Journal of Machine Learning Research 20 (2019) 1-22 Submitted 9/15; Revised 3/19; Published 6/19

Relative Error Bound Analysis for Nuclear Norm Regularized Matrix Completion

Lijun Zhang ZHANGLJ@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China

Tianbao Yang TIANBAO-YANG@UIOWA.EDU
Department of Computer Science
The University of Iowa, Iowa City, IA 52242, USA

Rong Jin JINRONG.JR@ALIBABA-INC.COM
Machine Intelligence Technology
Alibaba Group, Bellevue, WA 98004, USA

Zhi-Hua Zhou ZHOUZH@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China

Editor: Nathan Srebro

Abstract

In this paper, we develop a *relative* error bound for nuclear norm regularized matrix completion, with the focus on the completion of full-rank matrices. Under the assumption that the top eigenspaces of the target matrix are incoherent, we derive a relative upper bound for recovering the best low-rank approximation of the unknown matrix. Although multiple works have been devoted to analyzing the recovery error of full-rank matrix completion, their error bounds are usually additive, making it impossible to obtain the perfect recovery case and more generally difficult to leverage the skewed distribution of eigenvalues. Our analysis is built upon the optimality condition of the regularized formulation and existing guarantees for low-rank matrix completion. To the best of our knowledge, this is the first relative bound that has been proved for the regularized formulation of matrix completion.

Keywords: matrix completion, nuclear norm regularization, least squares, low-rank, full-rank, relative error bound

1. Introduction

Matrix completion is concerned with the problem of recovering an unknown matrix from a small fraction of its entries (Candès and Tao, 2010). Recently, the problem of low-rank matrix completion has received a great deal of interests due to the theoretical advances (Candès and Recht, 2009; Keshavan et al., 2010a), as well as its application to a wide range of real-world problems, including collaborative filtering (Goldberg et al., 1992), sensor networks (Biswas et al., 2006), computer vision (Cabral et al., 2011), and machine learning (Jalali et al., 2011).

Let A be an unknown matrix of size $m \times n$, and without loss of generality, we assume $m \leq n$. The information available about A is a sampled set of entries A_{ij}, $(i, j) \in \Omega$, where

Ω is a subset of the complete set of entries $[m] \times [n]$. Our goal is to recover A as precisely as possible. In a seminal work, Candès and Recht (2009) assume that A is low-rank, and propose to recover A from the observed entries in Ω by solving the following nuclear norm minimization problem

$$\begin{aligned} \min \quad & \|B\|_* \\ \text{s.t.} \quad & B_{ij} = A_{ij} \ \forall (i,j) \in \Omega. \end{aligned} \tag{1}$$

Under the incoherence condition, they prove that with a high probability the solution to (1) yields a perfect reconstruction of A, provided that a sufficiently large number of entries are observed randomly. When A is of full rank, a similar nuclear norm minimization problem has been proposed. Suppose $A = Z + N$, where Z is a low-rank matrix to recover, and N is the residual matrix. Candès and Plan (2010) introduce the following problem for recovering A

$$\begin{aligned} \min \quad & \|B\|_* \\ \text{s.t.} \quad & \sqrt{\textstyle\sum_{(i,j)\in\Omega} (B_{ij} - A_{ij})^2} \leq \delta \end{aligned} \tag{2}$$

where δ is an upper bound for $\sqrt{\sum_{(i,j)\in\Omega} N_{ij}^2}$. Although a relative error bound has been established for (2) when δ is large enough (Candès and Plan, 2010), the high computational cost with solving the optimization problem in (2), mostly due to the constraint and non-smooth objective function, makes it practically less attractive.

An alternative approach to (2) for matrix completion is to solve a nuclear norm regularized least squares problem

$$\min_{B \in \mathbb{R}^{m\times n}} \quad \frac{1}{2} \sum_{(i,j)\in\Omega} (B_{ij} - A_{ij})^2 + \lambda \|B\|_*. \tag{3}$$

This is the approach that is favored by practitioners because it can be solved significantly more efficiently than (2). In fact, a number of efficient optimization methods have been designed (Ji and Ye, 2009; Toh and Yun, 2010; Pong et al., 2010; Zhang et al., 2012; Hsieh and Olsen, 2014). Using the accelerated gradient method (Nesterov, 2013), the convergence rate for solving (3) is $O(1/T^2)$, where T is number of iterations, and can be even boosted to a linear convergence under mild conditions (Hou et al., 2013). In contrast, the convergence rate for (2) could be as low as $O(1/\sqrt{T})$.

Although (3) is computation-friendly, its recovery guarantee remains unclear. One may argue that (2) and (3) are equivalent by setting δ and λ appropriately, but the exact correspondence between them is unknown in general. To bridge the gap between practice and theory, in this paper we provide a *relative* error bound for the regularized formulation in (3). More specifically, assume A is a matrix of full rank to be recovered. Let A_r be the best rank-r approximation of A, and $\widehat{A}$ be the matrix recovered from the observed entries in Ω. A relative upper bound takes the following form

$$\|\widehat{A} - A_r\|_F \leq U(r, m, n, |\Omega|) \|A - A_r\|_F \tag{4}$$

where $U(\cdot)$ is a function of r, m, n and $|\Omega|$. [1] Note that this kind of bounds is very popular in compressive sensing (Cohen et al., 2009) and low-rank matrix approximation (Boutsidis

1. By the triangle inequality $\|\widehat{A} - A\|_F \leq \|\widehat{A} - A_r\|_F + \|A - A_r\|_F$, a relative upper bound for recovering A_r directly implies a relative upper bound for recovering A.

et al., 2009). Compared to the additive error bound, the key advantage of the relative error bound is that it bounds the error based on $\|A - A_r\|_F$, the approximation error between the original matrix A and its low-rank approximation A_r. As a result, when A is low-rank and $A - A_r = \mathbf{0}$, relative error bounds imply a perfect recovery of A, which will never be accomplished by additive bounds.

In this work, we are interested in bounding $\|B_* - A_r\|_F$ in the form of (4), where B_* is the optimal solution to (3). Similar to previous studies, we assume that the top eigenspaces of A satisfy the classical incoherence condition (Candès and Recht, 2009). Based on the celebrated result of low-rank matrix completion (Recht, 2011), we derive an upper bound for $\|B_* - A_r\|_F$, which induces a relative upper bound under favored conditions. We summarize the key features of our results as follows:

- We present a general theorem that allows us to bound the recovery error of (3) for any $\lambda > 0$. In contrast, Candès and Plan (2010) only analyze the performance of (2) when $\delta \geq \sqrt{\sum_{(i,j)\in\Omega} N_{ij}^2}$.
- By choosing λ appropriately, we obtain a relative upper bound of $O(\frac{mn\sqrt{r}}{|\Omega|}\|A - A_r\|_F)$ in general, and a tighter bound of $O(\sqrt{\frac{mnr}{|\Omega|}}\|A - A_r\|_F)$ when $A - A_r$ is flat, i.e., $\|A - A_r\|_\infty / \|A - A_r\|_F$ is not too large. Although Koltchinskii et al. (2011) and Negahban and Wainwright (2012) have analyzed some variants of (3), their bounds are additive in the sense that they are not proportional to $\|A - A_r\|_F$. To the best of our knowledge, this is the *first* relative error bound for the nuclear norm regularized matrix completion.
- Our relative upper bound for (3) is tighter than that for (2) developed by Candès and Plan (2010), and more general than those proved by Keshavan et al. (2010b) and Eriksson et al. (2012) under different conditions.
- Compared to the additive upper bounds of other methods (Keshavan et al., 2010b; Koltchinskii et al., 2011; Foygel and Srebro, 2011), our relative upper bound is tighter when $\|A - A_r\|_F$ is small. In addition, our relative error bound implies the perfect recovery case when the target matrix A is low-rank while the additive error never vanishes.

Notations For a matrix X, we use $\|X\|_*$, $\|X\|_F$, $\|X\|$, and $\|X\|_\infty$ to denote its nuclear norm, Frobenius norm, spectral norm, and the absolute value of the largest element in magnitude, respectively,

2. Related Work

In this section, we provide a brief review of existing work.

2.1. Low-rank Matrix Completion

The mathematical study of matrix completion began with Candès and Recht (2009). Specifically, they have proved that if A obeys the incoherence condition, $|\Omega| \geq Cn^{6/5} r \log(n)$ is sufficient to ensure that with a high probability, A is the unique solution to (1), where C is a constant independent from r, m, and n (Candès and Recht, 2009). The lower bound for the size of Ω is subsequently improved to $nr\log^6(n)$ under a stronger assumption (Candès and

Tao, 2010). These theoretical guarantees are without question great breakthroughs, but the proof techniques are highly involved. In two subsequent studies (Recht, 2011; Gross, 2011), the authors present a very elegant approach for analyzing (1), and give slightly better bounds. For example, Recht (2011) improves the bound for $|\Omega|$ to $rn\log^2(n)$ and requires the weakest assumptions on A. The simplification of the analysis also leads to better understanding of matrix completion, and lays the foundations of the study in this paper.

In an alternative line of work, Keshavan et al. (2010a) study matrix completion using a combination of spectral techniques and manifold optimization. The proposed algorithm named OPTSPACE, also achieves exact recovery if $|\Omega| \geq Cnr\max(\log(n), r)$. However, the constant C in their bound depends on many factors of A such as the aspect ratio and the condition number. After the pioneering work mentioned above, various algorithms and theories of matrix completion have been developed, including distributed matrix completion (Mackey et al., 2011), matrix completion with side information (Xu et al., 2013), 1-bit matrix completion (Cai and Zhou, 2013), noisy matrix completion (Klopp, 2014), coherent matrix completion (Chen et al., 2014), universal matrix completion (Bhojanapalli and Jain, 2014), and non-convex matrix completion (Sun and Luo, 2015), to name a few amongst many.

2.2. Full-rank Matrix Completion

Since existing studies for full-rank matrix completion differ significantly in their assumptions, their theoretical guarantees may not be directly comparable. In the following, we will state previous results in the most general form, and (if possible) characterize their behaviors with respect to m, n, r, and $|\Omega|$.

Denote the optimal solution of (2) by $\widehat{B}$. Under the assumption $\delta \geq \sqrt{\sum_{(i,j)\in\Omega} N_{ij}^2}$, Theorem 7 of Candès and Plan (2010) shows

$$\|\widehat{B} - Z\|_F \leq \left(1 + m\sqrt{\frac{n}{|\Omega|}}\right)\delta.$$

Let $Z = A_r$, $N = A - A_r$, and consider the optimal choice that $\delta = O\left(\sqrt{\sum_{(i,j)\in\Omega} N_{ij}^2}\right)$. The above bound becomes

$$\|\widehat{B} - A_r\|_F \leq \left(1 + m\sqrt{\frac{n}{|\Omega|}}\right)\sqrt{\sum_{(i,j)\in\Omega}(A - A_r)_{ij}^2}. \tag{5}$$

One limitation of this work is that the theoretical guarantee is only valid when δ is sufficiently large. On the other hand, if we use a very large δ, the upper bound becomes loose. Our result overcomes this limitation as our error bound holds for any positive regularization parameter $\lambda > 0$.

An investigation of OPTSPACE (Keshavan et al., 2010a) for full-rank matrix completion is discussed in Keshavan et al. (2010b). In particular, Theorem 1.1 of Keshavan et al. (2010b) implies the following additive upper bound

$$O\left(\|A_r\|_\infty m^{1/4}n^{5/4}\sqrt{\frac{r}{|\Omega|}} + \frac{mn\sqrt{r}}{|\Omega|}\|U\|\right) \tag{6}$$

where U is some matrix that depends on $A - A_r$ and Ω. Although it is possible to derive a relative upper bound from Theorem 1.2 of Keshavan et al. (2010b), it requires very strong assumptions about the coherence, the aspect ratio (n/m), the condition number of A_r and the r-th singular value of A. Thus, the bound derived from Theorem 1.2 of Keshavan et al. (2010b) is significantly more restricted than the bound proved here.

Foygel and Srebro (2011) study the problem of matrix completion from the view point of supervised learning. The optimization problem is formulated as least squares minimization subject to nuclear norm or max norm constraints. Their theoretical results follow from generic generalization guarantees based on the Rademacher complexity. Specifically, Theorem 6 of Foygel and Srebro (2011) implies the following additive upper bound

$$O\left(\|A - A_r\|_F + n\sqrt{\frac{rm}{|\Omega|}} + \sqrt{n\|A - A_r\|_F}\left(\frac{rm}{|\Omega|}\right)^{\frac{1}{4}}\right) \tag{7}$$

where logarithmic factors are ignored. The derivation of (7) is given in Appendix A.

Koltchinskii et al. (2011) have investigated a general trace regression model, which includes matrix completion as a special case. For matrix completion, they propose the following optimization problem

$$\min_{B \in \mathbb{R}^{m \times n}} \quad \frac{1}{2}\left\| B - \frac{mn}{|\Omega|}\sum_{(i,j)\in\Omega} A_{ij}\mathbf{e}_i\mathbf{e}_j^\top \right\|_F^2 + \lambda\|B\|_*.$$

Let $\widehat{B}$ be the optimal solution to the above problem. Under appropriate conditions, it has been proved that with a high probability (Koltchinskii et al., 2011, Corollary 2)

$$\|\widehat{B} - A\|_F^2 + \|\widehat{B} - X\|_F^2 \leq \|X - A\|_F^2 + \frac{Cmn^2\log(n)\mathrm{rank}(X)}{|\Omega|} \tag{8}$$

for all $X \in \mathbb{R}^{m \times n}$. However, due to the presence of the second term in the upper bound, it is impossible to obtain a relative error bound.

Negahban and Wainwright (2012) have analyzed a variant of (3), which contains an additional ℓ_∞-norm constraint. Based on assumptions about the spikiness and rank of the target matrix, they derive the restricted strong convexity condition, and establish the following additive bound (Negahban and Wainwright, 2012, Theorem 2)

$$\begin{aligned} &\max\left(\left(\frac{mn^2\log n}{|\Omega|}\right)^{1/4}\sqrt{\|A - A_r\|_*}, n\sqrt{\frac{rm\log n}{|\Omega|}}\right) \\ \leq &\max\left(\left(\frac{m^2n^2\log n}{|\Omega|}\right)^{1/4}\sqrt{\|A - A_r\|_F}, n\sqrt{\frac{rm\log n}{|\Omega|}}\right). \end{aligned} \tag{9}$$

Thus, their optimization problem, assumptions and theoretical guarantees are all different from ours.

In a recent work, Eriksson et al. (2012) consider a high-rank matrix completion problem in which the columns of A belong to a union of multiple low-rank subspaces. Under certain assumptions about the coherence as well as the geometrical arrangement of subspaces and the distribution of the columns in the subspaces, they develop a multi-step algorithm that is able to recover *each* column of A with a high probability, as long as $O(rn\log^2(m))$ entries of A are observed uniformly at random. However, the recovery guarantee of their algorithm for general full-rank matrices is unclear.

3. Our Results

We first describe theoretical guarantees and then provide some discussions.

3.1. Theoretical Guarantees

Let $U = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$ and $V = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ be two matrices that contain the first r left and right singular vectors of matrix A, respectively. Let $\mathbf{e}_i$ and $\mathbf{e}_j$ be the i-th and j-th standard basis in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Following the previous studies in matrix completion (Candès and Recht, 2009; Recht, 2011), we define the coherence measure μ_0 as

$$\mu_0 = \max\left(\frac{m}{r}\max_{1\le i\le m}\|P_U\mathbf{e}_i\|^2, \frac{n}{r}\max_{1\le j\le n}\|P_V\mathbf{e}_j\|^2\right)$$

where $P_U = UU^\top$ and $P_V = VV^\top$ are two projection operators. We also define μ_1 as

$$\mu_1 = \max_{i\in[m], j\in[n]}\sqrt{\frac{mn}{r}}\left|[UV^\top]_{ij}\right|.$$

Define two projection operators $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$ for matrices as

$$\mathcal{P}_T(Z) = P_U Z + Z P_V - P_U Z P_V, \text{ and } \mathcal{P}_{T^\perp}(Z) = (I - P_U)Z(I - P_V).$$

We assume the indices are sampled uniformly with replacement, and thus Ω is a *collection* that may contain duplicate indices. The linear operator $\mathcal{R}_\Omega : \mathbb{R}^{m\times n} \mapsto \mathbb{R}^{m\times n}$ is defined as

$$\mathcal{R}_\Omega(Z) = \sum_{(i,j)\in\Omega}\langle \mathbf{e}_i\mathbf{e}_j^\top, Z\rangle \mathbf{e}_i\mathbf{e}_j^\top.$$

To simplify the notation, we define

$$\varepsilon = \|A - A_r\|_F.$$

3.1.1. A General Result

Let B_* be the optimal solution to (3). Based on the optimality condition of B_* and the guarantee for low-rank matrix completion (Recht, 2011), we obtain the following theorem.

Theorem 1 *Assume*

$$|\Omega| \ge 114\max(\mu_0, \mu_1^2)r(m+n)\beta\log^2(2n) \tag{10}$$

for some $\beta > 1$*, and* $n \geq 5$*. With a probability at least* $1 - 6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}}$*, we have*

$$\|\mathcal{P}_{T^\perp}(B_*)\|_F \leq \|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \frac{256\beta\log^2(n)\varepsilon^2}{9\lambda} + \frac{3mnr\log(2n)\lambda}{|\Omega|},$$

$$\|\mathcal{P}_T(A_r - B_*)\|_F \leq \frac{16\log(n)\varepsilon}{3}\sqrt{\frac{2\beta mn}{|\Omega|}} + \frac{2mn\lambda}{|\Omega|}\sqrt{3r\log(2n)} + 64\log(n)\sqrt{\frac{mn\beta}{6|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F.$$

As can be seen, our upper bound is valid for any $\lambda > 0$. In contrast, the upper bound for (2) in (Candès and Plan, 2010) is limited to the case $\delta \geq \sqrt{\langle \mathcal{R}_\Omega(A - A_r, A - A_r)\rangle}$.

By choosing λ to minimize the upper bounds in the above theorem, we obtain the following corollary.

Corollary 2 *Under the condition in Theorem 1. Set*

$$\lambda = \frac{16\varepsilon}{3}\sqrt{\frac{\beta\log(2n)|\Omega|}{3mnr}}. \tag{11}$$

With a probability at least $1 - 6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}}$*, we have*

$$\|\mathcal{P}_{T^\perp}(B_*)\|_F \leq \frac{32\log(2n)}{3}\sqrt{\frac{3\beta mnr\log(2n)}{|\Omega|}}\varepsilon,$$

$$\|\mathcal{P}_T(A_r - B_*)\|_F \leq \left(19\log(2n)\sqrt{\frac{\beta mn}{|\Omega|}} + \frac{2048\beta\log^2(2n)mn}{3|\Omega|}\sqrt{\frac{r\log(2n)}{2}}\right)\varepsilon,$$

and thus

$$\|A_r - B_*\|_F \leq O\left(\log(n)\sqrt{\frac{mnr\log(n)}{|\Omega|}} + \frac{mn\log^2(n)}{|\Omega|}\sqrt{r\log(n)}\right)\varepsilon.$$

Corollary 2 shows that, with an appropriate choice of the parameter λ, we can obtain a relative upper bound. One way to estimate λ is to use the cross validation technique, an approach that is widely used in learning. More specifically, we can divide the observed entries into two separate sets: the training set and the validation set. We will use the training set to find the optimal solution to the recovered matrix, and use the validation set to determine the appropriate parameter λ. When (11) holds, we can also express the upper bounds in terms of λ. It is easy to verify that with a high probability, we have

$$\|A_r - B_*\|_F \leq O\left(\frac{mnr\log^2(n)}{|\Omega|}\sqrt{\frac{mn}{|\Omega|}}\right)\lambda. \tag{12}$$

In the special case when A is a rank-r matrix, i.e., $A = A_r$, (12) implies the smaller the λ, the better the bound. In other words, we have $\|A - B_*\|_F \to 0$ as $\lambda \to 0$.

Finally, we note that whether the upper bound in Corollary 2 is tight remains open. Although both Koltchinskii et al. (2011, Theorem 6) and Negahban and Wainwright (2012, Theorem 3) have established lower bounds for noisy full-rank matrix completion, their bounds become 0 in the noisy-free setting. Thus, existing lower bounds cannot be used to examine the optimality of our result, and we will investigate the lower bound for noisy-free setting in the future.

3.1.2. A Special Result with Tighter Bounds

In the case that the residual matrix $A-A_r$ is not too spiky, in other words, $\|A-A_r\|_\infty/\|A-A_r\|_F$ is not too large, we obtain a tighter theorem as stated below.

Theorem 3 *Assume*

$$|\Omega| \geq \max\left(114\max(\mu_0, \mu_1^2) r(m+n)\beta\log^2(2n), \frac{8mn\|A-A_r\|_\infty^2}{3\|A-A_r\|_F^2}\beta\log(n)\right) \tag{13}$$

for some $\beta > 1$, and $n \geq 5$. With a probability at least $1-6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}} - n^{-\beta}$, we have

$$\|\mathcal{P}_{T^\perp}(B_*)\|_F \leq \|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \frac{8|\Omega|\varepsilon^2}{mn\lambda} + \frac{3mnr\log(2n)\lambda}{|\Omega|},$$

$$\|\mathcal{P}_T(A_r - B_*)\|_F \leq 4\varepsilon + \frac{2mn\lambda}{|\Omega|}\sqrt{3r\log(2n)} + 64\log(n)\sqrt{\frac{mn\beta}{6|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F.$$

In this theorem, we have two lower bounds for $|\Omega|$ in (13). If A is low-rank, the second lower bound will vanish. Furthermore, it can be dropped when $(\|A-A_r\|_\infty^2/\|A-A_r\|_F^2) \leq O(r\log n/m)$, i.e., when the residual matrix does not concentrate on a small number of entries. Note that our flatness assumption is a condition over the residual matrix $A - A_r$. It is different from the ℓ_∞-norm constraint of Negahban and Wainwright (2012), which is a requirement over the target matrix A.

By choosing λ to minimize the upper bounds in Theorem 3, we obtain the following relative upper bounds.

Corollary 4 *Under the condition in Theorem 1. Set*

$$\lambda = \frac{2|\Omega|\varepsilon}{mn}\sqrt{\frac{2}{3r\log(2n)}}.$$

With a probability at least $1-6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}} - n^{-\beta}$, we have

$$\|\mathcal{P}_{T^\perp}(B_*)\|_F \leq 4\sqrt{6r\log(2n)}\varepsilon,$$

$$\|\mathcal{P}_T(A_r - B_*)\|_F \leq \left(10 + 256\sqrt{\frac{mnr\log^3(2n)\beta}{|\Omega|}}\right)\varepsilon,$$

and thus

$$\|A_r - B_*\|_F \leq O\left(\sqrt{r\log(n)} + \sqrt{\frac{mnr\log^3(n)}{|\Omega|}}\right)\varepsilon.$$

As can be seen, the upper bound for $\|A_r - B_*\|_F$ in the above corollary is tighter than that in Corollary 2 by a factor of $\log(n)\sqrt{\frac{mn}{|\Omega|}}$.

3.2. Comparisons

We compare our theoretical guarantees with previous results for matrix completion in this section. We focus on the practical scenario $|\Omega| \leq mn$, and for simplicity *ignore* logarithmic factors.

The most comparable study is the relative upper bound derived by Candès and Plan (2010) for the constrained problem in (2), since their analysis also relies on the incoherence condition. In the general case, Corollary 2 gives the following relative error bound

$$\|A_r - B_*\|_F \leq O\left(\frac{mn\sqrt{r}}{|\Omega|}\right)\varepsilon.$$

From (5) in Section 2.2, we observe that Candès and Plan (2010) give the following bound

$$\|A_r - \widehat{B}\|_F \leq O\left(m\sqrt{\frac{n}{|\Omega|}}\right)\varepsilon.$$

Because $|\Omega| \geq Cnr$ for some constant C, we have

$$m\sqrt{\frac{n}{|\Omega|}} \geq \sqrt{C}\frac{mn\sqrt{r}}{|\Omega|},$$

which implies our bound is always tighter than that of Candès and Plan (2010). In the case that (13) holds, Corollary 4 indicates our relative error bound can be improved to

$$\|A_r - B_*\|_F \leq O\left(\sqrt{\frac{mnr}{|\Omega|}}\right)\varepsilon.$$

Using Lemma 2 in Section 4.7, the error bound of Candès and Plan (2010) can also be improved and becomes

$$\|A_r - \widehat{B}\|_F \leq O\left(m\sqrt{\frac{n}{|\Omega|}}\right)\sqrt{\langle \mathcal{R}_\Omega(A - A_r), A - A_r\rangle} \overset{(43)}{\leq} O\left(\sqrt{m}\varepsilon\right),$$

which is again worse than our bound since $|\Omega| \geq Cnr$. Compared to the relative upper bounds of Keshavan et al. (2010b) and Eriksson et al. (2012), our result is applicable to a more general case as their bounds only hold for a very restricted class of matrices.

Next, we compare our relative error bound with the additive bounds in previous studies (Keshavan et al., 2010b; Foygel and Srebro, 2011; Koltchinskii et al., 2011; Negahban and Wainwright, 2012). Since those results are derived under different assumptions, the comparison should be treated conservatively. We remark that those assumptions are incomparable in general, since we can construct matrices to satisfy one assumption but violate others (Negahban and Wainwright, 2012, Setion 3.4.2). Our goal is to show that relative bounds could be tighter than additive bounds under certain conditions.

For brevity, we only provide the comparison using the tighter bound in Corollary 4. Our relative bound $O(\sqrt{\frac{mnr}{|\Omega|}}\varepsilon)$ is tighter than the additive bound in (6) derived by Keshavan et al. (2010b), if $\varepsilon \leq O(n^{3/4}/m^{1/4})$ and also tighter than the additive bound in (7) derived by Foygel and Srebro (2011), if $\varepsilon \leq O(\sqrt{n})$. To compare with the additive bound in (8) derived by Koltchinskii et al. (2011), we set $X = A_r$ and have, with a high probability,

$$\left\|\widehat{B} - A_r\right\|_F \leq \varepsilon + O\left(\sqrt{\frac{mn^2 r}{|\Omega|}}\right)$$

which is worse than our bound if $\varepsilon \leq O(\sqrt{n})$. Our bound is better than the additive bound in (9) of Negahban and Wainwright (2012), when

$$\epsilon \leq \max\left(\frac{\sqrt{\Omega}}{r}, \sqrt{n}\right).$$

Finally, we note that although our analysis is devoted to full-rank matrix completion, it can also be applied to noisy low-rank matrix completion. In this case, we have $A = Z + N$, where Z is a low-rank matrix and N is the matrix of noise. As long as the eigenspaces of Z satisfy the incoherence condition, our theoretical guarantees are valid by setting $A_r = Z$ and $\varepsilon = \|N\|_F$, even when Z may not be the best rank-r approximation of A. [2] To compare with previous studies, let's assume $\|Z\|_F = 1$ and entries of N are independent sampled from $\mathcal{N}(0, \sigma^2)$ where $\sigma^2 = 1/(mn)$. Based on the concentration inequality for χ^2-distributions (Laurent and Massart, 2000, Lemma 1), with a high probability, $\varepsilon^2 = O(\sigma^2 mn) = O(1)$. Then, our Corollary 2 and Corollary 4 imply that with a high probability

$$\|Z - B_*\|_F \leq O\left(\frac{mn\sqrt{r}}{|\Omega|}\right) \text{ and } \|Z - B_*\|_F \leq O\left(\sqrt{\frac{mnr}{|\Omega|}}\right),$$

respectively. In contrast, existing results for noisy low-rank matrix completion, e.g., Corollary 1 of Negahban and Wainwright (2012) and Theorem 7 of Klopp (2014), have established an $O(\sqrt{rn/|\Omega|})$ bound. Thus, our bounds are loose for noisy low-rank matrix completion, which is probably because our analysis did not exploit the fact that entries of N are i.i.d. sampled.

4. Analysis

Although the current analysis is built upon the result from Recht (2011) that requires the incoherence assumption, it can be extended to support other assumptions for matrix completion. The key is to replace Theorem 5 below with the corresponding theorem derived under other assumptions. With appropriate replacement of Theorem 5, we should still be able to obtain a relative error bound, of course with different dependence on m, n, r, and $|\Omega|$. For example, using the assumptions and theorems in Bhojanapalli and Jain (2014), we can generalize our result to a *universal* guarantee for full-matrix completion. We leave the extension of our analysis to other assumptions as a future work.

2. This can be easily verified because our analysis only requires the eigenspaces of A_r satisfy the incoherence condition.

4.1. Sketch of the Proof

As we mentioned before, our analysis is built upon the existing theoretical guarantee for low-rank matrix completion, which is summarized below (Recht, 2011).

Theorem 5 *Suppose*

$$|\Omega| \geq 32 \max(\mu_0, \mu_1^2) r(m+n)\beta \log^2(2n) \tag{14}$$

for some $\beta > 1$. Then, with a probability at least $1 - 6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}}$, the following statements are true:

- $$\left\| \frac{mn}{|\Omega|} \mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T - \mathcal{P}_T \right\| \leq \frac{1}{2}. \tag{15}$$
- $$\|\mathcal{R}_\Omega\| \leq \frac{8}{3}\sqrt{\beta}\log(n). \tag{16}$$
- *There exists a $Y \in \mathbb{R}^{m\times n}$ in the range of $\mathcal{R}_\Omega$ such that*

$$\left\| \mathcal{P}_T(Y) - UV^\top \right\|_F \leq \sqrt{\frac{r}{2n}}, \tag{17}$$

$$\left\| \mathcal{P}_{T^\perp(Y)} \right\| \leq \frac{1}{2}, \tag{18}$$

$$|\langle Y, A\rangle| \leq \sqrt{\frac{3mnr\log(2n)}{8|\Omega|}} \sqrt{\langle \mathcal{R}_\Omega(A), A\rangle}, \tag{19}$$

for all $A \in \mathbb{R}^{m\times n}$.

The first part of above theorem contains concentration inequalities for the random linear operator $\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T$ and $\mathcal{R}_\Omega$, and the second part describes some important properties of a special matrix Y, which is used as an (approximate) dual certificate of (1).

Next, we will examine the optimality of B_* based on techniques from convex analysis, leading to the following theorem.

Theorem 6 *Let B_* be the optimal solution to (3), we have*

$$\lambda\langle B_* - A_r, UV^\top\rangle + \lambda\|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \langle \mathcal{R}_\Omega(B_* - A), A_r - B_*\rangle. \tag{20}$$

Based on Theorems 5 and 6, we are ready to prove the main results. However, the analysis is a bit lengthy, so we split it into two parts, and will first show the following intermediate theorem.

Theorem 7 *Suppose (14) holds. With a probability at least $1 - 6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}}$, we have*

$$\begin{aligned} &\frac{1}{2}\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_*\rangle + \frac{\lambda}{2}\|\mathcal{P}_{T^\perp}(B_*)\|_* \\ \leq &\lambda\sqrt{\frac{r}{2n}}\|\mathcal{P}_T(A_r - B_*)\|_F + \langle \mathcal{R}_\Omega(A - A_r), A - A_r\rangle + \frac{3mnr\log(2n)\lambda^2}{8|\Omega|}. \end{aligned} \tag{21}$$

Then, we can prove Theorem 1 by further lower bounding and upper bounding the L.H.S. and R.H.S. of (21), respectively. If the residual matrix $A - A_r$ is not too spiky, we can apply Bernstein's inequality to derive a tighter bound for $\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle$ and obtain Theorem 3 in a similar way.

4.2. Property of the Linear Operator $\mathcal{R}_\Omega$

Before going to the detail, we first introduce a lemma that will be used throughout the analysis. Since Ω may contain duplicate indices, $\langle \mathcal{R}_\Omega(A), A \rangle \neq \|\mathcal{R}_\Omega(A)\|_F^2$ in general. We use the following lemma to take care of this issue.

Lemma 1

$$\langle \mathcal{R}_\Omega(A), A \rangle \leq \|\mathcal{R}_\Omega(A)\|_F^2, \tag{22}$$

$$|\langle \mathcal{R}_\Omega(A), B \rangle| \leq \sqrt{\langle \mathcal{R}_\Omega(A), A \rangle}\sqrt{\langle \mathcal{R}_\Omega(B), B \rangle} \tag{23}$$

for all $A, B \in \mathbb{R}^{m \times n}$.

Proof Denote the number of unique indices in Ω by u, and let $\Theta = \{(a_k, b_k)\}_{k=1}^u$ be a *set* that contains all the unique indices in Ω. Let t_k denote the times that (a_k, b_k) appears in Ω. Then, we have

$$\langle \mathcal{R}_\Omega(A), A \rangle = \sum_{k=1}^{u} t_k A_{a_k b_k}^2 \leq \sum_{k=1}^{u} t_k^2 A_{a_k b_k}^2 = \|\mathcal{R}_\Omega(A)\|_F^2.$$

To show (23), we have

$$|\langle \mathcal{R}_\Omega(A), B \rangle| = \left| \sum_{(i,j)\in\Omega} A_{ij} B_{ij} \right| \leq \sqrt{\sum_{(i,j)\in\Omega} A_{ij}^2}\sqrt{\sum_{(i,j)\in\Omega} B_{ij}^2} = \sqrt{\langle \mathcal{R}_\Omega(A), A \rangle}\sqrt{\langle \mathcal{R}_\Omega(B), B \rangle}$$

where the inequality is due to Cauchy–Schwarz inequality. ■

4.3. Proof of Theorem 5

Except for the last inequality in (19), all the others can be found directly from Section 4 of Recht (2011). Thus, we only provide the derivation of (19), which is based on some intermediate results of Recht (2011).

We first state those intermediate results. Following the construction of Recht (2011, Section 4), we partition Ω into p partitions of size q. By assumption, we can choose

$$q \geq \frac{128}{3} \max(\mu_0, \mu_1^2) r (m+n) \beta \log(m+n)$$

such that

$$p = \frac{\Omega}{q} = \frac{3}{4} \log 2n.$$

Let Ω_j denote the set of indices corresponding to the j-th partition. We define $W_0 = UV^\top$,

$$Y_k = \frac{mn}{q}\sum_{j=1}^{k}\mathcal{R}_{\Omega_j}(W_{j-1}), \text{ and } W_k = UV^\top - \mathcal{P}_T(Y_k)$$

for $k = 1, \ldots, p$. Then, we set $Y = Y_p$. Recht (2011) has proved that with a probability at least $1 - 6\log(n)(m+n)^{2-2\beta} - n^{2-2\beta^{1/2}}$,

$$\|W_k\|_F \leq 2^{-k}\sqrt{r}, \tag{24}$$

$$\left\|\frac{mn}{q}\mathcal{P}_T\mathcal{R}_{\Omega_k}\mathcal{P}_T - \mathcal{P}_T\right\| \leq \frac{1}{2}, \tag{25}$$

for $k = 1, \ldots, p$.

We proceed to prove (19). Since $W_j = \mathcal{P}_T(W_j)$, we have

$$\begin{aligned}
&\frac{mn}{q}\langle\mathcal{R}_{\Omega_j}(W_j), W_j\rangle = \left\langle W_j, \frac{mn}{q}\mathcal{P}_T\mathcal{R}_{\Omega_j}\mathcal{P}_T(W_j)\right\rangle \\
\overset{(25)}{\leq}& \frac{3}{2}\left\|\mathcal{P}_T(W_j)\right\|_F^2 = \frac{3}{2}\|W_j\|_F^2 \overset{(24)}{\leq} \frac{3r}{2}4^{-j}.
\end{aligned} \tag{26}$$

Then,

$$\begin{aligned}
|\langle Y, A\rangle| \leq & \frac{mn}{q}\sum_{j=1}^{p}\left|\langle\mathcal{R}_{\Omega_j}(W_{j-1}), A\rangle\right| \\
\overset{(23)}{\leq}& \frac{mn}{q}\sum_{j=1}^{p}\sqrt{\langle\mathcal{R}_{\Omega_j}(W_{j-1}), W_{j-1}\rangle}\sqrt{\langle\mathcal{R}_{\Omega_j}(A), A\rangle} \\
\leq & \frac{mn}{q}\sqrt{\sum_{j=1}^{p}\langle\mathcal{R}_{\Omega_j}(W_{j-1}), W_{j-1}\rangle}\sqrt{\sum_{j=1}^{p}\langle\mathcal{R}_{\Omega_j}(A), A\rangle} \\
= & \sqrt{\frac{mn}{q}}\sqrt{\sum_{j=1}^{p}\left\langle\frac{mn}{q}\mathcal{R}_{\Omega_j}(W_{j-1}), W_{j-1}\right\rangle}\sqrt{\langle\mathcal{R}_{\Omega}(A), A\rangle} \\
\overset{(26)}{\leq}& \sqrt{\frac{mn}{q}}\sqrt{\langle\mathcal{R}_{\Omega}(A), A\rangle}\sqrt{\frac{3r}{2}\sum_{j=1}^{p}4^{-j}} \\
\leq & \sqrt{\frac{mnr}{2q}}\sqrt{\langle\mathcal{R}_{\Omega}(A), A\rangle} = \sqrt{\frac{mnpr}{2|\Omega|}}\sqrt{\langle\mathcal{R}_{\Omega}(A), A\rangle}.
\end{aligned}$$

4.4. Proof of Theorem 6

Since B_* is the optimal solution to (3), we have

$$\langle\mathcal{R}_{\Omega}(B_* - A) + \lambda E, A_r - B_*\rangle \geq 0 \tag{27}$$

where $E \in \partial\|B_*\|_*$ is *certain* subgradient of $\|\cdot\|_*$ evaluated at B_*. Let $F \in \partial\|A_r\|_*$ be *any* subgradient of $\|\cdot\|_*$ evaluated at A_r. From the property of convexity, we have

$$\langle B_* - A_r, E - F\rangle \geq 0. \tag{28}$$

From (27) and (28), we get

$$\langle \mathcal{R}_\Omega(B_* - A) + \lambda F, A_r - B_*\rangle \geq 0. \tag{29}$$

Next, we consider bounding $\lambda\langle F, A_r - B_*\rangle$. From previous studies (Candès and Recht, 2009), we know that the set of subgradients of $\|A_r\|_*$ takes the following form:

$$\partial\|A_r\|_* = \left\{UV^\top + W : W \in \mathbb{R}^{m\times n}, U^\top W = 0, WV = 0, \|W\| \leq 1\right\}.$$

Thus, we can choose

$$F = UV^\top + \mathcal{P}_{T^\perp}(N),$$

where $N = \mathrm{argmax}_{\|X\|\leq 1}\langle \mathcal{P}_{T^\perp}(B_*), X\rangle$. Then, it is easy to verify that

$$\begin{aligned}\langle B_* - A_r, F\rangle =& \langle B_* - A_r, UV^\top\rangle + \langle B_* - A_r, \mathcal{P}_{T^\perp}(N)\rangle \\ =& \langle B_* - A_r, UV^\top\rangle + \|\mathcal{P}_{T^\perp}(B_*)\|_*.\end{aligned} \tag{30}$$

We complete the proof by combining (29) and (30).

4.5. Proof of Theorem 7

We continue the proof by lower bounding $\langle B_* - A_r, UV^\top\rangle$ in (20) of Theorem 6. To this end, we need the matrix Y given in Theorem 5.

$$\begin{aligned}\langle B_* - A_r, UV^\top\rangle =& \langle B_* - A_r, UV^\top - Y\rangle + \langle B_* - A_r, Y\rangle \\ =& \langle B_* - A_r, UV^\top - \mathcal{P}_T(Y)\rangle + \langle A_r - B_*, \mathcal{P}_{T^\perp}(Y)\rangle + \langle B_* - A_r, Y\rangle.\end{aligned}$$

Next, we bound the last three terms by utilizing the conclusions in Theorem 5.

$$\begin{aligned}&\langle B_* - A_r, UV^\top - \mathcal{P}_T(Y)\rangle = \langle \mathcal{P}_T(B_* - A_r), UV^\top - \mathcal{P}_T(Y)\rangle \\ \geq & -\|\mathcal{P}_T(B_* - A_r)\|_F\|UV^\top - \mathcal{P}_T(Y)\|_F \overset{(17)}{\geq} -\sqrt{\frac{r}{2n}}\|\mathcal{P}_T(A_r - B_*)\|_F.\end{aligned}$$

$$\begin{aligned}&\langle A_r - B_*, \mathcal{P}_{T^\perp}(Y)\rangle = \langle \mathcal{P}_{T^\perp}(A_r - B_*), \mathcal{P}_{T^\perp}(Y)\rangle \\ =& \langle \mathcal{P}_{T^\perp}(-B_*), \mathcal{P}_{T^\perp}(Y)\rangle \geq -\|\mathcal{P}_{T^\perp}(B_*)\|_*\|\mathcal{P}_{T^\perp}(Y)\| \overset{(18)}{\geq} -\frac{1}{2}\left\|\mathcal{P}_{T^\perp}(B_*)\right\|_*.\end{aligned}$$

$$\langle B_* - A_r, Y\rangle \overset{(19)}{\geq} -\sqrt{\frac{3mnr\log(2n)}{8|\Omega|}}\sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_*\rangle}.$$

Putting the above inequalities together, we have

$$\begin{aligned}\langle B_* - A_r, UV^\top\rangle \geq & -\sqrt{\frac{r}{2n}}\|\mathcal{P}_T(A_r - B_*)\|_F - \frac{1}{2}\left\|\mathcal{P}_{T^\perp}(B_*)\right\|_* \\ & - \sqrt{\frac{3mnr\log(2n)}{8|\Omega|}}\sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_*\rangle}.\end{aligned} \tag{31}$$

Substituting (31) into (20) and rearranging, we get

$$\begin{aligned}
&\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \frac{\lambda}{2} \left\| \mathcal{P}_{T^\perp}(B_*) \right\|_* \\
\leq &\langle \mathcal{R}_\Omega(A_r - A), A_r - B_* \rangle + \lambda \sqrt{\frac{r}{2n}} \| \mathcal{P}_T(A_r - B_*) \|_F \\
&+ \lambda \sqrt{\frac{3mnr\log(2n)}{8|\Omega|}} \sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle} \\
\overset{(23)}{\leq} &\sqrt{\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle} \sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle} + \lambda \sqrt{\frac{r}{2n}} \| \mathcal{P}_T(A_r - B_*) \|_F \\
&+ \lambda \sqrt{\frac{3mnr\log(2n)}{8|\Omega|}} \sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle}.
\end{aligned} \tag{32}$$

From the the basic inequality $\frac{1}{4}\alpha^2 - \alpha\beta + \beta^2 \geq 0$, we have

$$\begin{aligned}
&\sqrt{\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle} \sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle} \\
\leq &\frac{1}{4} \langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle,
\end{aligned} \tag{33}$$

$$\begin{aligned}
&\lambda \sqrt{\frac{3mnr\log(2n)}{8|\Omega|}} \sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle} \\
\leq &\frac{1}{4} \langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \lambda^2 \frac{3mnr\log(2n)}{8|\Omega|}.
\end{aligned} \tag{34}$$

We complete the proof by summing (32), (33), and (34) together.

4.6. Proof of Theorem 1

The lower bound of $|\Omega|$ in (10) is due to Theorem 5, but we use a larger constant (114 instead of 32) to ensure

$$\frac{8\log(n)}{3} \sqrt{\frac{rm\beta}{|\Omega|}} \leq \frac{1}{4} \tag{35}$$

which is used later.

Based on Lemma 1 and Theorem 5, we have

$$\sqrt{\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle} \overset{(22)}{\leq} \| \mathcal{R}_\Omega(A - A_r) \|_F \overset{(16)}{\leq} \frac{8}{3} \sqrt{\beta} \log(n) \varepsilon.$$

Substituting the above inequality into (21), we have

$$\frac{1}{2} \langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \frac{\lambda}{2} \left\| \mathcal{P}_{T^\perp}(B_*) \right\|_* \leq \lambda \sqrt{\frac{r}{2n}} \| \mathcal{P}_T(A_r - B_*) \|_F + \Gamma, \tag{36}$$

where

$$\Gamma = \frac{64\beta \log^2(n) \varepsilon^2}{9} + \frac{3mnr\log(2n)\lambda^2}{8|\Omega|}. \tag{37}$$

4.6.1. Upper Bound for $\|\mathcal{P}_{T^\perp}(B_*)\|_F$

We upper bound $\|\mathcal{P}_T(A_r - B_*)\|_F^2$ in (36) by

$$\|\mathcal{P}_T(A_r - B_*)\|_F^2 = \langle \mathcal{P}_T(A_r - B_*), A_r - B_* \rangle \overset{(15)}{\leq} 2\frac{mn}{|\Omega|}\langle \mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T(A_r - B_*), A_r - B_* \rangle.$$

Plugging the above inequality in (36), we have

$$\begin{aligned} &\frac{1}{2}\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \frac{\lambda}{2}\|\mathcal{P}_{T^\perp}(B_*)\|_* \\ \leq &\lambda\sqrt{\frac{rm}{|\Omega|}}\sqrt{\langle \mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T(A_r - B_*), A_r - B_* \rangle} + \Gamma. \end{aligned} \tag{38}$$

Since $\mathcal{P}_T + \mathcal{P}_{T^\perp} = I$, we have

$$\begin{aligned} &\frac{1}{2}\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle \\ =&\frac{1}{2}\underbrace{\langle \mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T(A_r - B_*), A_r - B_* \rangle}_{:=\Theta^2} + \frac{1}{2}\underbrace{\langle \mathcal{P}_{T^\perp} \mathcal{R}_\Omega \mathcal{P}_{T^\perp}(B_*), B_* \rangle}_{:=\Lambda^2} \\ &- \langle \mathcal{R}_\Omega(\mathcal{P}_T(A_r - B_*)), \mathcal{P}_{T^\perp}(B_*) \rangle \\ \overset{(23)}{\geq}&\frac{1}{2}\Theta^2 + \frac{1}{2}\Lambda^2 - \Theta\Lambda = \frac{1}{2}(\Theta - \Lambda)^2. \end{aligned} \tag{39}$$

Substituting (39) into (38), we have

$$\frac{1}{2}(\Theta - \Lambda)^2 + \frac{\lambda}{2}\|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \lambda\sqrt{\frac{rm}{|\Omega|}}\Theta + \Gamma.$$

Combining with the fact

$$\frac{1}{2}(\Theta - \Lambda)^2 - \lambda\sqrt{\frac{rm}{|\Omega|}}\Theta + \lambda\sqrt{\frac{rm}{|\Omega|}}\Lambda + \frac{rm\lambda^2}{2|\Omega|} = \frac{1}{2}\left(\Theta - \Lambda - \lambda\sqrt{\frac{rm}{|\Omega|}}\right)^2 \geq 0$$

we have

$$\begin{aligned} \frac{\lambda}{2}\|\mathcal{P}_{T^\perp}(B_*)\|_* \leq &\lambda\sqrt{\frac{rm}{|\Omega|}}\sqrt{\langle \mathcal{P}_{T^\perp} \mathcal{R}_\Omega \mathcal{P}_{T^\perp}(B_*), B_* \rangle} + \frac{rm\lambda^2}{2|\Omega|} + \Gamma \\ \overset{(22)}{\leq}& \lambda\sqrt{\frac{rm}{|\Omega|}}\|\mathcal{R}_\Omega \mathcal{P}_{T^\perp}(B_*)\|_F + \frac{rm\lambda^2}{2|\Omega|} + \Gamma \\ \overset{(16)}{\leq}& \frac{8\lambda\log(n)}{3}\sqrt{\frac{rm\beta}{|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F + \frac{rm\lambda^2}{2|\Omega|} + \Gamma \\ \overset{(35)}{\leq}& \frac{\lambda}{4}\|\mathcal{P}_{T^\perp}(B_*)\|_F + \frac{rm\lambda^2}{2|\Omega|} + \Gamma \\ \overset{(37)}{\leq}& \frac{\lambda}{4}\|\mathcal{P}_{T^\perp}(B_*)\|_* + \frac{64\beta\log^2(n)\varepsilon^2}{9} + \frac{3mnr\log(2n)\lambda^2}{4|\Omega|} \end{aligned} \tag{40}$$

where in the last line we use the fact

$$\frac{1}{2} \leq \frac{3n\log(2n)}{8}, \ \forall n \geq 2.$$

From (40), we immediately have

$$\|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \frac{256\beta\log^2(n)\varepsilon^2}{9\lambda} + \frac{3mnr\log(2n)\lambda}{|\Omega|}.$$

4.6.2. Upper Bound for $\|\mathcal{P}_T(A_r - B_*)\|_F$

Similar to (39), we have

$$\begin{aligned}
&\frac{1}{2}\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_*\rangle \\
\overset{(23)}{\geq}&\frac{1}{2}\langle \mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(A_r - B_*), A_r - B_*\rangle + \frac{1}{2}\Lambda^2 - \sqrt{\langle \mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(A_r - B_*), A_r - B_*\rangle}\Lambda \\
\overset{(15)}{\geq}&\frac{|\Omega|}{4mn}\|\mathcal{P}_T(A_r - B_*)\|_F^2 + \frac{1}{2}\Lambda^2 - \sqrt{\frac{3|\Omega|}{2mn}}\|\mathcal{P}_T(A_r - B_*)\|_F\Lambda
\end{aligned}$$

where

$$\Lambda = \sqrt{\langle \mathcal{P}_{T^\perp}\mathcal{R}_\Omega\mathcal{P}_{T^\perp}(B_*), B_*\rangle} \overset{(22)}{\leq} \|\mathcal{R}_\Omega\mathcal{P}_{T^\perp}(B_*)\|_F \overset{(16)}{\leq} \frac{8}{3}\sqrt{\beta}\log(n)\|\mathcal{P}_{T^\perp}(B_*)\|_F.$$

By plugging the above inequalities into (36), we have

$$\begin{aligned}
&\frac{|\Omega|}{4mn}\|\mathcal{P}_T(A_r - B_*)\|_F^2 + \frac{1}{2}\Lambda^2 + \frac{\lambda}{2}\|\mathcal{P}_{T^\perp}(B_*)\|_* \\
\leq&\lambda\sqrt{\frac{r}{2n}}\|\mathcal{P}_T(A_r - B_*)\|_F + \Gamma + 8\log(n)\sqrt{\frac{|\Omega|\beta}{6mn}}\|\mathcal{P}_{T^\perp}(B_*)\|_F\|\mathcal{P}_T(A_r - B_*)\|_F
\end{aligned}$$

and thus

$$\begin{aligned}
\|\mathcal{P}_T(A_r - B_*)\|_F^2 \overset{(37)}{\leq}& \frac{2m\lambda\sqrt{2rn}}{|\Omega|}\|\mathcal{P}_T(A_r - B_*)\|_F + \frac{256\beta mn\log^2(n)\varepsilon^2}{9|\Omega|} + \frac{3rm^2n^2\lambda^2\log(2n)}{2|\Omega|^2} \\
&+ 32\log(n)\sqrt{\frac{mn\beta}{6|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F\|\mathcal{P}_T(A_r - B_*)\|_F.
\end{aligned} \tag{41}$$

Recall that

$$x^2 \leq bx + c \Rightarrow x \leq 2b + \sqrt{2c}.$$

From (41), we have

$$\begin{aligned}
\|\mathcal{P}_T(A_r - B_*)\|_F \leq& \frac{4m\lambda\sqrt{2rn}}{|\Omega|} + 64\log(n)\sqrt{\frac{mn\beta}{6|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F \\
&+ \frac{16\log(n)\varepsilon}{3}\sqrt{\frac{2\beta mn}{|\Omega|}} + \frac{mn\lambda}{|\Omega|}\sqrt{3r\log(2n)}.
\end{aligned}$$

We complete the proof by noticing

$$4\sqrt{2n} \leq n\sqrt{3\log(2n)}, \ \forall n \geq 5.$$

4.7. Proof of Theorem 3

With the second lower bound of $|\Omega|$ in (13), we can prove the following upper bound for $\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle$.

Lemma 2 *Suppose*

$$|\Omega| \geq \frac{8mn\|A - A_r\|_\infty^2}{3\|A - A_r\|_F^2} \beta \log(n) \tag{42}$$

for some $\beta > 1$. Then, with a probability at least $1 - n^{-\beta}$, we have

$$\sqrt{\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle} \leq \varepsilon \sqrt{\frac{2|\Omega|}{mn}}. \tag{43}$$

Proof For each index $(a_k, b_k) \in \Omega$, we define a random variable

$$\xi_k = \langle \mathbf{e}_{a_k} \mathbf{e}_{b_k}^\top, A - A_r \rangle^2 - \frac{1}{mn} \|A - A_r\|_F^2.$$

Then, it is easy to verify that

$$\begin{aligned}
\mathrm{E}[\xi_k] =& 0, \\
|\xi_k| =& \left| \langle \mathbf{e}_{a_k} \mathbf{e}_{b_k}^\top, A - A_r \rangle^2 - \frac{1}{mn} \|A - A_r\|_F^2 \right| \\
\leq & \max \left(\langle \mathbf{e}_{a_k} \mathbf{e}_{b_k}^\top, A - A_r \rangle^2, \frac{1}{mn} \|A - A_r\|_F^2 \right) \leq \|A - A_r\|_\infty^2, \\
\mathrm{E}[\xi_k^2] =& \mathrm{E}\left[\langle \mathbf{e}_{a_k} \mathbf{e}_{b_k}^\top, A - A_r \rangle^4 \right] - \frac{1}{m^2 n^2} \|A - A_r\|_F^4 \\
\leq & \frac{1}{mn} \sum_{i,j} [A - A_r]_{ij}^4 \leq \frac{1}{mn} \|A - A_r\|_\infty^2 \|A - A_r\|_F^2.
\end{aligned}$$

From Bernstein's inequality, we have

$$\begin{aligned}
& \mathrm{P}\left[\langle \mathcal{R}_\Omega(A - A_r), A - A_r \rangle \geq 2 \frac{|\Omega|}{mn} \|A - A_r\|_F^2 \right] \\
= & \mathrm{P}\left[\sum_{k=1}^{|\Omega|} \xi_k \geq \frac{|\Omega|}{mn} \|A - A_r\|_F^2 \right] \leq \exp\left(-\frac{3|\Omega|}{8\|A - A_r\|_\infty^2} \frac{1}{mn} \|A - A_r\|_F^2 \right) \overset{(42)}{\leq} n^{-\beta}.
\end{aligned}$$

■

Following the derivation of (36), we have

$$\frac{1}{2} \langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_* \rangle + \frac{\lambda}{2} \left\| \mathcal{P}_{T^\perp}(B_*) \right\|_* \leq \lambda \sqrt{\frac{r}{2n}} \|\mathcal{P}_T(A_r - B_*)\|_F + \Gamma',$$

where

$$\Gamma' = \frac{2|\Omega|\varepsilon^2}{mn} + \frac{3mnr \log(2n) \lambda^2}{8|\Omega|}.$$

The rest of the analysis is almost identical to that of Theorem 1. In particular, (40) becomes

$$\frac{\lambda}{2} \|\mathcal{P}_{T^\perp}(B_*)\|_* \leq \frac{\lambda}{4}\|\mathcal{P}_{T^\perp}(B_*)\|_* + \frac{2|\Omega|\varepsilon^2}{mn} + \frac{3mnr\log(2n)\lambda^2}{4|\Omega|},$$

and (41) becomes

$$\begin{aligned}\|\mathcal{P}_T(A_r - B_*)\|_F^2 \leq & \frac{2m\lambda\sqrt{2rn}}{|\Omega|}\|\mathcal{P}_T(A_r - B_*)\|_F + 8\varepsilon^2 + \frac{3rm^2n^2\lambda^2\log(2n)}{2|\Omega|^2} \\ & + 32\log(n)\sqrt{\frac{mn\beta}{6|\Omega|}}\|\mathcal{P}_{T^\perp}(B_*)\|_F\|\mathcal{P}_T(A_r - B_*)\|_F.\end{aligned}$$

A complete proof can be found in an early version of this paper (Zhang et al., 2015).

5. Conclusion and Future Work

In this paper, we develop a relative error bound for the nuclear norm regularized matrix completion, under the assumption that the top eigenspaces of the target matrix are incoherent. To the best of our knowledge, this is the first work toward relative error bound for nuclear norm regularized matrix completion, and an extensive comparison shows that our bound is tighter than previous results under favored conditions.

In many real-world applications, it is appropriate to assume the observed entries are corrupted by noise. As we discussed in the end of Section 3.2, it is possible to extend our analysis to the noisy case. More specifically, let N be the matrix of noise. We just need to add $\langle \mathcal{R}_\Omega(N), A_r - B_*\rangle$ to the R.H.S. of (20), which leads to an additional term $\sqrt{\langle \mathcal{R}_\Omega(N), N\rangle}\sqrt{\langle \mathcal{R}_\Omega(A_r - B_*), A_r - B_*\rangle}$ in the R.H.S. of (32). The rest of the proof is almost the same, and finally we will obtain an upper bound that depends on both $A - A_r$ and N.

Acknowledgments

This work was partially supported by the National Key R&D Program of China (2018YFB-1004300), NSFC-NRF Joint Research Project (61861146001), JiangsuSF (BK20160658), YESS (2017QNRC001), and the Collaborative Innovation Center of Novel Software Technology and Industrialization.

Appendix A. Derivation of (7)

Let $\widehat{B}$ be the solution found by the algorithm in Foygel and Srebro (2011). Let $\epsilon > 0$ be the mean-squared reconstruction error. Using the notations in this paper, (5) of Foygel and Srebro (2011) becomes

$$\frac{1}{mn}\|\widehat{B} - A\|_F^2 \leq \frac{1}{mn}\|A - A_r\|_F^2 + \epsilon \tag{44}$$

under the condition

$$|\Omega| \geq O\left(\frac{r(n+m)}{\epsilon^2}\left(\epsilon + \frac{1}{mn}\|A - A_r\|_F^2\right)\right) \tag{45}$$

where logarithmic factors are ignored. (45) can be rewritten as

$$|\Omega|\epsilon^2 \geq O\left(r(n+m)\epsilon + \frac{r(n+m)}{mn}\|A - A_r\|_F^2\right).$$

Since we assume $m \leq n$, it can be further simplified to

$$|\Omega|\epsilon^2 \geq O\left(rn\epsilon + \frac{r}{m}\|A - A_r\|_F^2\right).$$

Let consider the optimal case, i.e.,

$$|\Omega|\epsilon^2 = Crn\epsilon + C\frac{r}{m}\|A - A_r\|_F^2$$

for some constant $C > 0$. Then, we have

$$\begin{aligned}\epsilon =& \frac{Crn + \sqrt{C^2r^2n^2 + 4|\Omega|C\frac{r}{m}\|A - A_r\|_F^2}}{2|\Omega|} \\ \leq& \frac{Crn + \sqrt{|\Omega|C\frac{r}{m}\|A - A_r\|_F^2}}{|\Omega|} = O\left(\frac{rn}{|\Omega|} + \sqrt{\frac{r}{m|\Omega|}\|A - A_r\|_F^2}\right).\end{aligned}$$

As a result, (44) becomes

$$\frac{1}{mn}\|\widehat{B} - A\|_F^2 \leq O\left(\frac{1}{mn}\|A - A_r\|_F^2 + \frac{rn}{|\Omega|} + \sqrt{\frac{r}{m|\Omega|}\|A - A_r\|_F^2}\right)$$

which implies

$$\|\widehat{B} - A\|_F^2 \leq O\left(\|A - A_r\|_F^2 + \frac{rmn^2}{|\Omega|} + \sqrt{\frac{rmn^2}{|\Omega|}\|A - A_r\|_F^2}\right)$$

from which we obtain (7).

References

Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

Pratik Biswas, Tzu-Chen Lian, Ta-Chung Wang, and Yinyu Ye. Semidefinite programming based algorithms for sensor network localization. *ACM Transactions on Sensor Networks*, 2(2):188–220, 2006.

Christos Boutsidis, Michael W. Mahoney, and Petros Drineas. An improved approximation algorithm for the column subset selection problem. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 968–977, 2009.

Ricardo S. Cabral, Fernando Torre, Joao P. Costeira, and Alexandre Bernardino. Matrix completion for multi-label image classification. In *Advances in Neural Information Processing Systems 24*, pages 190–198, 2011.

Tony Cai and Wen-Xin Zhou. A max-norm constrained minimization approach to 1-bit matrix completion. *Journal of Machine Learning Research*, 14:3619–3647, 2013.

Emmanuel J. Candès and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Yudong Chen, Srinadh Bhojanapalli, Sujay Sanghavi, and Rachel Ward. Coherent matrix completion. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211C–231, 2009.

Brian Eriksson, Laura Balzano, and Robert Nowak. High-rank matrix completion. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, pages 373–381, 2012.

Rina Foygel and Nathan Srebro. Concentration-based guarantees for low-rank matrix reconstruction. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 315–340, 2011.

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory,*, 57(3):1548–1566, 2011.

Ke Hou, Zirui Zhou, Anthony Man-Cho So, and Zhi-Quan Luo. On the linear convergence of the proximal gradient method for trace norm regularization. In *Advances in Neural Information Processing Systems 26*, pages 710–718, 2013.

Cho-Jui Hsieh and Peder Olsen. Nuclear norm minimization via active subspace selection. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

Ali Jalali, Yudong Chen, Sujay Sanghavi, and Huan Xu. Clustering partially observed graphs via convex optimization. In *Proceedings of the 28th International Conference on Machine Learning*, pages 1001–1008, 2011.

Shuiwang Ji and Jieping Ye. An accelerated gradient method for trace norm minimization. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 457–464, 2009.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010a.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *Journal of Machine Learning Research*, 11:2057–2078, 2010b.

Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20:282–303, 2014.

Vladimir Koltchinskii, Karim Lounici, and Alexandre B. Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

Lester W. Mackey, Michael I. Jordan, and Ameet Talwalkar. Divide-and-conquer matrix factorization. In *Advances in Neural Information Processing Systems 24*, pages 1134–1142, 2011.

Sahand Negahban and Martin J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13:1665–1697, 2012.

Yu. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

Ting Kei Pong, Paul Tseng, Shuiwang Ji, and Jieping Ye. Trace norm regularization: Reformulations, algorithms, and multi-task learning. *SIAM Journal on Optimization*, 20 (6):3465–3489, 2010.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011.

Ruoyu Sun and Zhi-Quan Luo. Guaranteed matrix completion via non-convex factorization. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pacific Journal of Optimization*, 6(3):615–640, 2010.

Miao Xu, Rong Jin, and Zhi-Hua Zhou. Speedup matrix completion with side information: Application to multi-label learning. In *Advances in Neural Information Processing Systems 26*, pages 2301–2309, 2013.

Lijun Zhang, Tianbao Yang, Rong Jin, and Zhi-Hua Zhou. Analysis of nuclear norm regularization for full-rank matrix completion. *ArXiv e-prints*, arXiv:1504.06817, 2015.

Xinhua Zhang, Yaoliang Yu, and Dale Schuurmans. Accelerated training for matrix-norm regularization: A boosting approach. In *Advances in Neural Information Processing Systems 25*, pages 2906–2914, 2012.