \leq \gamma_1 \leq \dots \leq \gamma_{r_1}$  and  $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_{r_1})$ ,  $0 \leq \sigma_1 \leq \dots \leq \sigma_{r_1}$ . Moreover,  $\gamma_i$  and  $\sigma_i$  satisfy  $\gamma_i^2 + \sigma_i^2 = 1$ ,  $i = 1, \dots, r_1$ .

According to Lemma 3.2, we can easily get the following result.

**Lemma 3.3** *Let  $A, D \in \mathbb{C}^{m \times n}$ ,  $\text{Rank}(A) = r$ ,  $\text{Rank}(D) = s$ ,  $r \leq s$ , then we have*

$$\|P_A^\perp P_D\| = \|P_D P_A^\perp\|. \quad (17)$$

*Proof* Since

$$P_A^\perp = I - P_A = Q^H \begin{pmatrix} 0 & 0 & 0 \\ 0 & I^{(s-r)} & 0 \\ 0 & 0 & I^{(m-s)} \end{pmatrix} Q, \quad (18)$$

and

$$P_D = Q^H \begin{pmatrix} \Gamma_r^2 & 0 & \Gamma_r \Sigma_r \\ 0 & I^{(s-r)} & 0 \\ \Sigma_r \Gamma_r & 0 & \Sigma_r^2 \end{pmatrix} Q, \quad (19)$$

then

$$P_A^\perp P_D = Q^H \begin{pmatrix} 0 & 0 & 0 \\ 0 & I^{(s-r)} & 0 \\ \Sigma_r \Gamma_r & 0 & \Sigma_r^2 \end{pmatrix} Q \quad (20)$$

and

$$P_D P_A^\perp = Q^H \begin{pmatrix} 0 & 0 & \Gamma_r \Sigma_r \\ 0 & I^{(s-r)} & 0 \\ 0 & 0 & \Sigma_r^2 \end{pmatrix} Q. \quad (21)$$

Therefore, they have the same singular values which yield that  $\|P_A^\perp P_D\| = \|P_D P_A^\perp\|$ .  $\square$

This is a useful lemma that we will use in the proof of the main result. In order to prove our main theorem, two lower bounds of  $D - A$  are required by the following lemma.

**Lemma 3.4** *For the unitarily invariant norm, if  $\text{Rank}(A) \leq \text{Rank}(D)$ , then the lower bound of  $D - A$  satisfies:*

Case I: For  $\|D\| \geq \|A\|$ , we have

$$\|D - A\| \geq \|D\| - \|A\| - \|D^\dagger - A^\dagger\| \|D\|_2 \|A\|_2. \quad (22)$$

Case II: For  $\|D\| \leq \|A\|$ , we have

$$\|D - A\| \geq \|A\| - \|D\| - \|D^\dagger - A^\dagger\| \|D\|_2^2. \quad (23)$$

*Proof Case I:* Since  $\|D\| \geq \|A\|$ , we have  $\|D - A\| \geq \|D\| - \|A\|$ . Using Theorem 2.3 and Lemma 3.1, we have  $\|AB\| \leq \|A\|_2 \|B\|$  and  $\|P_D^\perp P_A\| \leq \|P_D P_A^\perp\|$ , respectively. By Lemma 2.9, we have  $P_D^\perp D = 0$ ,  $A P_{A^H}^\perp = 0$  and  $A^\dagger P_A^\perp = 0$ , this also yields

$$\begin{aligned} \|D - A\| &\geq \|D\| - \|A\| = \|D\| - \|(P_D + P_D^\perp)(P_A + P_A^\perp)A\| \\ &= \|D\| - \|P_D P_A A + P_D^\perp P_A A\| \quad (\text{by Lemma 2.9}) \\ &= \|D\| - \|P_D P_A A\| - \|P_D^\perp P_A A\| \quad (\text{by } P_D \perp P_D^\perp) \\ &\geq \|D\| - \|A\| - \|P_D^\perp P_A\| \|A\|_2 \quad (\text{by Lemma 3.1}) \\ &\geq \|D\| - \|A\| - \|P_D P_A^\perp\| \|A\|_2 \\ &= \|D\| - \|A\| - \|D(D^\dagger - A^\dagger)P_A^\perp\| \|A\|_2 \\ &\geq \|D\| - \|A\| - \|D^\dagger - A^\dagger\| \|D\|_2 \|A\|_2 \quad (\text{by Theorem 2.3}). \end{aligned}$$

*Case II:* Since  $\|D\| \leq \|A\|$ , we have  $\|D - A\| \geq \|A\| - \|D\|$ . Similarly, by Lemma 3.3, using  $\|P_A^\perp P_D\| = \|P_D P_A^\perp\|$ , we have

$$\begin{aligned} \|D - A\| &\geq \|A\| - \|D\| = \|A\| - \|(P_A + P_A^\perp)(P_D + P_D^\perp)D\| \\ &= \|A\| - \|P_A P_D D + P_A^\perp P_D D\| \quad (\text{by Lemma 2.9}) \\ &= \|A\| - \|P_A P_D D\| - \|P_A^\perp P_D D\| \quad (\text{by } P_A \perp P_A^\perp) \\ &\geq \|A\| - \|D\| - \|P_A^\perp P_D\| \|D\|_2 \\ &\geq \|A\| - \|D\| - \|P_D P_A^\perp\| \|D\|_2 \quad (\text{by Lemma 3.3}) \\ &= \|A\| - \|D\| - \|D(D^\dagger - A^\dagger)P_A^\perp\| \|D\|_2 \\ &\geq \|A\| - \|D\| - \|D^\dagger - A^\dagger\| \|D\|_2^2 \quad (\text{by Theorem 2.3}). \end{aligned}$$

We complete the proof of Lemma 3.4.  $\square$

Our main results can be described as the following theorem.

**Theorem 3.5** *Suppose that  $D = A + E$ ,  $\text{Rank}(A) \leq \text{Rank}(D)$ , for the unitarily invariant norm  $\|\cdot\|$ , the error of  $D - A$  has the following bounds.*



Case I: For  $\|D\| \geq \|A\|$ , we have

$$\|D - A\| \geq \frac{\|D\| - \|A\|}{1 + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D\|_2 \|A\|_2}. \quad (24)$$

Case II: For  $\|D\| \leq \|A\|$ , we have

$$\|D - A\| \geq \frac{\|A\| - \|D\|}{1 + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D\|_2^2}, \quad (25)$$

where the value options for  $\gamma$  are the same as in Table 1.

*Proof Case I:* For  $\|D\| \geq \|A\|$ , by Theorem 2.11 and Lemma 3.4 (22), we can deduce

$$\begin{aligned} \|D\| - \|A\| &\leq \|D - A\| + \|D^\dagger - A^\dagger\| \|D\|_2 \|A\|_2 \\ &\leq \|D - A\| + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D - A\| \|D\|_2 \|A\|_2, \end{aligned}$$

this yields

$$\|D - A\| \geq \frac{\|D\| - \|A\|}{1 + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D\|_2 \|A\|_2}. \quad (26)$$

*Case II:* Similarly, for  $\|D\| \leq \|A\|$ , by Theorem 2.11 and Lemma 3.4 (23), we can deduce

$$\begin{aligned} \|A\| - \|D\| &\leq \|D - A\| + \|D^\dagger - A^\dagger\| \|D\|_2^2 \\ &\leq \|D - A\| + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D - A\| \|D\|_2^2, \end{aligned}$$

this yields

$$\|D - A\| \geq \frac{\|A\| - \|D\|}{1 + \gamma \max\{\|A^\dagger\|_2^2, \|D^\dagger\|_2^2\} \|D\|_2^2}, \quad (27)$$

where the value options for  $\gamma$  are the same as in Table 1. In summary, we prove the lower bounds of Theorem 3.5.  $\square$

**Remark 3.6** From the main theorem, we can see that if  $\|D\| = \|A\|$ , then  $\|D - A\| = 0$ . However, in the problem of low-rank matrix approximation,  $\|D\|$  is not necessarily equal to  $\|A\|$ , so the approximation error is present. Furthermore, when  $\|D\|$  is close to  $\|A\|$ , simulations demonstrate that the error has a very small magnitude (see Section 4).

In this section, we discuss the error bounds under different conditions for the unitarily invariant norm. Based on a useful decomposition of  $D^\dagger - A^\dagger$ , for  $\|D\| \geq \|A\|$  and  $\|D\| \leq \|A\|$ , we have bounds (26) and (27), respectively. The two error bounds are useful in low-rank matrix approximation. The following experiments illustrate our results when the approximation matrix  $A$  is low-rank and the perturbation matrix  $E$  is sparse.

## 4 Experiments

### 4.1 The singular value thresholding algorithm

Our results are obtained by a singular value thresholding (SVT [22]) algorithm. This algorithm is easy to implement and surprisingly effective both in terms of computational cost and storage requirement when the minimum nuclear norm solution is also the lowest-rank solution. The specific algorithm is described as follows.

For the low-rank matrix approximation problem which is contaminated with perturbation item  $E$ , we observe that the data matrix  $D = A + E$ . To approximate  $D$ , we can solve the convex optimization problem

$$\begin{aligned} \min \|A\|_* \\ \text{s.t. } \|D - A\|_F \leq \varepsilon, \end{aligned} \quad (28)$$

where  $\|\cdot\|_*$  denotes the nuclear norm of a matrix (*i.e.*, the sum of its singular values).

For solving (28), we introduce the soft-thresholding operator  $\mathcal{D}_\tau$  [22] which is defined as

$$\mathcal{D}_\tau(A) := U\mathcal{D}_\tau(S)V^*, \quad \mathcal{D}_\tau(S) = \text{diag}(\{(\sigma_i - \tau)_+\}),$$

where  $(\sigma_i - \tau)_+ = \max\{0, \sigma_i - \tau\}$ . In general, this operator can effectively shrink some singular values toward zero. The following theorem is with respect to the shrinkage operators [22–24], which will be used at each iteration of the proposed algorithms.

**Theorem 4.1** ([22]) *For each  $\tau > 0$  and  $W \in \mathbb{R}^{m \times n}$ , the singular value shrinkage operator  $\mathcal{D}_\tau(\cdot)$  obeys*

$$\mathcal{D}_\tau(W) = \arg \min_A \tau \|A\|_* + \frac{1}{2} \|A - W\|_F^2,$$

where  $\mathcal{D}_\tau(W) := U\mathcal{D}_\tau(S)V^*$ .

By introducing a Lagrange multiplier  $Y$  to remove the inequality constraint, one has the augmented Lagrangian function of (28)

$$L(A, Y) = \|A\|_* - \langle Y, A - D \rangle + \frac{\tau}{2} \|A - D\|_F^2.$$

The iterative scheme of the classical augmented Lagrangian multipliers method is

$$\begin{cases} A^{k+1} \in \arg \min_A L(A, Y^k), \\ Y^{k+1} = Y^k - \tau(D - A^{k+1}). \end{cases} \quad (29)$$

Based on the optimality conditions, (29) is equivalent to

$$\begin{cases} \mathbf{0} \in \frac{1}{\tau} \partial(\|A^{k+1}\|_*) + A^{k+1} - (D + \frac{1}{\tau} Y^k), \\ Y^{k+1} = Y^k - \tau(D - A^{k+1}), \end{cases} \quad (30)$$

---

**Algorithm 1** SVT

---

**Task:** Approximate the solution of (28).

**Input:** Observation matrix  $D = A + E$ , weight  $\tau$ .  $Y^0 = \text{zeros}(m, n)$

**while** the termination criterion is not met, **do**

$$A^{k+1} = \mathcal{D}_{1/\tau}(D + \frac{1}{\tau}Y^k),$$

$$Y^{k+1} = Y^k - \tau(D - A^{k+1}),$$

$$k \leftarrow k + 1.$$

**end while**

**Output:**  $A \leftarrow A^{k+1}$ .

---

**Table 2** Lower bound comparison results

$m = n$	Bound (24)		Bound (25)		Bound (12)	
	$\  \cdot \ _2$	$\  \cdot \ _F$	$\  \cdot \ _2$	$\  \cdot \ _F$	$\  \cdot \ _2$	$\  \cdot \ _F$
100	8.13e-7	1.89e-7	1.54e-7	3.31e-7	1.01e-4	1.27e-4
500	5.11e-8	3.71e-8	4.22e-8	4.62e-8	4.23e-4	5.22e-4
1,000	3.76e-8	2.14e-8	1.01e-8	1.19e-8	5.57e-4	7.48e-4

where  $\partial(\cdot)$  denotes the subgradient operator of a convex function. Then, by Theorem 4.1 above, we have the iterative solution

$$\begin{cases} A^{k+1} = \mathcal{D}_{1/\tau}(D + \frac{1}{\tau}Y^k), \\ Y^{k+1} = Y^k - \tau(D - A^{k+1}). \end{cases} \quad (31)$$

The SVT approach works as described in Algorithm 1.

## 4.2 Simulations

In this section, we use the SVT algorithm for the low-rank matrix approximation problem. Let  $D = A + E \in \mathbb{R}^{m \times n}$  be the available data. Simply, we restrict our examples to square matrices ( $m = n$ ). We draw  $A$  according to the independent random matrices and generate the perturbation matrix  $E$  to be sparse, which satisfies the i.i.d. Gaussian distribution. Specially, the rank of the matrix  $A$  and the sparse entries of the perturbation matrix  $E$  are selected to be  $5\%m$  and  $5\%m^2$ , respectively.

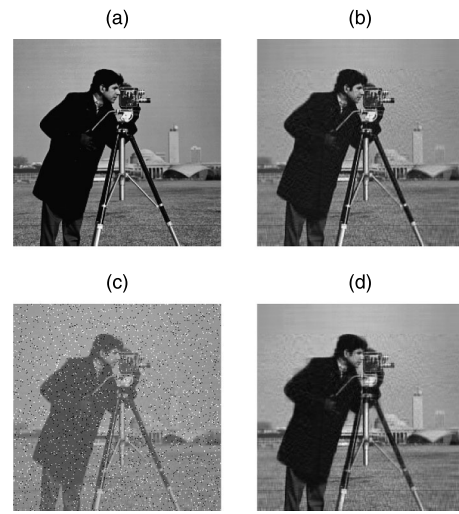
Table 2 reports the results obtained by lower bounds (24), (25) and (12), respectively. Bounds (24) and (25) are our new result, bound (12) is the previous result. Then, comparing the bounds with each other by numerical experiments, we find that lower bounds (24), (25) are smaller than lower bound (12).

## 4.3 Applications

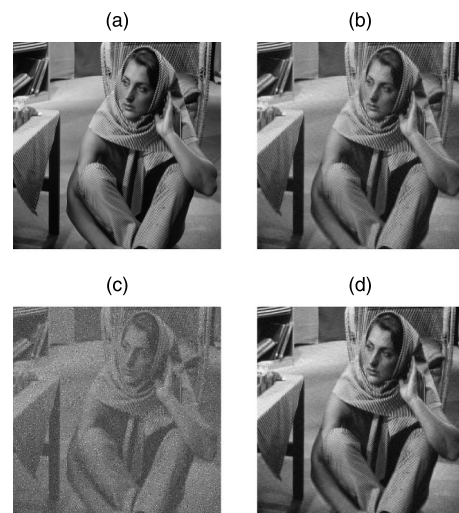
In this section, we use the SVT algorithm for the low-rank image approximation. From Figures 1 and 2, comparing with the original image (a), the low-rank image (b) loses some details. We can hardly get any detailed information from incomplete image (c). However, the output image (d) =  $A^k$ , which is obtained by the SVT algorithm, can recover the details of the low-rank image (b). If we denote image (b) to be a low-rank matrix  $A$ , then image (c) is the observed data matrix  $D$  which is perturbed by a sparse matrix  $E$ , that is,

$$\text{image (c)} = \text{image (b)} + E.$$

**Figure 1 Cameraman.** (a) Original  $256 \times 256$  image with full rank. (b) Original image truncated to be rank 50. (c) 50% randomly masked of (b). (d) Recovered image from (c).



**Figure 2 Barbara.** (a) Original  $512 \times 512$  image with full rank. (b) Original image truncated to be rank 100. (c) 50% randomly masked of (b). (d) Recovered image from (c).



**Table 3 Lower bound comparison results of low-rank image approximation**

	Cameraman	Barbara
$\ E\ _F$	8.71e-2	7.23e-2
Bound (25)	2.59e-5	1.09e-5
Iters	200	200

Using the SVT algorithm for the low-rank image approximation problem, the lower bound comparison results are shown in Table 3. We calculate  $\|E\|_F = \|D - A\|_F$  are 8.71e-2 and 7.23e-2 for images Cameraman and Barbara, respectively. But for F-norm of our lower bound (25), we can see that they are 2.59e-5 and 1.09e-5 for images Cameraman and Barbara, respectively. That is to say, our error bounds can verify that the SVT algorithm still can be improved.

## 5 Conclusion

Low-rank matrix approximation problem is a field which arises in a number of applications in model selection, system identification, complexity theory, and optics. Based on a useful decomposition of  $D^\dagger - A^\dagger$ , this paper reviewed the previous work and provided two sharp lower bounds for the low-rank matrices recovery problem with a unitarily invariant norm.

From our main Theorem 3.5, we can see that if  $\|D\| = \|A\|$ , then  $\|D - A\| = 0$ . However, in the problem of low-rank matrix approximation,  $\|D\|$  is not necessarily equal to  $\|A\|$ , so the approximation error is present. Furthermore, from the main results, we can clearly find the influence of the spectral norm ( $\|\cdot\|_2$ ) on the low-rank matrix approximation. For example, in Case II, when the maximum eigenvalue of the matrix  $D$  is larger, the error of  $D - A$  is smaller.

Finally, we use the SVT algorithm for the low-rank matrix approximation problem. Table 2 shows that our lower bounds (24), (25) are smaller than lower bound (12). Simulation results demonstrate that the lower bounds have a very small magnitude. In applications section, we use the SVT algorithm for the low-rank image approximation problem, the lower bounds comparison results are shown in Table 3. From the comparison results, we find that our lower bounds can verify whether the SVT algorithm can be improved.

## Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under grant No. 11671318, and the Fundamental Research Funds for the Central Universities (Xi'an Jiaotong University, Grant No. xkjc2014008).

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

All authors worked in coordination. All authors carried out the proof, read and approved the final version of the manuscript.

## Author details

<sup>1</sup>School of Mathematics and Statistics, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, 710049, China.

<sup>2</sup>School of Statistics, Henan University of Economics and Law, No. 180, Jinshui East Road, Zhengzhou, 450046, China.

<sup>3</sup>College of Mathematics and Statistics, Shenzhen University, No. 3688, Nanhai Ave, Shenzhen, 518060, China.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 August 2017 Accepted: 10 November 2017 Published online: 21 November 2017

## References

1. Fazel, M, Hindi, H, Boyd, S: A rank minimization heuristic with application to minimum order system approximation. In: Proceedings of the American Control Conference, vol. 6, pp. 4734-4739 (2002)
2. Linial, N, London, E, Rabinovich, Y: The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15**, 215-245 (1995)
3. Tomasi, C, Kanade, T: Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vis.* **9**, 137-154 (1992)
4. Chen, P, Suter, D: Recovering the missing components in a large noisy low-rank matrix: application to SFM. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(8), 1051-1063 (2004)
5. Liu, ZS, Li, JC, Li, G, Bai, JC, Liu, XN: A new model for sparse and low-rank matrix decomposition. *J. Appl. Anal. Comput.* **2**, 600-617 (2017)
6. Wright, J, Ganesh, A, Shankar, R, Yigang, P, Ma, Y: Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. In: Twenty-Third Annual Conference on Neural Information Processing Systems (NIPS 2009) (2009)
7. Deerwester, S, Dumains, ST, Landauer, T, Furnas, G, Harshman, R: Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci. Technol.* **41**(6), 391-407 (1990)
8. Papadimitriou, C, Raghavan, P, Tamaki, H, Vempala, S: Latent semantic indexing, a probabilistic analysis. *J. Comput. Syst. Sci.* **61**(2), 217-235 (2000)
9. Argyriou, A, Evgeniou, T, Pontil, M: Multi-task feature learning. *Adv. Neural Inf. Process. Syst.* **19**, 41-48 (2007)
10. Abernethy, J, Bach, F, Evgeniou, T, Vert, JP: Low-rank matrix factorization with attributes. *arXiv:cs/0611124* (2006)
11. Amit, Y, Fink, M, Srebro, N, Ullman, S: Uncovering shared structures in multiclass classification. In: Proceedings of the 24th International Conference on Machine Learning, pp. 17-24. ACM, New York (2007)

12. Zhang, HY, Lin, ZC, Zhang, C, Gao, J: Robust latent low rank representation for subspace clustering. *Neurocomputing* **145**, 369-373 (2014)
13. Mesbahi, M, Papavassilopoulos, GP: On the rank minimization problem over a positive semidefinite linear matrix inequality. *IEEE Trans. Autom. Control* **42**, 239-243 (1997)
14. Golub, GH, Van Loan, CF: *Matrix Computations*, 4th edn. Johns Hopkins University Press, Baltimore (2013)
15. Eckart, C, Young, G: The approximation of one matrix by another of lower rank. *Psychometrika* **1**(3), 211-218 (1936)
16. Hotelling, H: Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**(6), 417-520 (1932)
17. Jolliffe, I: *Principal Component Analysis*. Springer, Berlin (1986)
18. Stewart, GW, Sun, JG: *Matrix Perturbation Theory*. Academic Press, New York (1990)
19. Wedin, P-Å: Perturbation theory for pseudo-inverses. *BIT Numer. Math.* **13**(2), 217-232 (1973)
20. Chen, YM, Chen, XS, Li, W: On perturbation bounds for orthogonal projections. *Numer. Algorithms* **73**, 433-444 (2016)
21. Sun, JG: *Matrix Perturbation Analysis*, 2nd edn. Science Press, Beijing (2001)
22. Cai, J-F, Candès, EJ, Shen, ZW: A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20**(4), 1956-1982 (2010)
23. Candès, EJ, Li, X, Ma, Y, Wright, J: Robust principal component analysis? *J. ACM* **58**(3), 1-37 (2011)
24. Tao, M, Yuan, XM: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.* **20**(1), 57-81 (2011)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)