CONVERGENCE AND STABILITY OF ITERATIVELY
REWEIGHTED LEAST SQUARES FOR LOW-RANK
MATRIX RECOVERY

YUN CAI
Department of Mathematics, Nanjing University of Chinese Medicine
Nanjing, Jiangsu 210023, China

SONG LI*
School of Mathematical Science, Zhejiang University
Hangzhou, Zhejiang 310027, China

(Communicated by Hao-Min Zhou)

Abstract. In this paper, we study the theoretical properties of iteratively
reweighted least squares algorithm for recovering a matrix (IRLS-M for short)
from noisy linear measurements. The IRLS-M was proposed by Fornasier
et al. (2011) [17] for solving nuclear norm minimization and by Mohan
et al. (2012) [31] for solving Schatten-\(p\) (quasi) norm minimization \(0 < p \leq 1\)
in noiseless case, based on the iteratively reweighted least squares algorithm
for sparse signal recovery (IRLS for short) (Daubechies et al., 2010) [15], and
numerical experiments have been given to show its efficiency (Fornasier et
al. and Mohan et al.) [17, 31]. In this paper, we focus on providing con-
vergence and stability analysis of iteratively reweighted least squares algorithm
for low-rank matrix recovery in the presence of noise. The convergence of
IRLS-M is proved strictly for all \(0 < p \leq 1\). Furthermore, when the measurement map \(\mathcal{A}\)
satisfies the matrix restricted isometry property (M-RIP for short), we show
that the IRLS-M is stable for \(0 < p \leq 1\). Specially, when \(p = 1\), we prove
that the M-RIP constant \(\delta_{2r} < \sqrt{2} - 1\) is sufficient for IRLS-M to recover
an unknown (approximately) low rank matrix with an error that is propor-
tional to the noise level. The simplicity of IRLS-M, along with the theoretical
guarantees provided in this paper, make a compelling case for its adoption as a
standard tool for low rank matrix recovery.

1. Introduction.

1.1. Low rank matrix recovery. Low rank matrix recovery is a burgeoning topic
which has drawn attention of many researchers in different fields in the past ser-
val years. It has wide applications in quantum state tomography [21, 22], image
processing [2], system identification [29], recovering shape and motion from image
streams [32], etc. The fundamental problem of low rank matrix recovery is to recon-
struct an unknown (approximately) low rank matrix \(M \in \mathbb{R}^{n_1 \times n_2}\) from its linear
measurements \(b = \mathcal{A}(M) + z\), where \(\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m\) \((m < n_1 n_2)\) is a linear
measurement operator and \(z \in \mathbb{R}^m\) is the noise vector bounded by \(\|z\|_2 \leq \eta\). The direct
method to tackle this inverse problem is to solve the following rank minimization problem

\[
\min_{X \in \mathbb{R}^{n_1 \times n_2}} \text{rank}(X) \text{ subject to } \|A(X) - b\|_2 \leq \eta.
\]

A special case of rank minimization is the matrix completion problem which aims to recover a matrix from a small sample of its entries

\[
\min_{X \in \mathbb{R}^{n_1 \times n_2}} \text{rank}(X) \text{ subject to } \|\mathcal{P}_\Omega(X) - Y\|_F \leq \eta,
\]

where \(\mathcal{P}_\Omega(Y) = \mathcal{P}_\Omega(M) + \mathcal{P}_\Omega(Z)\) with \(Z\) being the noise matrix and satisfying \(\|\mathcal{P}_\Omega(Z)\|_F \leq \eta\), and \(\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}\) denotes the projection operator onto the index set \(\Omega \subset [1, \ldots, n_1] \times [1, \ldots, n_2]\), i.e., \((\mathcal{P}_\Omega(X))_{ij} = X_{ij}\) for \((i, j) \in \Omega\) and \((\mathcal{P}_\Omega(X))_{ij} = 0\), otherwise. Matrix completion problem has applications in online recommendation systems including the famous Netflix problem \[33\], video restoration \[24\], etc.

Note that when \(X\) is restricted to be a diagonal matrix, the rank minimization problem (1) reduces to the following sparse signal recovery problem in compressed sensing

\[
\min \|x\|_0 \text{ subject to } \|Ax - b\|_2 \leq \eta,
\]

where \(\|x\|_0\) denotes the number of non-zero entries of vector \(x\) \[7, 8\]. Therefore, rank minimization problem (1) includes the sparse signal recovery problem (2) as a special case. Unfortunately, the rank minimization problem (1) is NP-hard and thus computationally infeasible. To circumvent this intractability, a popular approach is to consider its relaxation problem, i.e., the following Schatten-\(p\) norm minimization problem (0 < \(p\) ≤ 1)

\[
\min_{X \in \mathbb{R}^{n_1 \times n_2}} \|X\|_p^p \text{ subject to } \|A(X) - b\|_2 \leq \eta,
\]

where the Schatten-\(p\) (quasi) norm of \(X\) is defined by \(\|X\|_p = (\sum \sigma_i^p(X))^\frac{1}{p}\) with \(\sigma_i(X)\) being the singular values of \(X\). When \(p = 1\), problem (3) is the convex nuclear norm minimization problem. Nuclear norm minimization is a semi-definite program (SDP) and can be solved efficiently by convex optimization methods \[16, 35, 4\]. When 0 < \(p\) < 1, \(\|X\|_p\) denotes a quasi-norm and problem (3) is a non-convex optimization problem \[34, 42\]. A number of numerical algorithms implementing on the non-convex Schatten-\(p\) norm minimization problem has been developed and they empirically have better performance than nuclear norm problem \[26, 30, 31\].

Theoretical analysis showed that Schatten-\(p\) norm minimization problem (0 < \(p\) ≤ 1) \[3\] can exactly recover a low rank matrix on the case of noiseless measurement (i.e., \(\eta = 0\)) and stably recover a low rank matrix in the presence of observation noise (i.e., \(\eta > 0\)) when the measurement map \(A\) satisfies matrix restricted isometry property (M-RIP) \[5, 35, 42, 41\]. Recht et al. \[35\] generalized the sparsity-restricted isometry property (RIP) \[7\] defined for sparse vectors in compressed sensing to low rank matrix recovery. The matrix restricted isometry property (M-RIP) is defined as follows:

**Definition 1.1.** Given a linear operator \(A : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m\), for every integer \(r\) with 1 ≤ \(r\) ≤ \(n_1\), the matrix restricted isometry constant \(\delta_r\) is defined as the minimum constant that satisfies

\[
(1 - \delta_r)\|X\|_F^2 \leq \|A(X)\|_2^2 \leq (1 + \delta_r)\|X\|_F^2.
\]
for all $X \in \mathbb{R}^{n_1 \times n_2}$ with rank($X$) $\leq r$.

Literature [41] showed that when the unknown matrix $M$ satisfies rank($M$) $\leq r$ and measurement map $A$ satisfies M-RIP of order $2r$, the optimal solution $X^*$ of $\|X - M\|_p \leq C\eta$, where $C$ is a constant only dependent on the M-RIP constant $\delta^{2r}$ and rank $r$. Candès et al. [12] proved that any random measurement map $A$, which satisfies the following concentration inequality: for any given $X \in \mathbb{R}^{n_1 \times n_2}$ and for fixed $0 < \delta < 1$,

$$P\left(\left|\|A(X)\|_2^2 - \|X\|_F^2\right| \geq \delta \|X\|_F^2\right) \leq ce^{-tm\delta^2},$$

for fixed constants $c, t > 0$, will satisfy the M-RIP with restricted isometry constant $\delta_r \leq \delta$ with large probability provided that $m \geq c\delta^2 n_2 r$. Thus many types of random measurement maps such as Gaussian, Subgaussian and random partial Fourier measurement maps have the M-RIP with high probability [12, 25]. However, in the matrix completion problem, the projection operator $P\Omega$ doesn’t satisfy the M-RIP provided that there are rank one matrices in the kernel of $P\Omega$. Literatures [9, 10] show that under certain incoherence assumptions on the singular vectors of the unknown low rank matrix, matrix completion problem has good recovery performance. More studies about matrix completion could be found in [11, 36, 23, 40], etc.

1.2. Related iteratively reweighted least squares algorithm. Although the Schatten-$p$ norm minimization problem $(0 < p \leq 1)$ (3) can be solved using standard optimization tools, the iteratively reweighted least squares algorithm has been suggested as an attractive alternative simple and efficient algorithm for low rank matrix recovery (abbreviated as IRLS-M) in the literatures [26, 17, 30, 31]. Indeed, iteratively reweighted least squares (abbreviated as IRLS) algorithm appeared for the first time in the approximation theory for solving uniform approximation problems [27]. Then literatures [11, 14, 15, 18, 26, 39] used iteratively reweighted least squares algorithm as an efficient algorithm to solve convex and nonconvex $l_p$ ($0 < p \leq 1$) minimization problems in compressed sensing. Particularly, IRLS algorithm has been studied in detail for recovering a sparse signal in noiseless compressed sensing by Daubechies et al. [15]. Then Ba et al. [1] has used the IRLS algorithm to solve noisy $l_p$ ($0 < p \leq 1$) minimization problems, and its $k$th iteration is given by

$$x^k = \arg\min_{x \in \mathbb{R}^n} \left\{ \sum_i w_i^{k-1} x_i^2 : \|Ax - b\|_2 \leq \eta \right\},$$

where $w^{k-1} \in \mathbb{R}^n$ is a weight vector with $w_i^{k-1} = (|x_i^{k-1}|^2 + \epsilon^2)^{p/2-1}$ ($\epsilon > 0$ being a smoothed parameter added to ensure that $w_i^{k-1}$ is well defined). They proved that the IRLS algorithm has well theoretical and numerical recovery performance for sparse signal recovery problem in the presence of observation noise.

Fornasier et al. [17] have extended the IRLS algorithm [15] in sparse signal recovery to the IRLS-M algorithm for solving the noiseless nuclear norm minimization problem. Then Mohan and Fazel have used IRLS-M to solve Schatten-$p$ minimization $(0 < p \leq 1)$ problem for noiseless low rank recovery [30, 31]. In addition, [17, 30, 31] showed that the IRLS-M algorithm actually minimizes a certain smooth
approximation to the Schatten-$p$ norm ($0 < p \leq 1$), i.e., IRLS-M solves the following smoothed noiseless Schatten-$p$ minimization ($0 < p \leq 1$) problem:

\[
\min_{X \in \mathbb{R}^{n_1 \times n_2}} \text{Tr}((XX^T + \epsilon^2 I)^{\frac{p}{2}}) \quad \text{subject to} \quad \mathcal{A}(X) = b,
\]

where $\epsilon > 0$ is the smoothed parameter. Note that when $\epsilon \to 0$, $\text{Tr}((XX^T + \epsilon^2 I)^{\frac{p}{2}})$ is the approximation to Schatten-$p$ norm $\|X\|_p^p$. Convergence of IRLS-M has been proved strictly, and when $p = 1$ they showed that IRLS-M algorithm allows for efficient implementations while having theoretical guarantees similar to nuclear norm minimization in the absence of noise [17, 30, 31]. Numerical experiments in [17, 30, 31] have shown that IRLS-M algorithm saves computational time and is also efficient for matrix completion problem in the presence of noise. However, both papers lack a rigorously theoretical analysis of the convergence and stability of IRLS-M algorithm for low rank matrix recovery in the presence of noise. In addition, the theoretical analysis of the algorithm is given only for $p = 1$.

1.3. Contribution. In this paper, inspired by the good performance of IRLS algorithm for solving noisy sparse signal recovery problem [1], we extend the utility of IRLS-M algorithm for low rank matrix recovery in the presence of observation noise. We use IRLS-M algorithm to solve the following smoothed noisy Schatten-$p$ ($0 < p \leq 1$) minimization problem:

\[
\min_{X \in \mathbb{R}^{n_1 \times n_2}} \text{Tr}((XX^T + \epsilon^2 I)^{\frac{p}{2}}) \quad \text{subject to} \quad \|\mathcal{A}(X) - b\|_2 \leq \eta.
\]

Given $X^{k-1}$ at $(k - 1)$-th iteration, our IRLS-M algorithm generates $X^k$ by

\[
X^k = \arg \min_{X \in \mathbb{R}^{n_1 \times n_2}} \{\|(W^{k-1})^{\frac{1}{2}} X\|_F^2 : \|\mathcal{A}(X) - b\|_2 \leq \eta\},
\]

where $W^{k-1} \in \mathbb{R}^{n_1 \times n_2}$ is a weight matrix with $W^{k-1} = (X^{k-1}(X^{k-1})^T + \epsilon^2 I)^{\frac{1}{2}} - 1$ and $\epsilon > 0$ ensures that $W^{k-1}$ is well-conditioned. To recover a low rank matrix or sparse vector, IRLS-M [17, 30, 31] and IRLS [15, 14, 26] always vary $\epsilon$, starting at a large value and gradually reducing it. The update rule of $\epsilon$ in our IRLS-M algorithm for noisy low rank matrix recovery is the same as the one in [15, 17] (see Algorithm 1 in Section 2.2). We first give a convergence analysis of the IRLS-M algorithm for all $0 < p \leq 1$. While the matrix null space property plays a central role in [17, 30, 31], it is no longer useful for our IRLS-M algorithm for noisy low rank matrix recovery as the iteration error matrix $X^k - X$ does not lie in the null space of measurement map $\mathcal{A}$ any more. We use the M-RIP of measurement map $\mathcal{A}$ for establishing the stability of our IRLS-M algorithm and giving a local error bound for all $0 < p \leq 1$. Specially, when $p = 1$, we prove that the M-RIP constant $\delta_{2r} < \sqrt{2} - 1$ is sufficient for IRLS-M to recover an unknown (approximately) low rank matrix with an error that is proportional to the noise level.

1.4. Organization. The rest of this paper is organized as follows. In Section 2, we firstly introduce some basic notations used throughout the whole paper. Subsequently, we will state our IRLS-M algorithm for low rank matrix recovery in the presence of noise and give some preliminaries for proving the main results in Section 3. Section 3 give the convergence and stability analysis of IRLS-M algorithm in the noisy case provided that the measurement map $\mathcal{A}$ satisfies the matrix restricted isometry property. Section 4 give some conclusions.
2. IRLS-M algorithm for noisy low rank matrix recovery.

2.1. Notation. For a general matrix $X \in \mathbb{R}^{n_1 \times n_2}$ (without loss of generality assume $n_1 \leq n_2$), let $X^T$ be the adjoint matrix of $X$. We also write the singular value decomposition (abbreviated as SVD) of $X$ as $X = U \Sigma V^T$ for unitary matrices $U \in \mathbb{R}^{n_1 \times n_1}$, $V \in \mathbb{R}^{n_2 \times n_2}$ and a diagonal matrix $\Sigma = \text{diag}(\sigma_1(X), \sigma_2(X), \ldots, \sigma_{n_1}(X)) \in \mathbb{R}^{n_1 \times n_2}$, where $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_{n_1}(X) \geq 0$ are the singular values of $X$. We denote $\sigma(X) = (\sigma_1(X), \sigma_2(X), \ldots, \sigma_{n_1}(X))$ the singular vector of $X$. The trace of a square matrix $X \in \mathbb{R}^{n_1 \times n_1}$ is the sum of its diagonal entries, i.e., $\text{Tr}(X) = \sum_{i=1}^{n_1} X_{ii}$. The trace satisfies $\text{Tr}(X) = \text{Tr}(X^T)$ and $\text{Tr}(XY) = \text{Tr}(YX)$ for matrices $X = (X_{ij})$, $Y = (Y_{ij}) \in \mathbb{R}^{n_1 \times n_2}$. The inner product of $X$ and $Y$ is defined by $\langle X, Y \rangle = \text{Tr}(XY^T) = \sum_{i,j} X_{ij} Y_{ij}$. Therefore, the Frobenius norm of $X$ induced by this inner product is $\|X\|_F = \sqrt{\langle X, X \rangle} = \sqrt{\sum_{i,j} X_{ij}^2} = \sqrt{\sum_{i} \sigma_i^2(X)}$.

For a positive definite and self-adjoint matrix $W \in \mathbb{R}^{n_1 \times n_1}$, we define the weighted Frobenius norm of $X$ related to $W$ by $\|WX\|_F = (\text{Tr}(WXX^T))^\frac{1}{2}$. More generally, we consider the Schatten-$p$ (quasi) norms of $X$: $\|X\|_p = \|\sigma(X)\|_p = (\sum \sigma_i^p(X))^\frac{1}{p}$ for $0 < p \leq 1$. In particular, when $p = 1$, $\|X\|_1 = \|X\|_*$ denotes the nuclear norm of $X$. The spectral norm of a matrix $X$ is defined as $\|X\| = \sup_{\beta \in \mathbb{R}^{n_2}} \|X\beta\|_2/\|\beta\|_2$, which equals to the largest singular value of $X$. The rank of $X$ equals to the number of nonzero singular values of $X$. For any integer $1 \leq r \leq n_1$, the best rank-$r$ approximation $X_{[r]}$ of $X$ is defined as

$$X_{[r]} = U \left( \begin{array}{cc} \text{diag}(\sigma_{[r]}(X)) & 0 \\ 0 & 0 \end{array} \right) V^T = U_{[r]} \Sigma_{[r]} V_{[r]}^T,$$

where $\sigma_{[r]}(X) = (\sigma_1(X), \sigma_2(X), \ldots, \sigma_r(X))$, $U_{[r]} \in \mathbb{R}^{n_1 \times n_1}$ and $V_{[r]} \in \mathbb{R}^{n_2 \times n_2}$ contain the left and right $r$ singular vectors of $X$ corresponding to the first $r$ largest singular values respectively, and $\Sigma_{[r]} \in \mathbb{R}^{n_1 \times n_2}$ is a diagonal matrix with the first $r$ largest singular values of $X$ on its diagonal.

2.2. IRLS-M algorithm. In this subsection, we extend the utility of IRLS-M algorithm in [17, 30, 31] for low rank matrix recovery in the presence of observation noise. We aim to solve the minimization problem in (6). Note that when $\epsilon > 0$, the function $F_p^\epsilon(X) = \text{Tr}((XX^T + \epsilon^2 I)^\frac{1}{2})$ for $0 < p \leq 1$ is the convex ($p = 1$) and nonconvex ($0 < p < 1$) smoothed approximation to the rank function of $X$. In addition, for $X$, $F_p^\epsilon(X)$ is differentiable for $p > 0$. We denote the feasible set by $\mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|\mathcal{A}(X) - b\|_2 \leq \eta\}$. Motivated by the algorithm in [11, 17, 30, 31], we give the IRLS-M algorithm to solve problem (6) (see Algorithm 1 in next page).

Next we give some remarks for our IRLS-M algorithm (Algorithm 1).

Remark 1. Each iteration of the IRLS-M algorithm corresponds to a weighted least squares constrained to the closed quadratic convex set $\mathcal{C}$ and can be efficiently solved using standard convex optimization methods [3, 19, 20]. Literature [38] also has given an efficient method to solve a general class of least-squares constrained problems. In addition, the estimated rank $r$ is one of the most important parameters. Literatures [20, 31, 17] give many methods of rank estimation such as $r = \lfloor 1.5r^* \rfloor$ where $r^*$ is the true rank of the unknown matrix $M$. 

Inverse Problems and Imaging Volume 11, No. 4 (2017), 643–661
Algorithm 1 Iteratively reweighted least squares minimization for noisy low rank matrix recovery

**Input:** vector $b$, linear operator $A$, error bound $\eta$ and estimated rank $r$;

**Output:** matrix $X \in \mathbb{R}^{n_1 \times n_2}$;

1. Initialize $W^0 = I \in \mathbb{R}^{n_1 \times n_1}$, set $\epsilon_0 = 1$ and $0 < \gamma < 1$, choose $p \in (0, 1]$ ;
2. For $k = 1, 2, \ldots$
   - Let $X^k = \arg \min_{X \in \mathcal{C}} \| (W^{k-1})^\frac{1}{2} X \|^2_F$;
   - Let $\epsilon_k = \min\{\epsilon_{k-1}, \gamma \sigma_{r+1}(X^k)\}$;
   - Set $W^k = (X^k(X^k)^T + \epsilon_k^2 I)^{\frac{1}{2}}$;
3. The algorithm stops if $\epsilon_k = 0$, in this case, define $X^j := X^k$ for $j > k$. In general, the algorithm generates an infinite sequence $\{X^k\}$ of matrices.

**Remark 2.** Note that the smoothed parameter $\epsilon > 0$ is used to prevent the weight matrix $W^k$ to become ill-conditioned. In addition, in this sense the limit of $X^k$ ($x^k$) is often a good approximation of the true matrix (vector) up to $O(|\epsilon|)$ in magnitude [15] [14] [26]. Empirical analysis has shown that when starting at a large value of $\epsilon$ and gradually reducing it in each step, the IRLS [15] [14] [26] and IRLS-M [26, 17, 30, 31] algorithm dramatically improves the ability to recover sparse signals and low rank matrices. There are various ways to choose $\epsilon_k$ adaptively or in a static fashion. Ba et al. [1] analyzed the IRLS algorithm with fixed $\epsilon_k > 0$ for the simplicity of presentation. Mohan and Fazel [31] used the update rule $0 < \epsilon_k \leq \epsilon_{k+1}$. Daubechies et al. [15], Fornasier et al. [17], Mohan and Fazel [30] and Lai et al. [26] used the same adaptively update rule of $\epsilon_k$ as ours, which makes $F^p_\epsilon(X)$ provides a better approximation to the Schatten-$p$ norm $\|X\|_p^p$. Actually, the theoretical analysis for fixed $\epsilon_k$ and $0 < \epsilon_k \leq \epsilon_{k+1}$ case can be seen as a a part of our analysis of IRLS-M algorithm.

To derive the convergence of IRLS-M algorithm, we need a useful auxiliary function $J^p(X, W, \epsilon)$. Assume $X \in \mathbb{R}^{n_1 \times n_2}$, $0 \prec W = W^T \in \mathbb{R}^{n_1 \times n_1}$ and for $0 < p \leq 1$ consider the function

$$J^p(X, W, \epsilon) = \frac{p}{2}(\|W^\frac{1}{2} X\|^2_F + \epsilon^2 \|W^\frac{1}{2}\|^2_F + \frac{2-p}{p} \|W^{\frac{p}{p-1}}\|_F^2).$$

We can express the iterations of IRLS-M as

$$(7) \quad X^k = \arg \min_{X \in \mathcal{C}} J^p(X, W^{k-1}, \epsilon_{k-1})$$

and

$$(8) \quad W^k = \arg \min_{W > 0} J^p(X^k, W, \epsilon_k),$$

where (7) is obvious and (8) comes from the discussions of Section 3.1 in [31]. Moreover, the function $J^p(X, W, \epsilon)$ has the following property [31], for all $k \in N$,

$$J^p(X^{k+1}, W^{k+1}, \epsilon_{k+1}) \leq J^p(X^{k+1}, W^k, \epsilon_{k+1})$$

$$\leq J^p(X^{k+1}, W^k, \epsilon_k)$$

$$\leq J^p(X^k, W^k, \epsilon_k).$$

Note that by Algorithm 1 for each $k \geq 0$, $X^{k+1}$ is completely determined by $W^k$. In particular, for $k = 0$, $X^1$ is determined solely by $W^0$. Therefore, inequality (9) for $k = 0$ thus holds for arbitrary $X^0 \in \mathcal{C}$.
Next, we give the definition of stationary point over a convex set.

**Definition 2.1.** Let \( S \subseteq \mathbb{R}^{n_1 \times n_2} \) be an arbitrary convex subset and \( F(X) : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R} \) a continuously differentiable function over \( S \). Then if \( \langle \nabla F(X^*) , X - X^* \rangle \geq 0 \) for arbitrary \( X \in S \), we say that \( X^* \) is a stationary point of \( F(X) \) over \( S \).

**Lemma 2.2.** If \( X^* \) is a local minimizer of \( F(X) \) over a convex subset \( S \subseteq \mathbb{R}^{n_1 \times n_2} \), then we have \( \langle \nabla F(X^*) , X - X^* \rangle \geq 0 \) for arbitrary \( X \in S \). Moreover, if \( F(X) \) is convex over \( S \), then the condition \( \langle \nabla F(X^*) , X - X^* \rangle \geq 0 \) for arbitrary \( X \in S \) is sufficient for \( X^* \) to minimize \( F(X) \) over \( S \).

**Proof.** Let \( 0 \leq t \leq 1 \), then for arbitrary \( X \in S \), \( X^* + t(X - X^*) \in S \). Consider the analytic function,

\[
G(t) = F(X^* + t(X - X^*)) - F(X^*) = t \langle \nabla F(X^*) , X - X^* \rangle + o(t\|X - X^*\|_F),
\]

where for \( E = t(X - X^*) \), \( o(\|E\|_F) \) is defined by \( \lim_{\|E\|_F \to 0} \frac{o(\|E\|_F)}{\|E\|_F} = 0 \). If \( X^* \) is a local minimizer of \( F(X) \) over \( S \), then by the convex analysis theory \([3]\), \( G'(0) \geq 0 \), that is \( \langle \nabla F(X^*) , X - X^* \rangle \geq 0 \).

If \( F(X) \) is convex over \( S \) and \( \langle \nabla F(X^*) , X - X^* \rangle \geq 0 \) for arbitrary \( X \in S \), note that

\[
F(X) \geq F(X^*) + \langle \nabla F(X^*) , X - X^* \rangle \geq F(X^*).
\]

Then \( X^* \) is a minimizer of \( F(X) \) over \( S \). \( \square \)

The following lemma is useful for analyzing the convergence of the IRLS-M algorithm for noisy low rank matrix recovery.

**Lemma 2.3.** Assume that \( W \in \mathbb{R}^{n_1 \times n_1} \) and \( W = W^T > 0 \). Let \( S \) denote an arbitrary non-empty, closed and convex subset of \( \mathbb{R}^{n_1 \times n_2} \). Then if \( \hat{X} = \arg \min_{X \in S} \|W^{1/2}X\|_F^2 \), we have \( \langle W\hat{X} , X - \hat{X} \rangle \geq 0 \) for all \( X \in S \).

**Proof.** We give a simple proof of Lemma 2.3. For arbitrary \( X \in S \) and all \( t \in [0,1] \), then \( \hat{X} + t(X - \hat{X}) \in S \). By the optimality of \( \hat{X} \), we have

\[
\|W^{1/2}\hat{X}\|_F^2 \leq \|W^{1/2}(\hat{X} + t(X - \hat{X}))\|_F^2,
\]

which equals to

\[
\text{Tr}(W\hat{X}(\hat{X})^T) \leq \text{Tr}(W(\hat{X} + t(X - \hat{X}))(\hat{X} + t(X - \hat{X}))^T).
\]

Note that (11) also equals to

\[
0 \leq 2t\text{Tr}(W(X - \hat{X})(X - \hat{X})) + t^2\text{Tr}(W(X - \hat{X})(X - \hat{X}))^T.
\]

This implies \( \text{Tr}(W(X - \hat{X})(X - \hat{X})) = \langle W\hat{X} , X - \hat{X} \rangle \geq 0 \). \( \square \)

The next lemma gives some simple properties of the iterations \( \{X^k\}_{k \in \mathbb{N}} \) in IRLS-M algorithm.

**Lemma 2.4.** Let \( \{X^k\}_{k \in \mathbb{N}} \) be the iterations of IRLS-M algorithm, then \( \{X^k\}_{k \in \mathbb{N}} \) is bounded and \( \lim_{k \to \infty} \|X^{k+1} - X^k\|_F^2 = 0 \).
Proof. The proof of Lemma 2.4 is similar to the proof of Theorem 2 in [31]. But the main difference is the use of Lemma 2.3 instead of Lemma 13 in [31]. For completeness, we give the details. A direct computation shows that

\[ \mathcal{J}^p(X^k, W^k, \epsilon_k) = \text{Tr}((X^k(X^k)^T + \epsilon_k^2 I)^{\frac{p}{2}}) = \sum_{i=1}^{n_1} (\sigma_i^2(X^k) + \epsilon_k^2)^{\frac{p}{2}}. \]

Notice that

\[ \|X^k\|_p^p = \sum_{i=1}^{n_1} \sigma_i^p(X^k) \leq \sum_{i=1}^{n_1} (\sigma_i^2(X^k) + \epsilon_k^2)^{\frac{p}{2}} = \mathcal{J}^p(X^k, W^k, \epsilon_k) \leq \mathcal{J}^p(X^1, W^0, \epsilon_0) = D, \]

where the last inequality follows from (9). Therefore, the sequence \( \{X^k\}_{k \in \mathbb{N}} \) is bounded. In addition, for each \( k \geq 1 \), using (9) we have that

\[
\frac{2}{p} \left[ \mathcal{J}^p(X^k, W^k, \epsilon_k) - \mathcal{J}^p(X^{k+1}, W^{k+1}, \epsilon_{k+1}) \right] 
\geq \frac{2}{p} \left[ \mathcal{J}^p(X^k, W^k, \epsilon_k) - \mathcal{J}^p(X^{k+1}, W^k, \epsilon_k) \right]
\]
\[
= \langle W^k X^k, X^k \rangle - \langle W^k X^{k+1}, X^{k+1} \rangle
\]
\[
= \langle W^k X^k + X^{k+1}, X^k - X^{k+1} \rangle
\]
\[
\geq \langle W^k (X^k + X^{k+1}), X^k - X^{k+1} \rangle - 2 \langle W^k X^{k+1}, X^k - X^{k+1} \rangle
\]
\[
= \langle W^k (X^k - X^{k+1}), X^k - X^{k+1} \rangle
\]
\[
= \| (W^k)^{\frac{1}{2}} (X^k - X^{k+1}) \|_F^2,
\]

where the second inequality follows from \( X^{k+1} = \arg \min_{X \in \mathbb{C}} \| (W^k)^{\frac{1}{2}} X \|_F^2 \) and Lemma 2.3. On the other hand, for any \( i = 1, 2, \ldots, n_1 \),

\[
(\sigma_i^2(X^k) + \epsilon_k^2)^{\frac{p}{2}} \leq \sum_{i=1}^{n_1} (\sigma_i^2(X^k) + \epsilon_k^2)^{\frac{p}{2}} \leq D
\]

and from the definition of \( W^k \)

\[
\| (W^k)^{-\frac{1}{2}} \|_F^2 = \text{Tr}((W^k)^{-1}) = \text{Tr}((X^k(X^k)^T + \epsilon_k^2 I)^{-\frac{1}{2}}) = \sum_{i=1}^{n_1} (\sigma_i^2(X^k) + \epsilon_k^2)^{-\frac{1}{2}}.
\]

Hence, there exists a positive constant \( D_0 \), such that \( \| (W^k)^{-\frac{1}{2}} \|_F^2 \leq D_0 \). Therefore,

\[
\| X^{k+1} - X^k \|_F^2 = \| (W^k)^{-\frac{1}{2}} (W^k)^{\frac{1}{2}} (X^{k+1} - X^k) \|_F^2
\]
\[
\leq \| (W^k)^{-\frac{1}{2}} \|_F^2 \| (W^k)^{\frac{1}{2}} (X^{k+1} - X^k) \|_F^2
\]
\[
\leq D_0 \| (W^k)^{\frac{1}{2}} (X^{k+1} - X^k) \|_F^2
\]
\[
\leq \frac{2D_0}{p} [\mathcal{J}^p(X^k, W^k, \epsilon_k) - \mathcal{J}^p(X^{k+1}, W^{k+1}, \epsilon_{k+1})].
\]
Summing the above inequalities over all \( k \geq 1 \), we have that \( \lim_{k \to \infty} \|X^{k+1} - X^k\|_p^2 = 0. \)

3. Convergence and stability of IRLS-M for noisy low rank matrix recovery. In this section, convergence and stability of IRLS-M algorithm for all \( 0 < p \leq 1 \) in the presence of observation noise are studied. Specially when \( p = 1 \), we give a weaker sufficient condition for IRLS-M algorithm to recover an unknown (approximately) low rank matrix for noisy low rank matrix recovery than \([17, 26]\), based on the matrix restricted isometry property. Our analysis is inspired by \([1, 15, 17, 30, 31, 26]\).

3.1. Convergence of IRLS-M algorithm. It is clear from Algorithm 1 that \( \{\epsilon_k\}_{k \in \mathbb{N}} \) is a nonincreasing sequence and lower bounded by zero, then \( \{\epsilon_k\}_{k \in \mathbb{N}} \) must converge to a nonnegative constant \( \tau \). In addition, because of the boundedness of \( \{X^k\}_{k \in \mathbb{N}} \) (see Lemma 2.4), it has at least a convergent subsequence. Below we will show that when \( \tau > 0 \), the limit of any convergent subsequence is a stationary point of \( F_\tau^p(X) \). When \( \tau = 0 \), we will show that there exists a convergent subsequence which converges to a low rank matrix. Theorem 3.1 gives the main convergence result.

**Theorem 3.1.** Suppose that the sequence of matrices \( \{X^k\}_{k \in \mathbb{N}} \) is produced by the IRLS-M algorithm (Algorithm 1) and \( \epsilon_k \to \tau \) when \( k \to \infty \). Then \( \{X^k\}_{k \in \mathbb{N}} \) has at least one convergent subsequence. If \( \tau > 0 \), the limit of any convergent subsequence of \( \{X^k\}_{k \in \mathbb{N}} \) is a stationary point of \( F_\tau^p(X) = \text{Tr}((XX^T + \tau^2 I)\hat{Z}) \) over \( \mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\} \). If \( \tau = 0 \), then there exists a convergent subsequence of \( \{X^k\}_{k \in \mathbb{N}} \) which converges to a rank-r matrix \( \overline{X} \).

**Proof.** By Lemma 2.4, \( \{X^k\}_{k \in \mathbb{N}} \) is a bounded sequence. Hence there exists a subsequence \( \{X^{k_j}\}_{j \in \mathbb{N}} \) of \( \{X^k\}_{k \in \mathbb{N}} \) converging to a matrix \( \overline{X} \). Note that \( \|A(\overline{X}) - b\|_2 = \lim_{j \to \infty} \|A(X^{k_j}) - b\|_2 \leq \eta \), then \( \overline{X} \in \mathcal{C} \). If \( \lim_{k \to \infty} \epsilon_k = \tau > 0 \), by the definition of \( W^{k_j} \), \( W^{k_j} = (X^{k_j}(X^{k_j})^T + \epsilon_k^2 I)^{\frac{1}{2}} \), then \( W^{k_j} \to W = (\overline{X}(\overline{X})^T + \tau^2 I)^{\frac{1}{2}} \) for \( j \to \infty \). By invoking Lemma 2.4 we also have \( X^{k_{j+1}} \to \overline{X} \) when \( j \to \infty \). Hence, for arbitrary \( X \in \mathcal{C} \), by the optimality of \( X^{k_{j+1}} \) and Lemma 2.3, we have

\[
\langle W^{k_j} X^{k_{j+1}}, X - X^{k_{j+1}} \rangle = \lim_{j \to \infty} \langle W^{k_j} X^{k_{j+1}}, X - X^{k_{j+1}} \rangle \geq 0.
\]

For the matrix \( \overline{X} \in \mathcal{C} \) and fixed \( \tau > 0 \), we can use Proposition 7.4 in \([17]\) to get

\[
\nabla F_\tau^p(X) = \overline{W} \overline{X},
\]

where \( \overline{W} = (\overline{X}(\overline{X})^T + \tau^2 I)^{\frac{1}{2}} \). Hence, \( \langle \nabla F_\tau^p(\overline{X}), X - \overline{X} \rangle \geq 0 \) and by the Definition 2.1, \( \overline{X} \) is a stationary point of \( F_\tau^p(X) \) over the constraint set \( \mathcal{C} \).

If \( \lim_{k \to \infty} \epsilon_k = \tau = 0 \), from the updating rule of \( \epsilon_k = \min\{\epsilon_{k-1}, \gamma \sigma_{r+1}(X^{k})\} \), then \( \epsilon_k = 0 \) for some integer \( k \), or \( \epsilon_k > 0 \) for all \( k \). If \( \epsilon_k = 0 \) for some integer \( k \), we can set \( \overline{X} = X^k \). If \( \epsilon_k > 0 \), there exists a subsequence \( \{\epsilon^{k_j}\}_{j \in \mathbb{N}} \) such that \( \epsilon^{k_{j+1}} < \epsilon^{k_j} \) and \( \epsilon^{k_{j+1}} = \gamma \sigma_{r+1}(X^{k_j}) \). By the boundedness of \( \{X^{k_j}\}_{j \in \mathbb{N}} \), we can further extract a convergent subsequence of \( \{X^{k_j}\}_{j \in \mathbb{N}} \) which we still denote it by \( \{X^{k_j}\}_{j \in \mathbb{N}} \). Let \( \lim_{j \to \infty} X^{k_j} = \overline{X} \). Then \( \sigma_{r+1}(X) = \lim_{j \to \infty} \sigma_{r+1}(X^{k_j}) = \lim_{j \to \infty} \frac{\epsilon^{k_{j+1}}}{\gamma} = 0 \). Thus \( \overline{X} \) is an rank-r matrix. \( \square \)
3.2. Stability of IRLS-M algorithm. In this subsection, we give the stability of IRLS-M algorithm for noisy low rank matrix recovery. Under the M-RIP assumption, Theorem 3.2 and 3.3 consider the case for all $0 < p \leq 1$ and Theorem 3.4 focuses on the special case $p = 1$. In the following, we assume that the true matrix $M$ satisfies $\mathcal{A}(M) + z = b$ and $\|z\|_2 \leq \eta$.

**Theorem 3.2.** Suppose that the measurement map $\mathcal{A}$ satisfies the M-RIP of order $2r \leq n_1$ with $\delta_{2r} < 1$. If $\lim_{k \to \infty} \epsilon_k = \tau = 0$, then for arbitrary matrix $X \in \mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\}$, any limit point $\bar{X}$ of the convergent subsequence of $\{X^k\}_{k \in \mathbb{N}}$ generated by IRLS-M algorithm with all $0 < p \leq 1$ satisfies

$$\|\bar{X} - X\|_F \leq C_1 \eta + C_2 \|X - X_{[\tau]}\|_F,$$

where $C_1$, $C_2$ are positive constants and $X_{[\tau]}$ is the best rank-$r$ approximation of $X$.

**Remark 3.** According to Theorem 3.2 if the true matrix $M$ satisfies rank($M$) $\leq r$, then $\bar{X}$ differs from the true matrix $M$ by an amount proportional to the noise level $\eta$.

When $\lim_{k \to \infty} \epsilon_k = \tau > 0$, the next theorem states that IRLS-M algorithm for $0 < p \leq 1$ is stable if the limit point $\bar{X}$ of the convergent subsequence of $\{X^k\}_{k \in \mathbb{N}}$ coincides with a minimizer of $F_p^r(X) = \text{Tr}((XX^T + \tau^2 I)^{\frac{r}{2}})$ over $\mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\}$.

**Theorem 3.3.** Let $\theta > 1$ and $k$ be $2r\theta^{\frac{2}{\sqrt{7} - \epsilon}}$ rounded up to an integer, i.e., $k = \lceil 2r\theta^{\frac{2}{\sqrt{7} - \epsilon}} \rceil$ for $0 < p \leq 1$. Suppose the measurement map $\mathcal{A}$ satisfies the M-RIP with $\delta_k + \theta^{\frac{1}{2}}\delta_{2r+k} < \theta^{\frac{3}{2}} - 1$. Consider the IRLS-M algorithm for $0 < p \leq 1$ with parameter $\gamma = (n_1)^{\frac{1}{2} - \frac{1}{2p}}$ and $\lim_{k \to \infty} \epsilon_k = \tau > 0$. If the limit point $\bar{X}$ of the convergent subsequence of $\{X^k\}_{k \in \mathbb{N}}$ obeying $\bar{X} = \arg\min_{X \in \mathcal{C}} \text{Tr}((XX^T + \tau^2 I)^{\frac{r}{2}})$, then for any integer $t < r + 1 - C_0$ ($C_0$ is a positive constant explicitly given in (10)) and arbitrary matrix $X \in \mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\}$, we have

$$\|\bar{X} - X\|_p \leq C_1 \eta^p + C_2 \|X - X_{[\tau]}\|_p,$$

where $C_1$, $C_2$ are positive constants and $X_{[\tau]}$ is the best rank-$t$ approximation of $X$.

**Remark 4.** When $\epsilon_k > 0$ being fixed in each iterative step, it is a special case of Theorem 3.3.

When $p = 1$ and $\lim_{k \to \infty} \epsilon_k = \tau > 0$, the function $F_1^r(X) = \text{Tr}((XX^T + \tau^2 I)^{\frac{r}{2}})$ is strictly convex with respect to $X$. Hence, the limit point $\bar{X}$ of the convergent subsequence of $\{X^k\}_{k \in \mathbb{N}}$ generated by IRLS-M algorithm is the unique global minimizer of $F_1^r(X)$ over the convex set $\mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\}$. Next theorem gives a special case in $p = 1$.

**Theorem 3.4.** Suppose that the measurement map $\mathcal{A}$ satisfies the matrix M-RIP of order $2r \leq n_1$ with $\delta_{2r} \leq \sqrt{2} - 1$. Consider the IRLS-M algorithm for $p = 1$ with parameter $\gamma = \frac{1}{n_1}$ and $\lim_{k \to \infty} \epsilon_k = \tau > 0$. For any integer $t < r - \frac{2}{1 - \beta}$ with $\beta = \frac{\sqrt{2}n_1}{1 - \sqrt{2}r}$ and arbitrary matrix $X \in \mathcal{C} := \{X \in \mathbb{R}^{n_1 \times n_2} : \|A(X) - b\|_2 \leq \eta\}$, the limit point $\bar{X}$ of the convergent subsequence of $\{X^k\}_{k \in \mathbb{N}}$ generated by IRLS-M algorithm satisfies

$$\|\bar{X} - X\|_* \leq C_1 \eta + C_2 \|X - X_{[\tau]}\|_*,$$

where $C_1$, $C_2$ are positive constants and $X_{[\tau]}$ is the best rank-$t$ approximation of $X$. 


Remark 5. Note that when \( p = 1 \), [31] used the matrix NSP to give the matrix recovery guarantee, while we use the M-RIP to obtain the recovery guarantee. Our sufficient condition \( \delta_{2r} < \sqrt{2} - 1 \) is slightly weaker than \( \delta_{2r} < \frac{\sqrt{3} - 1}{\sqrt{2} - 1} < 1 \) in [26] and \( \frac{\sqrt{3} - 1}{\sqrt{2} - 1} < 1 \) in [17]. We have not tried to optimize the M-RIP constant, we expect with a more complicated proof as in [5], one can still improve this condition. In addition, literatures [26, 17] have successfully used the Woodbury matrix identity to give the error analysis of IRLS-M algorithm for matrix completion problem. We think it is also possible to extend our analysis of IRLS-M algorithm to the noisy matrix completion problem by using Woodbury matrix identity. We leave it to the interested readers.

3.2.1. Proof of Theorem 3.2

Proof. In the case \( \lim k \to \infty \epsilon_k = \tau = 0 \), by Theorem 3.1 there exists a subsequence \( \{X^k\}_{j \in \mathbb{N}} \) of \( \{X^k\}_{k \in \mathbb{N}} \) converging to an rank-\( r \) matrix \( X \in C \). For any matrix \( X \in C \), let \( X_{[r]} \) be the best rank-\( r \) approximation of \( X \). Note that \( \text{rank}(X) = \text{rank}(X) + \|A(X) - b\|_2 \). In addition, if \( \epsilon_k \) converges to an rank-\( r \) matrix \( X \in C \), then for arbitrary matrices \( X \), we use the triangle inequality to have

\[
\|X - X\|_F \leq \|X - X_{[r]}\|_F + \|X_{[r]} - X\|_F
\]

\[
\leq \frac{1}{\sqrt{1 - \delta_{2r}}} \|A(X - X_{[r]}\|_2 + \|X - X_{[r]}\|_F
\]

\[
\leq \frac{1}{\sqrt{1 - \delta_{2r}}} \left( \|A(X) - b\|_2 + \|A(X) - b\|_2 \right)
\]

\[
+ \frac{1}{\sqrt{1 - \delta_{2r}}} \|A\| \|X - X_{[r]}\|_F + \|X - X_{[r]}\|_F
\]

\[
\leq \frac{2}{\sqrt{1 - \delta_{2r}}} \eta + \left( \frac{1}{\sqrt{1 - \delta_{2r}}} + 1 \right) \|X - X_{[r]}\|_F,
\]

where \( \|A\| \) is the operator norm of \( A \).

3.2.2. Proof of Theorem 3.3

For arbitrary \( X \in C \), let \( H = X - X \) and \( X_{[r]} \) be the best rank-\( r \) approximation of \( X \). The next lemma is related to the matrix block decomposition. The details of the proof can be found in [6, 17, 35].

Lemma 3.5. The error matrix \( H \) can be decomposed as \( H = H^{(r)} + H_c^{(r)} \), where \( \text{rank}(H^{(r)}) \leq 2r \), \( (X_{[r]}^T)H^{(r)} = 0 \), \( X_{[r]}(H^{(r)}_c)^T = 0 \) and \( \langle H^{(r)}, H^{(r)}_c \rangle = 0 \). Moreover, \( H^{(r)}_c \) can be decomposed as \( H^{(r)}_c = \sum_{j \geq 1} H_{T_j} \) with \( \text{rank}(H_{T_j}) \leq k \), where \( k \) is defined as in Theorem 3.3 \( \langle H^{(r)}, H_{T_j} \rangle = 0 \) for all \( j = 1, 2, \ldots, (H_{T_j})^T H_{T_j} = 0 \) and \( H_{T_i} H_{T_j}^T = 0 \) for all \( i, j = 1, 2, \ldots, i \neq j \). In addition, we assume that \( H_{T_k} \) corresponds to the first \( k \) largest singular values of \( H^{(r)}_c \), \( H_{T_2} \) corresponds to the next \( k \) largest singular values of \( H^{(r)}_c \), . . .

The following lemma is useful for our proof (see [42]).

Lemma 3.6. Let \( 0 < p \leq 1 \), then for arbitrary matrices \( Y_1, Y_2 \in \mathbb{R}^{n_1 \times n_2} \), \( \|Y_1 + Y_2\|_p \leq \|Y_1\|_p + \|Y_2\|_p \). In addition, if \( Y_1^T Y_2 = 0 \) and \( Y_1 Y_2^T = 0 \), then \( \|Y_1 + Y_2\|_p = \|Y_1\|_p + \|Y_2\|_p \).
The following three lemmas are related to the error matrix $H$.

**Lemma 3.7.** $\|H^{(r)}\|^p_p \leq \|H^{(r)}\|^p_p + 2\|X - X_{[r]}\|^p_p + n_1\varepsilon^p$.

**Proof.** Note that for any matrix $X \in C$ and the minimizer $X$ of $F_p^p(X) = \text{Tr}((X X^T + \varepsilon^2 I)^{\frac{p}{2}})$ subject to the constraint set $\mathcal{C}$,

\[
\|X\|^p_p \leq F_p^p(X) \leq F_p^p(X) = \sum_{i=1}^{n_1} (\sigma_i^2(X) + \varepsilon^2)^{\frac{p}{2}} \leq \|X\|^p_p + n_1\varepsilon^p.
\]

On the other hand,

\[
\|X\|^p_p = \|X - H\|^p_p
\]
\[
= \|X - X_{[r]} + X_{[r]} - H^{(r)} - H^{(r)}\|^p_p
\]
\[
\geq \|X_{[r]} - H^{(r)}\|^p_p - \|X - X_{[r]} - H^{(r)}\|^p_p
\]
\[
= \|X_{[r]}\|^p_p + \|H^{(r)}\|^p_p - \|X - X_{[r]} - H^{(r)}\|^p_p
\]
\[
\geq \|X_{[r]}\|^p_p + \|H^{(r)}\|^p_p - \|X - X_{[r]}\|^p_p - \|H^{(r)}\|^p_p,
\]

where the inequalities follow from the triangle inequality and the third equality uses Lemma 3.5 and Lemma 3.6. Combining (12), (13) and using $\|X\|^p_p \leq \|X_{[r]}\|^p_p + \|X - X_{[r]}\|^p_p$, we finally get

\[
\|H^{(r)}\|^p_p \leq \|H^{(r)}\|^p_p + 2\|X - X_{[r]}\|^p_p + n_1\varepsilon^p.
\]

\[\blacksquare\]

**Lemma 3.8.** $\sum_{j \geq 2} \|H_{T_j}\|^p_p \leq \frac{1}{2} \|H^{(r)} + H_{T_1}\|^p_p + 2k^{\frac{p}{2}} - 1 \|X - X_{[r]}\|^p_p + n_1\varepsilon^p k^{\frac{p}{2}} - 1$.

**Proof.** By (23) in [12], for $j \geq 2$, $\|H_{T_j}\|^p_p \leq k^{\frac{p}{2}} - 1 \|H_{T_{j-1}}\|^p_p$. Therefore, combining Lemma 3.5 and Lemma 3.7 yields

\[
\sum_{j \geq 2} \|H_{T_j}\|^p_F \leq \sum_{j \geq 1} k^{\frac{p}{2}} - 1 \|H_{T_j}\|^p_p = k^{\frac{p}{2}} - 1 \|H^{(r)}\|^p_p
\]
\[
\leq k^{\frac{p}{2}} - 1 \|H^{(r)}\|^p_p + 2\|X - X_{[r]}\|^p_p + n_1\varepsilon^p
\]
\[
\leq k^{\frac{p}{2}} - 1 (\|H^{(r)}\|^p_p + 2\|X - X_{[r]}\|^p_p + n_1\varepsilon^p)
\]
\[
= (\frac{k}{2r})^{\frac{p}{2}} - 1 \|H^{(r)}\|^p_p + 2k^{\frac{p}{2}} - 1 \|X - X_{[r]}\|^p_p + k^{\frac{p}{2}} - 1 n_1\varepsilon^p
\]
\[
\leq \frac{1}{2} \|H^{(r)} + H_{T_1}\|^p_p + 2k^{\frac{p}{2}} - 1 \|X - X_{[r]}\|^p_p + n_1\varepsilon^p k^{\frac{p}{2}} - 1.
\]

\[\blacksquare\]

**Lemma 3.9.** Let $H_1 = \sum_{j \geq 2} H_{T_j}$, then $\|H_1\|^p_F \leq \frac{1}{2c(p)} \|H^{(r)} + H_{T_1}\|^p_p + 2k^{\frac{p}{2}} - 1 c(p) \|X - X_{[r]}\|^p_p + n_1\varepsilon^p k^{\frac{p}{2}} - 1$, where $c(p) = (\frac{2}{p} - 1)^{\frac{p}{2}}$.

**Proof.** The proof of this lemma mainly follows the idea of Theorem 2.1 in [37] for sparse signal recovery. First note that for arbitrary $1 \leq j \leq n_1$,

\[
\|H^{(r)}_j\|^p_p = \sum_{i=1}^{j} \sigma_i^p(H^{(r)}_j) \geq \sum_{i=1}^{j} \sigma_i^p(H^{(r)}_j) \geq j \sigma_j^p(H^{(r)}_j).
\]
Therefore, using Lemma 3.5 and (14) we have
\[ \|H_1\|_F^p \leq \sum_{j \geq 2} H_{T_1,j} \|F\|_F^p \leq \sum_{j \geq k+1} \sigma_j^2(H^{(r)}) \leq \sum_{j \geq k+1} \|H^{(r)}\|_F^p \leq \frac{\|H^{(r)}\|_F^2}{(\frac{1}{p} - 1)k^{\frac{k}{p} - 1}}, \]
where the last inequality follows from for 0 < p ≤ 1, \[ \sum_{j \geq k+1} j^{-\frac{k}{p}} \leq \int_1^\infty t^{-\frac{k}{p}} dt = \frac{1}{(\frac{1}{p} - 1)k^{\frac{k}{p} - 1}}. \]
Hence, using Lemma 3.7 we obtain
\[ \|H_1\|_F^p \leq \frac{\|H^{(r)}\|_F^p}{(\frac{1}{p} - 1)k^{\frac{k}{p} - 1}} \leq \frac{1}{c(p)} k^{\frac{k}{p} - 1} \left( \|H^{(r)}\|_F^p + 2\|X - X_{[r]}\|_p^p + n_1\tau_p \right) \]
\[ \leq \frac{1}{c(p)} k^{\frac{k}{p} - 1} \left( (2r)^{\frac{1}{2} - \frac{k}{p}} \|H^{(r)}\|_F^p + 2\|X - X_{[r]}\|_p^p + \frac{k^{\frac{k}{p} - 1}n_1\tau_p}{c(p)} \right) \]
\[ \leq \frac{1}{c(p)} \|H^{(r)} + H_{T_1}\|_F^p + \frac{2k^{\frac{k}{p} - 1}}{c(p)} \|X - X_{[r]}\|_p^p + \frac{n_1\tau_p k^{\frac{k}{p} - 1}}{c(p)}. \]

Next we give the proof of Theorem 3.3.

**Proof of Theorem 3.3.** By the feasibility of X and \( \overline{X} \),
\[ \|A\|_2 = \|A(X - \overline{X}) - b\|_2 \leq \|A(X) - b\|_2 + \|A(X) - b\|_2 \leq 2\eta. \]
Therefore, using the M-RIP condition and triangle inequality, we have
\[ (2\eta)^p \geq \|A\|_2^p = \|A(H^{(r)} + H_{T_1})\|_2^p - \sum_{j \geq 2} \|A(H_{T_1})\|_2^p \]
\[ \geq \|A(H^{(r)} + H_{T_1})\|_2^p - \sum_{j \geq 2} \|A(H_{T_1})\|_2^p \]
\[ \geq (1 - \delta_{2r+k})^{\frac{k}{p}} \|H^{(r)} + H_{T_1}\|_F^p - (1 + \delta_k)^{\frac{k}{p}} \sum_{j \geq 2} \|H_{T_1}\|_F^p, \]
which together with Lemma 3.3 implies
\[ \left( (1 - \delta_{2r+k})^{\frac{k}{p}} - \frac{1}{\theta}(1 + \delta_k)^{\frac{k}{p}} \right) \|H^{(r)} + H_{T_1}\|_F^p \]
\[ \leq (1 + \delta_k)^{\frac{k}{p}} \left( (2r)^{\frac{1}{2} - \frac{k}{p}} \|X - X_{[r]}\|_p^p + \frac{k^{\frac{k}{p} - 1}n_1\tau_p}{c(p)} \right) + (2\eta)^p. \]
The condition \( \delta_k + \theta^{\frac{k}{p}}\delta_{2r+k} < \theta^{\frac{k}{p}} - 1 \) in Theorem 3.3 implies that \( (1 - \delta_{2r+k})^{\frac{k}{p}} - \frac{1}{\theta}(1 + \delta_k)^{\frac{k}{p}} > 0. \) Thus
\[ \|H^{(r)} + H_{T_1}\|_F^p \leq c_0\eta^p + c_1n_1\tau_p + c_2\|X - X_{[r]}\|_p^p, \]
where \( c_0 = \frac{2^p}{(1 - \delta_{2r+k})^{\frac{k}{p}} - \frac{1}{\theta}(1 + \delta_k)^{\frac{k}{p}}} \), \( c_1 = \frac{k^{\frac{k}{p} - 1}(1 + \delta_k)^{\frac{k}{p}}}{(1 - \delta_{2r+k})^{\frac{k}{p}} - \frac{1}{\theta}(1 + \delta_k)^{\frac{k}{p}}} \) and \( c_2 = 2c_1. \) Combining Lemma 3.9 and (15), we obtain
\[ \|H\|_F^p \leq (n_1)^{1 - \frac{k}{p}} \|H\|_F^p \]
\[ \leq (n_1)^{1 - \frac{k}{p}} \left( \|H^{(r)} + H_{T_1}\|_F^p + \|H_1\|_F^p \right) \]
\( \leq (n_1)^{1-\frac{2}{p}} \left( 1 + \frac{1}{\theta c(p)} \right) \| H^{(r)} + HT_1 \|_F^p \)

\[ \leq (n_1)^{1-\frac{2}{p}} \left( 1 + \frac{1}{\theta c(p)} \right) \left[ c_0 \eta^p + c_1 n_1 \tilde{\Xi} + c_2 \|X - X[\|_p \right] \]

\[ \leq (n_1)^{1-\frac{2}{p}} \left( 1 + \frac{1}{\theta c(p)} \right) \left[ c_0 \eta^p + c_1 \eta \tilde{\Xi} + c_2 \|X - X[\|_p \right] \]

where \( c_3 = (n_1)^{1-\frac{2}{p}} (1 + \frac{1}{\theta c(p)})c_0 \), \( c_4 = (n_1)^{1-\frac{2}{p}} (1 + \frac{1}{\theta c(p)})c_2 + \frac{2(n_1)^{1-\frac{2}{p}}}{c(p)k^{1-\frac{2}{p}}} \), and

\[ (n_1)^2 \tilde{\Xi} = \lim_{k \to \infty} (n_1)^2 \tilde{\Xi}^p \leq \lim_{k \to \infty} (n_1)^2 \tilde{\Xi}^p \left( \gamma \sigma_{r+1}(X^k) \right)^p = \sigma_{r+1}(X). \]

For integer \( t \leq r \), we have \( \|X - X[\|_p \leq \|X - X[\|_p \) and

\[ (r - t + 1)\sigma_{r+1}(X) \leq \sum_{i=t+1}^{r+1} \sigma_i(X) \]

\[ \leq \|X - X[\|_p \]

Then if \( t < r + 1 - C_0 (C_0 = c_4 > 0) \), we could substitute (17) and (18) into (16) to get

\[ \|X - X[\|_p \leq C_1 \eta + C_2 \|X - X[\|_p \]

where \( C_1 = \frac{c_3}{1 - \frac{2}{p}} \), \( C_2 = \frac{c_4 + \frac{2c_3}{c(p)k^{1-\frac{2}{p}}} + \frac{2(n_1)^{1-\frac{2}{p}}}{c(p)k^{1-\frac{2}{p}}} \) are positive constants and \( X[\) is the best rank-\( t \) approximation of \( X \). \( \square \)

3.2.3. Proof of Theorem 3.4. The proof makes use of the ideas from [8, 28, 26, 17, 34]. Let \( H = X - \tilde{X} \), the SVD of \( X \) be \( X = USV^T \) and the best rank-\( r \) approximation of \( X \) be \( X[ = U[SV[ \). Define a projection \( P = U[SV[ \) and its complementary projection \( P^c = I - P \). Let \( H[ = P(H) \) and \( H[ = P^c(H) \). Then we can decompose \( H \) as \( H = H[ + H[c] \). Furthermore, let the SVD of \( H[c] \) be \( H[c] = \tilde{U} \tilde{\Sigma} \tilde{V}^T \). We can write \( \tilde{\Sigma} = \tilde{\Sigma}_1 + \tilde{\Sigma}_2 + \ldots \), where \( \tilde{\Sigma}_1 \) is a diagonal matrix with diagonal entries corresponding to the first \( r \) largest singular values of \( H[c] \) and \( \tilde{\Sigma}_2 \) is a diagonal matrix with diagonal entries corresponding to the next \( r \) largest singular values of \( H[c], \ldots \) Similarly, we write \( \tilde{U} = U[ + U[ + \ldots \) and \( \tilde{V} = V[ + V[ + \ldots \) accordingly. Then we can further decompose \( H[c] \) as \( H[c] = H[ + H[ + \ldots \) with

\[ H_i = \tilde{U}_i \tilde{\Sigma}_i \tilde{V}_i^T \] for \( i \geq 1 \). We further let \( H[01 = H[ + H[ \). The following lemmas are useful for proving Theorem 3.4.

Inverse Problems and Imaging Volume 11, No. 4 (2017), 643–661
Lemma 3.10. Suppose the measurement map $\mathcal{A}$ satisfies the M-RIP of order $2r \leq n_1$ with $\delta_{2r} < 1$, then

$$
\|H_{01}\|_F \leq \beta \sum_{j \geq 2} \|H_j\|_F + C\eta,
$$

where $\beta = \frac{\sqrt{2\delta_{2r}}}{1-\delta_{2r}}, \ C = \frac{2\sqrt{1+\delta_{2r}}}{1-\delta_{2r}}$.

Proof. Note that

$$
\|H_{01}\|_F^2 = \langle H_{01}, H \rangle = \langle H_{01}, H - \mathcal{A}^T \mathcal{A}(H) + \mathcal{A}^T \mathcal{A}(H) \rangle \\
= \langle H_{01}, (I - \mathcal{A}^T \mathcal{A})(H) \rangle + \langle H_{01}, \mathcal{A}^T \mathcal{A}(H) \rangle.
$$

Moreover,

$$
\langle H_{01}, (I - \mathcal{A}^T \mathcal{A})(H) \rangle = \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H \rangle \\
= \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_{01} + \sum_{j \geq 2} H_j \rangle \\
= \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_{01} \rangle + \sum_{j \geq 2} \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_j \rangle \\
= \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_{01} \rangle + \sum_{j \geq 2} \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_j \rangle \\
+ \sum_{j \geq 2} \langle (I - \mathcal{A}^T \mathcal{A})(H_{01}), H_j \rangle \\
\leq \delta_{2r} \|H_{01}\|_F^2 + \delta_{2r} \|H_{01}\|_F \cdot \sum_{j \geq 2} \|H_j\|_F \\
+ \delta_{2r} \|H_1\|_F \cdot \sum_{j \geq 2} \|H_j\|_F \\
\leq \delta_{2r} \|H_{01}\|_F^2 + \sqrt{2\delta_{2r}} \|H_{01}\|_F \cdot \sum_{j \geq 2} \|H_j\|_F,
$$

where we have used Lemma 3.4 in [28] to get the first inequality. For estimating the second term in (19), by the feasibility of $X$ and $\overline{X}$, we get

$$
\|\mathcal{A}(H)\|_2 = \|\mathcal{A}(X - \overline{X})\|_2 \leq \|\mathcal{A}(X) - b\|_2 + \|\mathcal{A}(X) - b\|_2 \leq 2\eta.
$$

Then using the M-RIP condition and (21), we obtain

$$
\langle H_{01}, \mathcal{A}^T \mathcal{A}(H) \rangle = \langle \mathcal{A}(H_{01}), \mathcal{A}(H) \rangle \\
\leq \|\mathcal{A}(H_{01})\|_2 \cdot \|\mathcal{A}(H)\|_2 \\
\leq 2\eta \sqrt{1 + \delta_{2r}} \|H_{01}\|_F.
$$

Substituting (20) and (22) into (19), we have

$$
\|H_{01}\|_F^2 \leq \delta_{2r} \|H_{01}\|_F^2 + \sqrt{2\delta_{2r}} \|H_{01}\|_F \cdot \sum_{j \geq 2} \|H_j\|_F + 2\eta \sqrt{1 + \delta_{2r}} \|H_{01}\|_F.
$$

Rearranging the terms in the last inequality and dividing both sides by $\|H_{01}\|_F$ yields

$$
\|H_{01}\|_F \leq \beta \sum_{j \geq 2} \|H_j\|_F + C\eta,
$$
where $\beta = \frac{\sqrt{2} \delta_r}{1-\delta_r}$, $C = \frac{2\sqrt{1+\delta_2}}{1-\delta_2}$. \hfill \Box

**Lemma 3.11.** Assume the measurement map $A$ satisfies the M-RIP of order $2r \leq n_1$ with $\delta_2 < 1$, then

$$
\|H_{[r]}\|_* \leq \beta \|H^c_{[r]}\|_* + \sqrt{r} C \eta,
$$

where $\beta = \frac{\sqrt{2} \delta_r}{1-\delta_r}$, $C = \frac{2\sqrt{1+\delta_2}}{1-\delta_2}$.

**Proof.** Note that for $j \geq 2$, $\|H_j\|_F \leq \sqrt{r} \cdot \max\{\sigma(H_j)\} \leq \sqrt{r} \|H_{j-1}\|_* = \frac{\|H_{j-1}\|_*}{\sqrt{r}}$, then

(22)

$$
\sum_{j \geq 2} \|H_j\|_F \leq \sum_{j \geq 1} \frac{\|H_j\|_*}{\sqrt{r}} = \frac{1}{\sqrt{r}} \|H^c_{[r]}\|_*.
$$

Then combining Lemma 3.10 and (22), we get

$$
\|H_{[r]}\|_* \leq \sqrt{r} \|H_{[r]}\|_F \\
\leq \sqrt{r} \|H_{01}\|_F \\
= \sqrt{r} \left( \beta \cdot \frac{1}{\sqrt{r}} \|H^c_{[r]}\|_* + C \eta \right) \\
= \beta \|H^c_{[r]}\|_* + \sqrt{r} C \eta.
$$

The following lemma, which is related to the matrix nuclear norm inequality, is useful for our proof.

**Lemma 3.12.** Assume the measurement map $A$ satisfies the M-RIP of order $2r \leq n_1$ with $\delta_2 < \sqrt{2} - 1$. Let $\beta = \frac{\sqrt{2} \delta_r}{1-\delta_2}$ and $C = \frac{2\sqrt{1+\delta_2}}{1-\delta_2}$, then we have

$$
\|H^c_{[r]}\|_* \leq \frac{1}{1 - \beta} \left( \|\mathbf{X}\|_* - \|X\|_* + 2\|X - X_{[r]}\|_* + \sqrt{r} C \eta \right).
$$

**Proof.** Note that $H^c_{[r]} = \mathcal{P}^c(H)$. We can use $H = X - \mathbf{X}$ and the triangle inequality to get

$$
\|H^c_{[r]}\|_* = \|\mathcal{P}^c(H)\|_* = \|\mathcal{P}^c(X - \mathbf{X})\|_* \\
\leq \|\mathcal{P}^c(X)\|_* + \|\mathcal{P}^c(\mathbf{X})\|_* \\
= \|\mathbf{X}\|_* - \|\mathcal{P}(\mathbf{X})\|_* + 2\|\mathcal{P}(X)\|_* + \|\mathcal{P}(X)\|_* + 2\|\mathcal{P}^c(X)\|_* \\
\leq \|\mathbf{X}\|_* - \|X\|_* + \|\mathcal{P}(X - \mathbf{X})\|_* + 2\|\mathcal{P}^c(X)\|_* \\
= \|\mathbf{X}\|_* - \|X\|_* + \|H_{[r]}\|_* + 2\|X - X_{[r]}\|_*.
$$

Substituting Lemma 3.11 into (23) yields

$$
\|H^c_{[r]}\|_* \leq \beta \|H^c_{[r]}\|_* + \sqrt{r} C \eta + \|\mathbf{X}\|_* - \|X\|_* + 2\|X - X_{[r]}\|_*.
$$

When $\delta_2 < \sqrt{2} - 1$, $1 - \beta > 0$, we can rearrange the terms in the last inequality to get

$$
\|H^c_{[r]}\|_* \leq \frac{1}{1 - \beta} \left( \|\mathbf{X}\|_* - \|X\|_* + 2\|X - X_{[r]}\|_* + \sqrt{r} C \eta \right).
$$

The following lemma, which is related to the matrix nuclear norm inequality, is useful for our proof.

**Lemma 3.12.** Assume the measurement map $A$ satisfies the M-RIP of order $2r \leq n_1$ with $\delta_2 < \sqrt{2} - 1$. Let $\beta = \frac{\sqrt{2} \delta_r}{1-\delta_2}$ and $C = \frac{2\sqrt{1+\delta_2}}{1-\delta_2}$, then we have

$$
\|H^c_{[r]}\|_* \leq \frac{1}{1 - \beta} \left( \|\mathbf{X}\|_* - \|X\|_* + 2\|X - X_{[r]}\|_* + \sqrt{r} C \eta \right).
$$

Next we will give the proof of Theorem 3.4.
Proof of Theorem 3.4. By the optimality of $X$, we get

$$\|X\|_* \leq F_T^*(X) \leq F_T^*(X) = \sum_{i=1}^{n_1} (\sigma_{i}^2(X) + \sigma_i^2) \leq \|X\|_* + n_1 \epsilon.$$

In addition, $n_1 = \frac{1}{\epsilon}$, then

$$n_1 \epsilon = \lim_{k \to \infty} n_1 \epsilon_k \leq \lim_{k \to \infty} n_1 \gamma \sigma_{r+1}(X^k) = \sigma_{r+1}(X).$$

Therefore, for any integer $t < r$, we use (24), (25) and (18) to obtain

$$|X|_r - \|X\|_* \leq \sigma_{r+1}(X) \leq \frac{1}{r-t+1} (\|X - X_{[t]}\|_* + \|X - X_{[t]}\|_*).$$

When $t < r$, $\|X - X_{[t]}\|_* \leq \|X - X_{[t]}\|_*$. Combining Lemma 3.10, Lemma 3.11 and (26) we get

$$\|H\|_* \leq \|H_{[r]}\|_* + \|H_{[r]}^\nu\|_*$$

$$\leq (1 + \beta) \|H_{[r]}\|_* + \sqrt{r} C \eta$$

$$\leq \frac{1 + \beta}{1 - \beta} (\|X\|_* - \|X\|_* + 2\|X - X_{[t]}\|_* + \sqrt{r} C \eta) + \sqrt{r} C \eta$$

$$\leq \frac{1 + \beta}{1 - \beta} \left( \frac{1}{r - t + 1} \|X - X\|_* + \left( 1 + \frac{1}{r - t + 1} \right) \|X - X_{[t]}\|_* + \sqrt{r} C \eta \right)$$

$$+ \sqrt{r} C \eta$$

$$= \frac{1 + \beta}{(1 - \beta)(r - t + 1)} \|X - X\|_* + \frac{1 + \beta}{1 - \beta} \left( 1 + \frac{1}{r - t + 1} \right) \|X - X_{[t]}\|_*$$

$$+ \frac{2}{1 - \beta} \sqrt{r} C \eta.$$

If $t < r - \frac{\beta \alpha}{1 - \beta}$, then $\alpha = \frac{1 + \beta}{1 + \beta(r - t + 1)} < 1$, we have

$$\|X - X\|_* \leq C_1 \eta + C_2 \|X - X_{[t]}\|_*,$$

where $C_1 = \frac{2\sqrt{r}}{(1 - \beta)(1 - \alpha)}$ and $C_2 = \frac{(1 - t + 2)\alpha}{(1 - \beta)(1 - \alpha)}.$ \hfill \Box

4. Conclusions. In this paper, we have given a rigorously theoretical analysis of iteratively reweighted least squares minimization algorithm (abbreviated as IRLS-M) which has been proposed in the literatures for low rank matrix recovery [17, 30, 31] in the presence of noise. We proved that IRLS-M algorithm is convergent and showed that it can recover an unknown (approximately) low rank matrix with an error that is proportional to the noise level, provided the measurement operator satisfies the matrix restricted isometry property (M-RIP). Future studies may be extending the iteratively reweighted least squares minimization algorithm to solve other problems of large practical interest such as phase retrieval problem [13] which has numerous applications in the imaging science and low-rank plus sparse matrix decomposition problem [43] which has been successfully applied to computer vision and image processing, etc.

Acknowledgments. The authors thank the reviewers and the editor for their valuable comments and suggestions.

This work is partially supported by the General Program of Natural Science Research Fund for Colleges and Universities in Jiangsu Province under Grant No. 16KJB110017 and the Youth Fund Project of Zhejiang Provincial NSF under Grant No. LQ14A010005.
REFERENCES

[1] D. Ba, B. Babadi, P. L. Purdon and E. N. Brown, Convergence and stability of iteratively re-weighted least squares algorithms, IEEE Trans. Signal. Process., 62 (2014), 183–195.

[2] R. Basri and D. Jacobs, Lambertian reflectance and linear subspaces IEEE Trans. Pattern Anal. Mach. Intell., 25 (2003), 218–233.

[3] D. P. Bertsekas, A. Nedic and A. E. Ozdaglar, Convex Analysis and Optimization, Athena Scientific, 2003.

[4] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[5] T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, IEEE Trans. Inform. Theory., 60 (2014), 122–132.

[6] Y. Cai and S. Li, Convergence analysis of projected gradient descent for Schatten-p nonconvex matrix recovery Sci. China Math., 58 (2015), 845–858.

[7] E. J. Candès and T. Tao, Decoding by linear programming IEEE Trans. Inform. Theory., 51 (2005), 4203–4215.

[8] E. J. Candès, The restricted isometry property and its implications for compressed sensing C. R. Math. Acad. Sci. Paris, Ser. I, 346 (2008), 589–592.

[9] E. J. Candès and B. Recht, Exact Matrix completion via convex optimization Found. Comput. Math., 9 (2009), 717–772.

[10] E. J. Candès and T. Tao, The power of convex relaxation: Near-optimal matrix completion IEEE Trans. Inform. Theory., 56 (2010), 2053–2080.

[11] E. J. Candès and Y. Plan, Matrix completion with noise, Proceedings of the IEEE, 98 (2009), 925–936.

[12] E. J. Candès and Y. Plan, Tight oracle bounds for low-rank recovery from a minimal number of random measurements IEEE Trans. Inform. Theory., 57 (2011), 2342–2359.

[13] E. J. Candès, Y. Eldar, T. Strohmer and M. V. Stojnic, Phase retrieval via matrix completion Comm. Pure. Appl. Math., 66 (2013), 199–225.

[14] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, International Conference on Acoustics, Speech and Signal Processing, (2008), 3869–3872.

[15] I. Daubechies, R. Devore, M. Fornasier and C. S. Güntük, Iteratively reweighted least squares minimization for sparse recovery Commu. Pure. Appl. Math., 63 (2010), 1–38.

[16] M. Fazel, Matrix Rank Minimization with Applications, Ph.D thesis, Stanford University, 2002.

[17] M. Fornasier, H. Rauhut and R. Ward, Low-rank matrix completion via iteratively reweighted least squares minimization SIAM J. Optim., 21 (2011), 1614–1640.

[18] S. Foucart and M. J. Lai, Sparse solutions of underdetermined linear systems via $l_q$-minimization for $0 < q \leq 1$, Appl. Comput. Harmon. Anal., 26 (2009), 395–407.

[19] M. Grant and S. Boyd, Graph implementations for nonsmooth convex programs, in Recent Advances in Learning and Control (tribute to M. Vidyasagar) (eds. V. Blondel, S. Boyd and H. Kimura), Springer, 2008, 95–110.

[20] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, Available from: http://web.stanford.edu/~boyd/cvxbook

[21] D. Gross, Y. K. Liu, S. T. Flammia, S. Becker and J. Eisert, Quantum state tomography via compressed sensing Phys. Rev. Lett., 105 (2010), 150401.

[22] D. Gross, Recovery Low-rank matrices from few coefficients in any basis IEEE Trans. Inform. Theory., 57 (2011), 1548–1566.

[23] P. Jain, P. Netrapalli and S. Sanghavi, Low-rank matrix completion using alternating minimization in Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, (2013), 665–674.

[24] H. Ji, C. Liu, Z. W. Shen and Y. H. Xu, Robust video denoising using low rank matrix completion, IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), San Francisco, (2010), 1548–1566.

[25] F. Kramer and R. Ward, New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property SIAM J. Math. Anal., 43 (2011), 1269–1281.

[26] M. J. Lai, Y. Xu and W. Yin, Improved iteratively reweighted least square for unconstrained smoothed $l_q$ minimization SIAM J. Numer. Anal., 51 (2013), 927–957.

[27] C. L. Lawson, Contributions to the Theory of Linear Least Maximum Approximation, Ph.D thesis, University of California, Los Angeles, 1961.
Iteratively reweighted least squares

[29] J. Lin and S. Li, Convergence of projected Landweber iteration for matrix rank minimization, Appl. Comput. Harmon. Anal., 36 (2014), 316–325.

[30] Z. Liu and L. Vandenberghe, Interior-point method for nuclear norm approximation with application to system identification, SIAM J. Matrix Anal. Appl., 31 (2009), 1235–1256.

[31] K. Mohan and M. Fazel, Iterative reweighted least squares for matrix rank minimization, In Proc. 48th Allerton Conference on Controls and Communications, Allerton, IL, 2010b.

[32] K. Mohan and M. Fazel, Iterative reweighted algorithms for matrix rank minimization, J. Mach. Learn. Res., 13 (2012), 3441–3473.

[33] T. Morita and T. Kanade, A sequential factorization method for recovering shape and motion from image streams, IEEE Trans. Pattern Anal. Mach. Intell., 19 (1997), 858–867.

[34] Netflix Prize, Available from: http://www.netflixprize.com/

[35] S. Oymak, K. Mohan, M. Fazel and B. Hassibi, A simplified approach to recovery conditions for low rank matrices, in: IEEE International Symposium on Information Theory Proceedings, ISIT, (2011), 2318–2322.

[36] B. Recht, M. Fazel and P. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev., 52 (2010), 471–501.

[37] B. Recht, A simpler approach to matrix completion, J. Mach. Learn. Res., 12 (2011), 3413–3430.

[38] R. Saab and O. Yilmaz, Sparse recovery by non-convex optimization-instance optimality, Appl. Comput. Harmon. Anal., 29 (2010), 30–48.

[39] E. van den Berg and M. P. Friedlander, Sparse optimization with least-squares constraints, SIAM J. Optim., 21 (2011), 1201–1229.

[40] W. Wang, W. Xu and A. Tang, On the performance of sparse recovery via $l_p$-minimization, IEEE Trans. Inform. Theory, 57 (2011), 7255–7278.

[41] K. Wei, J-F. Cai, T. F. Chan and S. Leung, Guarantees of riemannian optimization for low rank matrix recovery, SIAM J. Matrix Anal. Appl., 37 (2016), 1198–1222.

[42] M-C. Yue and A. Man-Cho So, A perturbation inequality for concave functions of singular values and its applications in low-rank matrix recovery, Appl. Comput. Harmon. Anal., 40 (2016), 396–416.

[43] Z Zhou, J. Wright, X. Li, E. J. Candés and Y. Ma, Stable Principal Component Pursuit, Proceedings of International Symposium on Information Theory, (2010).

Received November 2015; revised April 2017.

E-mail address: ycai05@163.com
E-mail address: songli@zju.edu.cn