## A Truncation Algorithms for the IHT

In this section, we present the algorithms for the truncation step in the IHT algorithm, i.e., Line 6 in Algorithm 1. The truncation algorithms for sparse vector and low-rank matrix recovery are tabulated in Algorithms 2 and 3, respectively.

**Algorithm 2** Evaluation of the truncation function $\mathrm{Trunc}(\mathbf{\Theta}, s)$ in Line 6 of Algorithm 1 for nonlinear sparse vector recovery

**Input:** Truncation level $s > 0$, a vector $\mathbf{\Theta} \in \mathbb{R}^d$
Sort $\{|\mathbf{\Theta}_j|\}_{j=1}^d$ such that $|\mathbf{\Theta}_{j_1}| \geq |\mathbf{\Theta}_{j_2}| \geq \ldots \geq |\mathbf{\Theta}_{j_d}|$
$\mathcal{S} \leftarrow \{j_1, \ldots, j_s\}$
**for** $j$ in $\{1, \ldots, d\}$ **do**
    $\mathbf{\Theta}_j \leftarrow 0$ if $j \notin \mathcal{S}$
**end for**
**return** $\mathbf{\Theta}$

**Algorithm 3** Evaluation of the truncation function $\mathrm{Trunc}(\mathbf{\Theta}, s)$ in Line 6 of Algorithm 1 for nonlinear low-rank matrix recovery

**Input:** Truncation level $s > 0$, a low rank matrix $\mathbf{\Theta} \in \mathbb{R}^{m_1 \times m_2}$ with $\mathrm{rank}(\mathbf{\Theta}) = r$
Perform singular value decomposition $\mathbf{\Theta} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{m_1 \times r}$, $\mathbf{\Lambda} \in \mathbb{R}^{r \times r}$, $\mathbf{V} \in \mathbb{R}^{r \times m_2}$. The diagonal elements of $\mathbf{\Lambda}$ are in decreasing order
**for** $j$ in $\{1, \ldots, r\}$ **do**
    $\mathbf{\Lambda}_{jj} \leftarrow 0$ if $j > s$
**end for**
$\mathbf{\Theta} \leftarrow \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$
**return** $\mathbf{\Theta}$

# B Proof of the Main Results

In this section we give a detailed proof of the main results.

## B.1 Proof of Theorem 3.5

*Proof.* We first define $\mathrm{supp}(\mathbf{V}) := \{j : v_j \neq 0\}$, for any $\mathbf{V} = (v_1, \cdots, v_d)^\top \in \mathbb{R}^d$ as the support of $\mathbf{V}$. When proceeding Algorithm 1 for sparse vector recovery, for $t \geq 0$, we denote $\mathcal{S}^{(t)} = \mathrm{supp}(\mathbf{\Theta}^{(t)})$, $\mathcal{S}^* = \mathrm{supp}(\mathbf{\Theta}^*)$ and $\mathcal{F}^{(t)} = \mathcal{S}^{(t)} \cup \mathcal{S}^{(t+1)} \cup \mathcal{S}^*$. By definition, the cardinality of $\mathcal{F}^{(t)}$ is no larger than $2s+s^*$, i.e., $\|\mathbf{\Theta}_{\mathcal{F}^{(t)}}\|_0 \leq 2s+s^*$. We denote $\widetilde{\mathbf{\Theta}}^{(t+1)} = \mathbf{\Theta}^{(t)} - \eta \cdot \nabla_{\mathcal{F}^{(t)}} \ell(\mathbf{\Theta}^{(t)})$, then by definition we have $\mathbf{\Theta}^{(t+1)} = \mathrm{Trunc}(\widetilde{\mathbf{\Theta}}^{(t+1)}, s)$. By denoting $\mathbf{\Delta}^{(t)} = \mathbf{\Theta}^{(t)} - \mathbf{\Theta}^*$, we have

$$\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2 = \|\mathbf{\Delta}^{(t)} - \eta \cdot \nabla_{\mathcal{F}^{(t)}} \ell(\mathbf{\Theta}^{(t)})\|_2. \tag{B.1}$$

According to the definition of the least-squares loss function in (1.2), $\nabla \ell(\mathbf{\Theta}^{(t)})$ is represented as:

$$\nabla \ell(\mathbf{\Theta}^{(t)}) = \frac{1}{n} \sum_{i=1}^{n} \big[f(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle) - Y_i\big] \cdot f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle\big) \cdot \mathbf{X}_i = \mathbf{E}^{(t)} + \mathbf{G}^{(t)}, \tag{B.2}$$

where $\mathbf{E}^{(t)}$ and $\mathbf{G}^{(t)}$ are defined as

$$\mathbf{E}^{(t)} := \frac{1}{n} \sum_{i=1}^{n} \big[f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle\big) - f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^* \rangle\big)\big] \cdot \epsilon_i \cdot \mathbf{X}_i, \text{ and} \tag{B.3}$$

$$\mathbf{G}^{(t)} := \frac{1}{n} \sum_{i=1}^{n} f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle\big) \cdot \big[f\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle\big) - f\big(\langle \mathbf{X}_i, \mathbf{\Theta}^* \rangle\big)\big] \cdot \mathbf{X}_i. \tag{B.4}$$

Triangle inequality yields that

$$\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2 \leq \|\mathbf{\Delta}^{(t)} - \eta \cdot \mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\|_2 + \eta \cdot \|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\|_2. \tag{B.5}$$

Using Mean Value Theorem, $\mathbf{G}^{(t)}$ can be written as

$$\mathbf{G}^{(t)} = \frac{1}{n} \sum_{i=1}^{n} f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle\big) \cdot f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}_1 \rangle\big) \cdot \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)} \rangle \cdot \mathbf{X}_i = \mathbf{A}^{(t)} \cdot \mathbf{\Delta}^{(t)}, \tag{B.6}$$

where $\mathbf{\Theta}_1$ lies between $\mathbf{\Theta}^{(t)}$ and $\mathbf{\Theta}^*$ and $\mathbf{A}^{(t)} = n^{-1} \sum_{i=1}^{n} f'(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle) \cdot f'(\langle \mathbf{X}_i, \mathbf{\Theta}_1 \rangle) \cdot \mathbf{X}_i \mathbf{X}_i^\top$. Since $\mathrm{supp}(\mathbf{G}^{(t)}) \subseteq \mathcal{F}^{(t)}$, we actually have

$$\mathbf{G}^{(t)} = \mathbf{A}^{(t)}_{\cdot, \mathcal{F}^{(t)}} \cdot \mathbf{\Delta}^{(t)}. \tag{B.7}$$

Therefore, combining (B.5), (B.6), and (B.7), we obtain

$$\begin{aligned} \big\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\big\|_2 &\leq \big\|\mathbf{\Delta}^{(t)} - \eta \cdot \mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\big\|_2 + \eta \cdot \big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_2 \\ &\leq \big\|\mathbf{I} - \eta \cdot \mathbf{A}^{(t)}_{\mathcal{F}^{(t)}, \mathcal{F}^{(t)}}\big\|_2 \cdot \big\|\mathbf{\Delta}^{(t)}\big\|_2 + \eta \cdot \big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_2, \end{aligned} \tag{B.8}$$

where $\mathbf{I} \in \mathbb{R}^{|\mathcal{F}^{(t)}| \times |\mathcal{F}^{(t)}|}$ is the identity matrix. Here the second inequality follows from the definition of operator norm. Since $|\mathcal{F}^{(t)}| \leq 2s + s^*$, by Assumptions 3.1 and 3.2 we have

$$a^2[1 - \delta(2s+s^*)] \cdot \|\mathbf{V}\|_2 \leq \mathbf{V}^\top \mathbf{A}^{(t)}_{\mathcal{F}^{(t)}, \mathcal{F}^{(t)}} \mathbf{V} \leq b^2[1 + \delta(2s+s^*)] \cdot \|\mathbf{V}\|_2$$

for any $\mathbf{V} \in \mathbb{R}^{|\mathcal{F}^{(t)}|}$. Therefore, given $\eta$ we can bound $\big\|\mathbf{I} - \eta \cdot \mathbf{A}^{(t)}_{\mathcal{F}^{(t)}, \mathcal{F}^{(t)}}\big\|_2$ by

$$\lambda_1(\eta) := \max\big\{1 - \eta a^2[1 - \delta(2s+s^*)], \eta b^2[1 + \delta(2s+s^*)] - 1\big\}. \tag{B.9}$$

By choosing $3/7 \cdot \{a^2[1-\delta(2s+s^*)]\}^{-1} < \eta < 11/7 \cdot \{b^2[1+\delta(2s+s^*)]\}^{-1}$, we have $\lambda_1(\eta) \in (0, 4/7)$. Note that given $a$ and $b$, such $\eta$ always exists as long as the constant $\delta(2s+s^*)$ is sufficiently small.

Moreover, due to the property of the norms,

$$\big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_2 \le \sqrt{2s+s^*} \cdot \big\|\mathbf{E}^{(t)}\big\|_\infty. \tag{B.10}$$

Therefore, combining (B.8), (B.9), and (B.10), we obtain that

$$\big\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\big\|_2 \le \lambda_1(\eta) \cdot \big\|\mathbf{\Delta}^{(t)}\big\|_2 + \eta\sqrt{2s+s^*} \cdot \big\|\mathbf{E}^{(t)}\big\|_\infty. \tag{B.11}$$

The following lemma further upper bounds $\big\|\mathbf{E}^{(t)}\big\|_\infty$ with high probability.

**Lemma B.1.** For $\mathbf{E}^{(t)}$ defined in (B.3), if we assume $\mathbf{X}_i \in \mathbb{R}^d$ satisfies Assumption 3.4, then it holds with probability at least $1-d^{-1}$ that

$$\big\|\mathbf{E}^{(t)}\big\|_\infty \le 2b\sigma D \cdot \sqrt{\log d/n}, \quad \forall t \ge 0. \tag{B.12}$$

*Proof.* Recall that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. sub-Gaussian random variables. By the definition of $\mathbf{E}^{(t)}$, for $j \in \{1, \ldots, d\}$, each entry $E_j^{(t)}$ is a sub-Gaussian random variable with variance proxy given by

$$\tau_t := \frac{\sigma^2}{n^2} \sum_{i=1}^n [f'(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)} \rangle) - f'\langle \mathbf{X}_i, \mathbf{\Theta}^* \rangle)]^2 X_{ij}^2 \le 4b^2\sigma^2 D^2/n, \quad \forall t \ge 0,$$

where $X_{ij}$ denotes the $j$-th entry in $\mathbf{X}_i$. Here the inequality follows from the assumption that $f'(u) \le b$ for all $u \in \mathbb{R}$ and Assumption 3.4. By the definition of sub-Gaussian random vectors, for any $\boldsymbol{A} \in \mathbb{R}^d$, we have

$$\max_{t \ge 0} \mathbb{E}[\exp(\boldsymbol{A}^\top \mathbf{E}^{(t)})] \le \max_{t \ge 0} \exp(\|\boldsymbol{A}\|_2^2 \cdot \tau_t^2) \le \exp(\|\boldsymbol{A}\|_2^2 \cdot 4b^2\sigma^2 D^2/n).$$

By the tail inequality for sub-Gaussian random variables, we conclude that

$$\|\mathbf{E}^{(t)}\|_\infty \le 2b\sigma D \cdot \sqrt{\log d/n}, \quad \forall t \ge 0$$

with probability at least $1-d^{-1}$. Thus, we conclude the proof of the lemma. □

Now we use the following lemma from Jain et al. (2014) to characterize the relation between $\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2$ and $\|\mathbf{\Theta}^{(t+1)} - \mathbf{\Theta}^*\|_2$.

**Lemma B.2.** For any $\mathbf{\Theta} \in \mathbb{R}^k$ and integer $s \le k$, let $\mathbf{\Theta}_t = \text{Trunc}(\mathbf{\Theta}, s)$. Then for any $\mathbf{\Theta}^* \in \mathbb{R}^k$ such that $\|\mathbf{\Theta}^*\|_0 \le s^*$, we have $\|\mathbf{\Theta}_t - \mathbf{\Theta}\|_2^2 \le (k-s)/(k-s^*) \cdot \|\mathbf{\Theta}^* - \mathbf{\Theta}\|_2^2$.

*Proof.* See Lemma 1 of Jain et al. (2014) for a detailed proof. □

Applying Lemma B.2 with $\mathbf{\Theta} = \widetilde{\mathbf{\Theta}}^{(t+1)}$, we have

$$\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^{(t+1)}\|_2^2 \le \frac{|\mathcal{F}^{(t)}| - s}{|\mathcal{F}^{(t)}| - s^*} \cdot \|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2^2 \le \frac{s+s^*}{2s} \cdot \|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2^2$$

where the second inequality follows from $|\mathcal{F}^{(t)}| \le 2s + s^*$. From Assumption 3.2 that $s \ge 8s^*$, we further have

$$\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^{(t+1)}\|_2^2 \le 9/16 \cdot \|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2^2,$$

Therefore, we obtain

$$\|\mathbf{\Theta}^{(t+1)} - \mathbf{\Theta}^*\|_2 \le \|\mathbf{\Theta}^{(t+1)} - \widetilde{\mathbf{\Theta}}^{(t+1)}\|_2 + \|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2 \le 7/4 \cdot \|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\|_2. \tag{B.13}$$

Finally we obtain the main result by combining (B.11), (B.12), and (B.13):

$$\|\mathbf{\Theta}^{(t+1)} - \mathbf{\Theta}^*\|_2 \le \mu_1^t \cdot \|\mathbf{\Theta}^{(0)} - \mathbf{\Theta}^*\|_2 + C_1 \cdot \sqrt{(2s+s^*)\log d/n},$$

where $\mu_1 \in (0,1)$ and $C_1 > 0$ is an absolute constant. Therefore, we conclude the proof of the theorem. □

### B.2 Proof of Theorem 3.6

*Proof.* The proof has the similar procedure as the proof of Theorem 3.5. When Algorithm 1 is adopted for low-rank matrix recovery, for $t \geq 0$, we define $\mathcal{S}^{(t)}$ and $\mathcal{S}^*$ as

$$\mathcal{S}^{(t)} := \big\{\mathbf{V} \in \mathbb{R}^{m_1 \times m_2} \ : \mathrm{row}\big(\mathbf{V}\big) \subseteq \mathrm{row}\big(\mathbf{\Theta}^{(t)}\big), \mathrm{col}\big(\mathbf{V}\big) \subseteq \mathrm{col}\big(\mathbf{\Theta}^{(t)}\big)\big\},$$
$$\mathcal{S}^* := \big\{\mathbf{V} \in \mathbb{R}^{m_1 \times m_2} \ : \mathrm{row}\big(\mathbf{V}\big) \subseteq \mathrm{row}\big(\mathbf{\Theta}^*\big), \mathrm{col}\big(\mathbf{V}\big) \subseteq \mathrm{col}\big(\mathbf{\Theta}^*\big)\big\}.$$

Therefore, any matrix in subspace $\mathcal{F}^{(t)} := \mathcal{S}^{(t)} \cup \mathcal{S}^{(t+1)} \cup \mathcal{S}^*$ has rank no larger than $2s+s^*$, i.e., $\mathrm{rank}\big(\mathbf{\Theta}_{\mathcal{F}^{(t)}}\big) \leq 2s+s^*$, for any $\mathbf{\Theta} \in \mathbb{R}^{m_1 \times m_2}$. Therefore, by denoting $\widetilde{\mathbf{\Theta}}^{(t+1)} = \mathbf{\Theta}^{(t)} - \eta \cdot \nabla_{\mathcal{F}^{(t)}} \ell(\mathbf{\Theta}^{(t)})$ and $\mathbf{\Theta}^{(t+1)} = \mathrm{Trunc}(\widetilde{\mathbf{\Theta}}^{(t+1)}, s)$, we obtain the same results as (B.1) and (B.2) for low-rank matrix recovery and have the same definition for $\mathbf{E}^{(t)}$ and $\mathbf{G}^{(t)}$ as (B.3) and (B.4), respectively. With triangle inequality for the Frobenius norm, we obtain

$$\big\|\widetilde{\mathbf{\Theta}}^{(t+1)} - \mathbf{\Theta}^*\big\|_F \leq \big\|\mathbf{\Delta}^{(t)} - \eta \cdot \mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F + \eta \cdot \big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F. \tag{B.14}$$

We need to upper bound the two terms on the right hand side.

Using the Mean Value Theorem, $\mathbf{G}^{(t)}$ can be written as

$$\mathbf{G}^{(t)} = \frac{1}{n}\sum_{i=1}^{n} f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)}\rangle\big) \cdot f'\big(\langle \mathbf{X}_i, \mathbf{\Theta}_1\rangle\big) \cdot \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)}\rangle \cdot \mathbf{X}_i = \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)}\rangle \cdot \mathbf{X}_i,$$

where $\mathbf{\Theta}_1$ lies between $\mathbf{\Theta}^{(t)}$ and $\mathbf{\Theta}^*$ and $B_i = f'(\langle \mathbf{X}_i, \mathbf{\Theta}^{(t)}\rangle) \cdot f'(\langle \mathbf{X}_i, \mathbf{\Theta}_1\rangle)$. Hence by definition, $\big\|\mathbf{\Delta}^{(t)} - \eta \cdot \mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F$ can be written as

$$\big\|\mathbf{\Delta}^{(t)} - \eta \cdot \mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F = \sup_{\widetilde{\mathbf{\Delta}} \in \widetilde{\mathcal{S}}^{(t)}} \bigg|\langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle - \eta \cdot \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)}\rangle\langle \widetilde{\mathbf{\Delta}}, \mathbf{X}_i\rangle\bigg| \tag{B.15}$$

where $\widetilde{\mathcal{S}}^{(t)} := \big\{\widetilde{\mathbf{\Delta}} \in \mathcal{F}^{(t)} : \|\widetilde{\mathbf{\Delta}}\|_F = 1\big\}$. Using triangle inequality, we have

$$\begin{aligned}
&\sup_{\widetilde{\mathbf{\Delta}} \in \widetilde{\mathcal{S}}^{(t)}} \bigg|\langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle - \eta \cdot \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)}\rangle\langle \widetilde{\mathbf{\Delta}}, \mathbf{X}_i\rangle\bigg| \\
&\leq \sup_{\widetilde{\mathbf{\Delta}} \in \widetilde{\mathcal{S}}^{(t)}} \bigg|\langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle - \eta \cdot \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle\bigg| + \\
&\qquad \bigg|\eta \cdot \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \big(\langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle - \langle \mathbf{X}_i, \mathbf{\Delta}^{(t)}\rangle\langle \widetilde{\mathbf{\Delta}}, \mathbf{X}_i\rangle\big)\bigg|
\end{aligned} \tag{B.16}$$

The first term on the right hand side of (B.16) is bounded as

$$\sup_{\widetilde{\mathbf{\Delta}} \in \widetilde{\mathcal{S}}^{(t)}} \bigg|\langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle - \eta \cdot \frac{1}{n}\sum_{i=1}^{n} B_i \cdot \langle \widetilde{\mathbf{\Delta}}, \mathbf{\Delta}^{(t)}\rangle\bigg| \leq \max\{1 - \eta a^2, \eta b^2 - 1\} \cdot \|\mathbf{\Delta}^{(t)}\|_F, \tag{B.17}$$

due to the boundness of the derivative $f'$.

For the second term on the right hand side of (B.16), we introduce the following lemma from Carpentier and Kim (2015) to bound it.

**Lemma B.3.** Under Assumption 3.2, for any $s \leq 8s^*$, we have that

$$\sup_{\mathbf{A},\mathbf{B} \in \mathcal{R}(s)} \bigg|\frac{1}{n}\sum_{i=1}^{n} \langle \mathbf{X}_i, \mathbf{A}\rangle\langle \mathbf{X}_i, \mathbf{B}\rangle - \langle \mathbf{A}, \mathbf{B}\rangle\bigg| \leq 2\delta(s) \cdot \|\mathbf{A}\|_F \|\mathbf{B}\|_F, \tag{B.18}$$

where we denote $\mathcal{R}(s) := \{\mathbf{V} \in \mathbb{R}^{m_1 \times m_2} \colon |\mathrm{row}(\mathbf{V})| \leq s, |\mathrm{col}(\mathbf{V})| \leq s\}$.

*Proof.* See Lemma 5.1. of Carpentier and Kim (2015) for a detailed proof. □

By applying Lemma B.3, we have

$$\sup_{\widetilde{\boldsymbol{\Delta}}\in\widetilde{\mathcal{S}}^{(t)}}\left|\eta\cdot\frac{1}{n}\sum_{i=1}^{n}B_i\cdot\big(\langle\widetilde{\boldsymbol{\Delta}},\boldsymbol{\Delta}^{(t)}\rangle-\langle\mathbf{X}_i,\boldsymbol{\Delta}^{(t)}\rangle\langle\widetilde{\boldsymbol{\Delta}},\mathbf{X}_i\rangle\big)\right|\leq 2\eta b^2\delta(2s+s^*)\cdot\|\boldsymbol{\Delta}^{(t)}\|_F. \tag{B.19}$$

Thus (B.15) can be upper bounded by combining (B.17) and (B.19)

$$\big\|\boldsymbol{\Delta}^{(t)}-\eta\cdot\mathbf{G}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F\leq\lambda_2(\eta):=\max\{1-\eta a^2,\eta b^2-1\}+2\eta b^2\delta(2s+s^*). \tag{B.20}$$

By choosing $b^{-2}<\eta<11/7\cdot b^{-2}\cdot[1+2\delta(2s+s^*)]^{-1}$, or $3/7\cdot[a^2-2b^2\delta(2s+s^*)]^{-1}<\eta<a^{-2}$, we have $\lambda_2(\eta)<4/7$. Note that such $\eta$ always exists as long as the constant $\delta(2s+s^*)$ in the RIP condition is sufficiently small.

For the term $\big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F$, we also have

$$\big\|\mathbf{E}^{(t)}_{\mathcal{F}^{(t)}}\big\|_F\leq\sqrt{2s+s^*}\cdot\big\|\mathbf{E}^{(t)}\big\|_2 \tag{B.21}$$

due to the relation between the Frobenius norm and the operator norm. Moreover, we can further upper bound $\|\mathbf{E}^{(t)}\|_2$ with high probability for low-rank matrix recovery depending on different assumptions on $\boldsymbol{X}$. For example, under the assumption that $\mathbf{X}_i$ are i.i.d. sampled from the $\boldsymbol{\Sigma}$-ensemble with some positive definite $\boldsymbol{\Sigma}$, we have $\big\|\mathbf{E}^{(t)}\big\|_2=\mathcal{O}_{\mathbb{P}}(\sqrt{m_1+m_2}/\sqrt{n})$ (Negahban and Wainwright, 2011) (See more discussions in Remark 3.7).

To characterize the relation between $\|\widetilde{\boldsymbol{\Theta}}^{(t+1)}-\boldsymbol{\Theta}^*\|_F$ and $\|\boldsymbol{\Theta}^{(t+1)}-\boldsymbol{\Theta}^*\|_F$, we use another lemma in Jain et al. (2014) for matrix recovery.

**Lemma B.4.** For any $\boldsymbol{\Theta}\in\mathbb{R}^{m_1\times m_2}$ with rank $k$ and integer $s\leq k$, let $\boldsymbol{\Theta}_1=\text{Trunc}(\boldsymbol{\Theta},s)$. Then for any $\boldsymbol{\Theta}^*\in\mathbb{R}^{m_1\times m_2}$ with $\text{rank}(\boldsymbol{\Theta}^*)\leq s^*$, we have $\|\boldsymbol{\Theta}_1-\boldsymbol{\Theta}\|_F^2\leq(k-s)/(k-s^*)\cdot\|\boldsymbol{\Theta}^*-\boldsymbol{\Theta}\|_F^2$.

*Proof.* See Lemma 2 of Jain et al. (2014) for a detailed proof. □

By applying Lemma B.4, we arrive at the similar result as (B.13),

$$\|\boldsymbol{\Theta}^{(t+1)}-\boldsymbol{\Theta}^*\|_F\leq\|\boldsymbol{\Theta}^{(t+1)}-\widetilde{\boldsymbol{\Theta}}^{(t+1)}\|_F+\|\widetilde{\boldsymbol{\Theta}}^{(t+1)}-\boldsymbol{\Theta}^*\|_F\leq 7/4\cdot\|\widetilde{\boldsymbol{\Theta}}^{(t+1)}-\boldsymbol{\Theta}^*\|_F. \tag{B.22}$$

Finally we obtain the main results for low-rank matrix recovery by combining (B.14), (B.20), (B.21), and (B.22):

$$\begin{aligned}\|\boldsymbol{\Theta}^{(t)}-\boldsymbol{\Theta}^*\|_F&\leq\mu_2^t\cdot\|\boldsymbol{\Theta}^{(0)}-\boldsymbol{\Theta}^*\|_F+\eta\sqrt{2s+s^*}\cdot\big\|\mathbf{E}^{(t)}\big\|_2\\&\leq\mu_2^t\cdot\|\boldsymbol{\Theta}^{(0)}-\boldsymbol{\Theta}^*\|_F+C_2\sqrt{2s+s^*}\cdot\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\mathbf{X}_i\right\|_2,\end{aligned}$$

with $\mu_2\in(0,1)$ and $C_2=2b\eta>0$ as an absolute constant. Thus, we conclude the proof of the theorem. □