Supplementary Material for *Fast and Provable Algorithms for Spectrally Sparse Signal Reconstruction via Low-Rank Hankel Matrix Completion*

Jian-Feng Cai *   Tianming Wang †   Ke Wei ‡

March 12, 2017

Abstract

We establish theoretical recovery guarantees of FIHT for multi-dimensional spectrally sparse signal reconstruction problems, which are straightforward extensions of what we have proved for one-dimensional signals in [1]. Assume the underlying multi-dimensional spectrally sparse signal is of model order r and total dimension N. We show that $O(r^2 \log^2(N))$ number of measurements are sufficient for FIHT with resampling initialization to achieve reliable reconstruction provided the signal satisfies the incoherence property.

1 Recovery Guarantees

Without loss of generality, we discuss the three-dimensional setting. Recall that a three-dimensional array $\boldsymbol{X} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$ is spectrally sparse if

$$\boldsymbol{X}(l_1, l_2, l_3) = \sum_{k=1}^{r} d_k y_k^{l_1} z_k^{l_2} w_k^{l_3}, \quad \forall \, (l_1, l_2, l_3) \in [N_1] \times [N_2] \times [N_3]$$

with

$$y_k = \exp(2\pi\imath f_{1k} - \tau_{1k}), \; z_k = \exp(2\pi\imath f_{2k} - \tau_{2k}), \text{ and } w_k = \exp(2\pi\imath f_{3k} - \tau_{3k})$$

for frequency triples $\boldsymbol{f}_k = (f_{1k}, f_{2k}, f_{3k}) \in [0,1)^3$ and dampling factor triples $\boldsymbol{\tau}_k = (\tau_{1k}, \tau_{2k}, \tau_{3k}) \in \mathbb{R}_+^3$. Concatenating the columns of $\boldsymbol{X}$, we get a signal $\boldsymbol{x}$ of length $N_1 N_2 N_3$. Define $N = N_1 N_2 N_3$. We form a three-fold Hankel matrix $\mathcal{H}\boldsymbol{x}$, which has Vandermonde decomposition in the form $\mathcal{H}\boldsymbol{x} = \boldsymbol{E}_L \boldsymbol{D} \boldsymbol{E}_R^T$, where the k-th columns $(1 \leq k \leq r)$ of $\boldsymbol{E}_L$ and $\boldsymbol{E}_R$ are given by

$$\begin{aligned}
\boldsymbol{E}_L^{(:,k)} &= \left\{ y_k^{l_1} z_k^{l_2} w_k^{l_3}, \; (l_1, l_2, l_3) \in [p_1] \times [p_2] \times [p_3] \right\}, \\
\boldsymbol{E}_R^{(:,k)} &= \left\{ y_k^{l_1} z_k^{l_2} w_k^{l_3}, \; (l_1, l_2, l_3) \in [q_1] \times [q_2] \times [q_3] \right\},
\end{aligned}$$

*Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China. Email: jfcai@ust.hk

†Department of Mathematics, University of Iowa, Iowa City, Iowa, USA. Email: tianming-wang@uiowa.edu

‡Department of Mathematics, University of California, Davis, California, USA. Email: kewei@math.ucdavis.edu

where $p_i + q_i = N_i + 1$ for $1 \leq i \leq 3$ and $\boldsymbol{D} = \mathrm{diag}(d_1, \cdots, d_r)$ is a diagonal matrix. It can be verified that if all d_k's are non-zeros and there exists i, $1 \leq i \leq 3$, such that all f_{ik}'s are distinct, $\mathcal{H}\boldsymbol{x}$ is a rank r matrix. The incoherence property is defined similarly.

Definition 1. *The rank r three-fold Hankel matrix $\mathcal{H}\boldsymbol{x}$ with the Vandermonde decomposition $\mathcal{H}\boldsymbol{x} = \boldsymbol{E}_L \boldsymbol{D} \boldsymbol{E}_R^T$ is said to be μ_0-incoherent if there exists a numerical constant $\mu_0 > 0$ such that*

$$\sigma_{\min}(\boldsymbol{E}_L^* \boldsymbol{E}_L) \geq \frac{p_1 p_2 p_3}{\mu_0}, \ \sigma_{\min}(\boldsymbol{E}_R^* \boldsymbol{E}_R) \geq \frac{q_1 q_2 q_3}{\mu_0}.$$

From [3, Thm. 1], in the undamping case, if the minimum wrap-around distance between the frequencies $\{f_{ik}\}_{k=1}^r$ is greater than about $\frac{2}{N_i}$ for $1 \leq i \leq 3$, this property can be satisfied. Let $\mathcal{H}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ be the reduced SVD of $\mathcal{H}\boldsymbol{x}$ and $\mathcal{P}_{\boldsymbol{U}}(\cdot)$ and $\mathcal{P}_{\boldsymbol{V}}(\cdot)$ respectively be the orthogonal projections onto the subspaces spanned by $\boldsymbol{U}$ and $\boldsymbol{V}$. The following lemma follows directly from Def. 1.

Lemma 1. *Let $\mathcal{H}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* = \boldsymbol{E}_L \boldsymbol{D} \boldsymbol{E}_R^T$. Define $c_s = \max\{\frac{N_1}{p_1}\frac{N_2}{p_2}\frac{N_3}{p_3}, \frac{N_1}{q_1}\frac{N_2}{q_2}\frac{N_3}{q_3}\}$. Assume $\mathcal{H}\boldsymbol{x}$ is μ_0 incoherent, then*

$$\left\|\boldsymbol{U}^{(i,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N} \quad \textit{and} \quad \left\|\boldsymbol{V}^{(j,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N}, \tag{1}$$

$$\|\mathcal{P}_{\boldsymbol{U}}(\boldsymbol{H}_a)\|_F^2 \leq \frac{\mu_0 c_s r}{N} \quad \textit{and} \quad \|\mathcal{P}_{\boldsymbol{V}}(\boldsymbol{H}_a)\|_F^2 \leq \frac{\mu_0 c_s r}{N}, \tag{2}$$

where $\{\boldsymbol{H}_a\}_{a=0}^{N-1}$ forms an orthonormal basis of the three-fold Hankel matrices.

Proof. The proof of (2) can be found in [2]. We include the proof here to be self-contained. We only prove the left inequalities of (1) and (2) as the right ones can be similarly established. Since $\boldsymbol{U} \in \mathbb{C}^{(p_1p_2p_3)\times r}$ and $\boldsymbol{E}_l \in \mathbb{C}^{(p_1p_2p_3)\times r}$ spans the same subspace and $\boldsymbol{U}$ is orthogonal, there exists an orthonormal matrix $\boldsymbol{Q} \in \mathbb{C}^{r\times r}$ such that $\boldsymbol{U} = \boldsymbol{E}_L(\boldsymbol{E}_L^*\boldsymbol{E}_L)^{-1/2}\boldsymbol{Q}$. So

$$\left\|\boldsymbol{U}^{(i,:)}\right\|^2 = \left\|\boldsymbol{e}_i^*\boldsymbol{E}_L(\boldsymbol{E}_L^*\boldsymbol{E}_L)^{-1/2}\right\|^2 \leq \|\boldsymbol{e}_i^*\boldsymbol{E}_L\|^2 \left\|(\boldsymbol{E}_L^*\boldsymbol{E}_L)^{-1}\right\| \leq \frac{\mu_0 r}{p_1p_2p_3} \leq \frac{\mu_0 c_s r}{N}$$

and

$$\|\mathcal{P}_{\boldsymbol{U}}(\boldsymbol{H}_a)\|_F^2 = \|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{H}_a\|_F^2 = \left\|\boldsymbol{E}_L(\boldsymbol{E}_L^*\boldsymbol{E}_L)^{-1}\boldsymbol{E}_L^*\boldsymbol{H}_a\right\|_F^2 \leq \frac{\|\boldsymbol{E}_L^*\boldsymbol{H}_a\|_F^2}{\sigma_{\min}(\boldsymbol{E}_L^*\boldsymbol{E}_L)} \leq \frac{\mu_0 r}{p_1p_2p_3} \leq \frac{\mu_0 c_s r}{N},$$

where we have used the fact that $\boldsymbol{H}_a$ has at most one nonzero element in every row and every column and it only has w_a nonzero entries of magnitude $1/\sqrt{w_a}$ and the magnitudes of the entries of $\boldsymbol{E}_L$ is bounded above by one for both the damped and undamped case. □

1.1 Initialization via One Step Hard Thresholding

Our first initial guess is $\boldsymbol{L}_0 = p^{-1}\mathcal{T}_r(\mathcal{H}\mathcal{P}_\Omega(\boldsymbol{x}))$, which is obtained by truncating the three-fold Hankel matrix constructed from m observed entries of $\boldsymbol{x}$. The following lemma which is of independent interest bounds the deviation of $\boldsymbol{L}_0$ from $\mathcal{H}\boldsymbol{x}$.

Lemma 2. *Assume $\mathcal{H}\boldsymbol{x}$ is μ_0-incoherent. Then there exists a universal constant $C > 0$ such that*

$$\|\boldsymbol{L}_0 - \mathcal{H}\boldsymbol{x}\| \leq C\sqrt{\frac{\mu_0 c_s r \log(N)}{m}}\|\mathcal{H}\boldsymbol{x}\|$$

with probability at least $1 - N^{-2}$.

The following theoretical recovery guarantee can be established for FIHT based on this lemma.

Theorem 1 (Guarantee I)**.** *Assume $\mathcal{H}\boldsymbol{x}$ is μ_0-incoherent. Let $0 < \varepsilon_0 < \frac{1}{10}$ be a numerical constant and $\nu = 10\varepsilon_0 < 1$. Then with probability at least $1 - 3N^{-2}$, the iterates generated by FIHT with the initial guess $\boldsymbol{L}_0 = p^{-1}\mathcal{T}_r(\mathcal{H}\mathcal{P}_\Omega(\boldsymbol{x}))$ satisfy*

$$\|\boldsymbol{x}_l - \boldsymbol{x}\| \leq \nu^l \|\boldsymbol{L}_0 - \mathcal{H}\boldsymbol{x}\|_F,$$

provided

$$m \geq C \max\left\{\varepsilon_0^{-2}\mu_0 c_s, (1+\varepsilon_0)\varepsilon_0^{-1}\mu_0^{1/2}c_s^{1/2}\right\}\kappa r N^{1/2}\log^{3/2}(N)$$

for some universal constant $C > 0$, where $\kappa = \frac{\sigma_{\max}(\mathcal{H}\boldsymbol{x})}{\sigma_{\min}(\mathcal{H}\boldsymbol{x})}$ denotes the condition number of $\mathcal{H}\boldsymbol{x}$.

Remark 1. Since $\mathcal{H}\boldsymbol{x} = \boldsymbol{E}_L\boldsymbol{D}\boldsymbol{E}_R^T$, we have

$$\kappa \leq \frac{\sigma_{\max}(\boldsymbol{E}_L)}{\sigma_{\min}(\boldsymbol{E}_L)} \cdot \frac{\max_k |d_k|}{\min_k |d_k|} \cdot \frac{\sigma_{\max}(\boldsymbol{E}_R)}{\sigma_{\min}(\boldsymbol{E}_R)}.$$

It follows from [3, Thm. 1] that $\sigma_{\max}(\boldsymbol{E}_L)$ (resp. $\sigma_{\max}(\boldsymbol{E}_R)$) and $\sigma_{\min}(\boldsymbol{E}_L)$ (resp. $\sigma_{\min}(\boldsymbol{E}_R)$) are both proportional to $\sqrt{p_1p_2p_3}$ (resp. $\sqrt{q_1q_2q_3}$) when the frequencies of $\boldsymbol{x}$ are well separated. Thus the condition number of $\mathcal{H}\boldsymbol{x}$ is essentially proportional to the dynamical range $\max_k |d_k| / \min_k |d_k|$.

Since the number of measurements required in Thm. 1 is proportional to c_s, it makes sense to set p_i to be about the same as q_i for $1 \leq i \leq 3$.

1.2 Initialization via Resampling and Trimming

To eliminate the dependence on $\sqrt{N}$, we investigate another initialization procedure via resampling and trimming. The following lemma provides an estimation of the approximation accuracy of the initial guess returned by the Alg. 3.

Lemma 3. *Assume $\mathcal{H}\boldsymbol{x}$ is μ_0-incoherent. Then with probability at least $1-(2L+1)N^{-2}$, the output of Alg. 3 satisfies*

$$\|\widetilde{\boldsymbol{L}}_L - \mathcal{H}\boldsymbol{x}\|_F \leq \left(\frac{5}{6}\right)^L \frac{\sigma_{\min}(\mathcal{H}\boldsymbol{x})}{256\kappa^2}$$

provided $\widehat{m} \geq C\mu_0 c_s \kappa^6 r^2 \log(N)$ for some universal constant $C > 0$.

We can obtain the following recovery guarantee for FIHT with $\boldsymbol{L}_0$ being the output of Alg. 3.

Theorem 2 (Guarantee II)**.** *Assume $\mathcal{H}\boldsymbol{x}$ is μ_0-incoherent. Let $0 < \varepsilon_0 < \frac{1}{10}$ and $L = \left\lceil 6\log\left(\frac{\sqrt{N}\log(N)}{16\varepsilon_0}\right)\right\rceil$. Define $\nu = 10\varepsilon_0 < 1$. Then with probability at least $1 - (2L+3)N^{-2}$, the iterates generated by FIHT with $\boldsymbol{L}_0 = \widetilde{\boldsymbol{L}}_L$ (the output of Alg. 3) satisfies*

$$\|\boldsymbol{x}_l - \boldsymbol{x}\| \leq \nu^l \|\boldsymbol{L}_0 - \mathcal{H}\boldsymbol{x}\|_F,$$

provided

$$m \geq C\mu_0 c_s \kappa^6 r^2 \log(N) \log\left(\frac{\sqrt{N}\log(N)}{16\varepsilon_0}\right)$$

for some universal constant $C > 0$.

2 Proofs

We first introduce several new variables and notation. Recall that $\mathcal{H}$ is an which maps a vector to a three-fold Hankel matrix and $\mathcal{H}^*$ is the adjoint of $\mathcal{H}$. Moreover, $\mathcal{D}^2 = \mathcal{H}^*\mathcal{H} = \operatorname{diag}(w_0, \cdots, w_{N-1})$ is a diagonal operator which multiply the a-th entry of a vector by the number of nonzero elements in $\boldsymbol{H}_a$. Define $\mathcal{G} = \mathcal{H}\mathcal{D}^{-1}$. Then the adjoint of $\mathcal{G}$ is given by $\mathcal{G}^* = \mathcal{D}^{-1}\mathcal{H}^*$. It can be easily verified that $\mathcal{G}$ and $\mathcal{G}^*$ have the following properties:

- $\mathcal{G}^*\mathcal{G} = \mathcal{I}$, $\|\mathcal{G}\| = 1$ and $\|\mathcal{G}^*\| \leq 1$;
- $\mathcal{G}\boldsymbol{z} = \sum_{a=0}^{N-1} z_a \boldsymbol{H}_a, \forall \boldsymbol{z} \in C^N$;
- $\mathcal{G}^*\boldsymbol{Z} = \{\langle \boldsymbol{Z}, \boldsymbol{H}_a\rangle\}_{a=0}^{N-1}, \forall \boldsymbol{Z} \in C^{(p_1p_2p_3)\times(q_1q_2q_3)}$.

Notice that the iteration of FIHT can be written in a compact form

$$\boldsymbol{x}_{l+1} = \mathcal{H}^\dagger \mathcal{T}_r \mathcal{P}_{\mathcal{S}_l} \mathcal{H}(\boldsymbol{x}_l + p^{-1}\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{x}_l)). \tag{3}$$

So if we define $\boldsymbol{y} = \mathcal{D}\boldsymbol{x}$ and $\boldsymbol{y}_l = \mathcal{D}\boldsymbol{x}_l$, the following iteration can be established for $\boldsymbol{y}_l$

$$\boldsymbol{y}_{l+1} = \mathcal{G}^* \mathcal{T}_r \mathcal{P}_{\mathcal{S}_l} \mathcal{G}(\boldsymbol{y}_l + p^{-1}\mathcal{P}_\Omega(\boldsymbol{y} - \boldsymbol{y}_l)) \tag{4}$$

since $\mathcal{P}_\Omega$ and $\mathcal{D}^{-1}$ commute with each other. *For ease of exposition, we will prove the lemmas and theorems in terms of* $\boldsymbol{y}_l$ *and* $\boldsymbol{y}$ *but note that the results in terms of* $\boldsymbol{x}_l$ *and* $\boldsymbol{x}$ *follow immediately since* $\mathcal{H}\boldsymbol{x} = \mathcal{G}\boldsymbol{y}$ *and*

$$\|\boldsymbol{x}_l - \boldsymbol{x}\| = \left\|\mathcal{D}^{-1}(\boldsymbol{y}_l - \boldsymbol{y})\right\| \leq \|\boldsymbol{y}_l - \boldsymbol{y}\|. \tag{5}$$

The following supplementary results from the literature but using our notation will be used repeatedly in the proofs of the main results.

Lemma 4 ([4, Proposition 3.3])**.** *Under the sampling with replacement model, the maximum number of repetitions of any entry in* Ω *is less than* $8\log(N)$ *with probability at least* $1 - N^{-2}$ *provided* $N \geq 9$.

Lemma 5 ([2, Lemma 3])**.** *Let* $\boldsymbol{U} \in \mathbb{C}^{(p_1p_2p_3)\times r}$ *and* $\boldsymbol{V} \in \mathbb{C}^{(q_1q_2q_3)\times r}$ *be two orthogonal matrices which satisfy*

$$\|\mathcal{P}_{\boldsymbol{U}}(\boldsymbol{H}_a)\|_F^2 \leq \frac{\mu c_s r}{N} \quad \text{and} \quad \|\mathcal{P}_{\boldsymbol{V}}(\boldsymbol{H}_a)\|_F^2 \leq \frac{\mu c_s r}{N}.$$

Then

$$\|\mathcal{P}_\mathcal{S}\mathcal{G}\mathcal{G}^*\mathcal{P}_\mathcal{S} - p^{-1}\mathcal{P}_\mathcal{S}\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_\mathcal{S}\| \leq \sqrt{\frac{32\mu c_s r\log(N)}{m}} \tag{6}$$

holds with probability at least $1 - N^{-2}$ *provided that*

$$m \geq 32\mu c_s r \log(N).$$

Lemma 6 ([6, Lemma 4.1]). *Let $\boldsymbol{L}_l = \boldsymbol{U}_l \boldsymbol{\Sigma}_l \boldsymbol{V}_l^*$ be another rank r matrix and $\mathcal{S}_l$ be the tangent space of the rank r matrix manifold at $\boldsymbol{L}_l$. Then*

$$\|(\mathcal{I} - \mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F \le \frac{\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F^2}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}, \quad \|\mathcal{P}_{\mathcal{S}_l} - \mathcal{P}_{\mathcal{S}}\| \le \frac{2\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}.$$

Lemma 7 ([5, Theorem 1.6]). *Consider a finite sequence $\{\boldsymbol{Z}_k\}$ of independent, random matrices with dimensions $d_1 \times d_2$. Assume that each random matrix satisfies*

$$\mathbb{E}\left(\boldsymbol{Z}_k\right) = 0 \quad \textit{and} \quad \|\boldsymbol{Z}_k\| \le R \quad \textit{almost surely.}$$

Define

$$\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^*\right)\right\|, \left\|\sum_k \mathbb{E}\left(\boldsymbol{Z}_k^*\boldsymbol{Z}_k\right)\right\|\right\}.$$

Then for all $t \ge 0$,

$$\mathbb{P}\left\{\left\|\sum_k \boldsymbol{Z}_k\right\| \ge t\right\} \le (d_1 + d_2)\exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

2.1 Local Convergence

We begin with a deterministic convergence result which characterizes the "basin of attraction" for FIHT. If the initial guess is located in this attraction region, FIHT will converge linearly to the underlying true solution.

Theorem 3. *Assume $0 < \varepsilon_0 < \frac{1}{10}$ and the following conditions*

$$\|\mathcal{P}_\Omega\| \le 8\log(N), \tag{7}$$
$$\|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}} - p^{-1}\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\| \le \varepsilon_0, \tag{8}$$
$$\frac{\|\boldsymbol{L}_0 - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} \le \frac{p^{1/2}\varepsilon_0}{16\log(N)(1+\varepsilon_0)} \tag{9}$$

are satisfied. Then the iterate $\boldsymbol{y}_l$ in (4) *satisfies $\|\boldsymbol{y}_l - \boldsymbol{y}\| \le \nu^l\|\boldsymbol{L}_0 - \mathcal{G}\boldsymbol{y}\|_F$ with $\nu = 10\varepsilon_0 < 1$.*

The proof of Thm. 3 makes use of the restricted isometry property of $\mathcal{P}_\Omega(\cdot)$ on $\mathcal{S}_l$ when $\boldsymbol{L}_l$ is in a small neighborhood of $\mathcal{G}\boldsymbol{y}$.

Lemma 8. *Suppose* (7), (8) *hold and*

$$\frac{\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} \le \frac{p^{1/2}\varepsilon_0}{16\log(N)(1+\varepsilon_0)}. \tag{10}$$

Then we have

$$\|\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| \le 8\log(N)(1+\varepsilon_0)p^{1/2} \tag{11}$$

and

$$\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l} - p^{-1}\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| \le 4\varepsilon_0. \tag{12}$$

Proof. Since $\|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\| = \|(\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega})^*\| = \|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\|$, for any $\boldsymbol{Z} \in \mathbb{C}^{(p_1p_2p_3)\times(q_1q_2q_3)}$,

$$\begin{aligned}
\|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z})\|^2 &= \langle \mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z}), \mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z})\rangle \\
&\le 8\log(N)\langle \mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z}), \mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z})\rangle \\
&= 8\log(N)\langle \boldsymbol{Z}, \mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z})\rangle \\
&\le 8\log(N)(1+\varepsilon_0)p\|\boldsymbol{Z}\|_F^2
\end{aligned}$$

where the first inequality follows from (7) and the second inequality follows from (8). So it follows that $\|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\| = \|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\| \le \sqrt{8\log(N)(1+\varepsilon_0)p}$ and

$$\begin{aligned}
\|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| &\le \|\mathcal{P}_{\Omega}\mathcal{G}^*(\mathcal{P}_{\mathcal{S}_l} - \mathcal{P}_{\mathcal{S}})\| + \|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\| \\
&\le 8\log(N)\frac{2\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} + \|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\| \\
&\le 8\log(N)\frac{p^{1/2}\varepsilon_0}{8\log(N)(1+\varepsilon_0)} + \sqrt{8\log(N)(1+\varepsilon_0)p} \\
&\le 8\log(N)(1+\varepsilon_0)p^{1/2},
\end{aligned}$$

where the second inequality follows from (7) and Lem. 6, the third inequality follows from (10).

Finally,

$$\begin{aligned}
&\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l} - p^{-1}\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| \\
&\le \|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}} - p^{-1}\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}}\| + \|(\mathcal{P}_{\mathcal{S}} - \mathcal{P}_{\mathcal{S}_l})\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| + \|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{G}^*(\mathcal{P}_{\mathcal{S}} - \mathcal{P}_{\mathcal{S}_l})\| \\
&\quad + \|p^{-1}(\mathcal{P}_{\mathcal{S}} - \mathcal{P}_{\mathcal{S}_l})\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| + \|p^{-1}\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*(\mathcal{P}_{\mathcal{S}} - \mathcal{P}_{\mathcal{S}_l})\| \\
&\le \varepsilon_0 + \frac{4\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} + p^{-1}\cdot\frac{2\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}\cdot(\|\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\| + \|\mathcal{P}_{\mathcal{S}}\mathcal{G}\mathcal{P}_{\Omega}\|) \\
&\le 4\varepsilon_0,
\end{aligned}$$

which completes the proof of (12). □

Proof of Theorem 3. First note that $\boldsymbol{L}_{l+1} = \mathcal{T}_r(\boldsymbol{W}_l)$, where

$$\begin{aligned}
\boldsymbol{W}_l &= \mathcal{P}_{\mathcal{S}_l}\mathcal{H}(\boldsymbol{x}_l + p^{-1}\mathcal{P}_{\Omega}(\boldsymbol{x} - \boldsymbol{x}_l)) \\
&= \mathcal{P}_{\mathcal{S}_l}\mathcal{G}(\boldsymbol{y}_l + p^{-1}\mathcal{P}_{\Omega}(\boldsymbol{y} - \boldsymbol{y}_l)).
\end{aligned}$$

So we have

$$\begin{aligned}
\|\boldsymbol{L}_{l+1} - \mathcal{G}\boldsymbol{y}\|_F &\le \|\boldsymbol{W}_l - \boldsymbol{L}_{l+1}\|_F + \|\boldsymbol{W}_l - \mathcal{G}\boldsymbol{y}\|_F \le 2\|\boldsymbol{W}_l - \mathcal{G}\boldsymbol{y}\|_F \\
&= 2\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}(\boldsymbol{y}_l + p^{-1}\mathcal{P}_{\Omega}(\boldsymbol{y} - \boldsymbol{y}_l)) - \mathcal{G}\boldsymbol{y}\|_F \\
&\le 2\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\boldsymbol{y} - \mathcal{G}\boldsymbol{y}\|_F + 2\|(\mathcal{P}_{\mathcal{S}_l}\mathcal{G} - p^{-1}\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_{\Omega})(\boldsymbol{y}_l - \boldsymbol{y})\|_F \\
&= 2\|(\mathcal{I} - \mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F + 2\|(\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{G}^* - p^{-1}\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*)(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F \\
&\le 2\|(\mathcal{I} - \mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F + 2\|(\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l} - p^{-1}\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F \\
&\quad + 2\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{G}^*(\mathcal{I} - \mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F + 2p^{-1}\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_{\Omega}\mathcal{G}^*(\mathcal{I} - \mathcal{P}_{\mathcal{S}_l})(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\|_F, \\
&:= I_1 + I_2 + I_3 + I_4,
\end{aligned}$$

where the second inequality comes from the fact that $\boldsymbol{L}_{l+1}$ is the best rank r approximation to $\boldsymbol{W}_l$, the second equality follows from $(\mathcal{I}-\mathcal{P}_{\mathcal{S}_l})\boldsymbol{L}_l = 0$, $\boldsymbol{y}_l = \mathcal{G}^*\boldsymbol{L}_l$ and $\mathcal{G}^*\mathcal{G} = \mathcal{I}$.

Let us first assume (10) holds. Then the application of Lem. 6 gives

$$\begin{aligned}
I_1 + I_3 + I_4 &\leq \left(\frac{4\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} + 2p^{-1}\|\mathcal{P}_\Omega \mathcal{G}^* \mathcal{P}_{\mathcal{S}_l}\| \frac{\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})} \right) \|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F \\
&\leq 2\varepsilon_0 \|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F,
\end{aligned}$$

where the last inequality follows from (8), (11) and the fact $\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\mathcal{P}_\Omega\| = \|\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_{\mathcal{S}_l}\|$. Moreover, (12) implies

$$I_2 \leq 8\varepsilon_0 \|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F.$$

Therefore putting the bounds for $I_1,\ I_2,\ I_3,$ and I_4 together gives

$$\|\boldsymbol{L}_{l+1} - \mathcal{G}\boldsymbol{y}\|_F \leq \nu \|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F,$$

where $\nu = 10\varepsilon_0 < 1$. Since (10) holds for $l = 0$ by the assumption of Thm. 3 and $\|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F$ is a contractive sequence, (10) holds for all $l \geq 0$. Thus

$$\|\boldsymbol{y}_l - \boldsymbol{y}\| = \|\mathcal{G}^*(\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y})\| \leq \|\boldsymbol{L}_l - \mathcal{G}\boldsymbol{y}\|_F \leq \nu^l \|\boldsymbol{L}_0 - \mathcal{G}\boldsymbol{y}\|_F,$$

where we have utilized the facts $\boldsymbol{y}_l = \mathcal{G}^*\boldsymbol{L}_l$, $\mathcal{G}^*\mathcal{G} = \mathcal{I}$ and $\|\mathcal{G}^*\| \leq 1$. □

2.2 Proofs of Lemma 2 and Theorem 1

Proof of Lemma 2. Recall that $\boldsymbol{L}_0 = \mathcal{T}_r(p^{-1}\mathcal{H}\mathcal{P}_\Omega(\boldsymbol{x})) = \mathcal{T}_r(p^{-1}\mathcal{G}\mathcal{P}_\Omega(\boldsymbol{y}))$ and $\mathcal{H}\boldsymbol{x} = \mathcal{G}\boldsymbol{y}$. Let us first bound $\left\|p^{-1}\mathcal{G}\mathcal{P}_\Omega(\boldsymbol{y}) - \mathcal{G}\boldsymbol{y}\right\|$. Since $p = \frac{m}{N}$, we have

$$p^{-1}\mathcal{G}\mathcal{P}_\Omega(\boldsymbol{y}) - \mathcal{G}\boldsymbol{y} = \sum_{k=1}^{m} \left(\frac{N}{m} y_{a_k} \boldsymbol{H}_{a_k} - \frac{1}{m}\mathcal{G}\boldsymbol{y} \right) := \sum_{k=1}^{m} \boldsymbol{Z}_{a_k}.$$

Because each a_k is drawn uniformly from $\{0, \cdots, N-1\}$, it is trivial that $\mathbb{E}\left(\boldsymbol{Z}_{a_k}\right) = 0$. Moreover, we have

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{Z}_{a_k}\boldsymbol{Z}_{a_k}^*\right) &= \mathbb{E}\left(\frac{N^2}{m^2} |y_{a_k}|^2 \boldsymbol{H}_{a_k}\boldsymbol{H}_{a_k}^* \right) - \frac{1}{m^2}(\mathcal{G}\boldsymbol{y})(\mathcal{G}\boldsymbol{y})^* \\
&= \frac{N}{m^2} \sum_{a=0}^{N-1} |y_a|^2 \boldsymbol{H}_a \boldsymbol{H}_a^* - \frac{1}{m^2}(\mathcal{G}\boldsymbol{y})(\mathcal{G}\boldsymbol{y})^* \\
&= \frac{N}{m^2}\boldsymbol{C} - \frac{1}{m^2}(\mathcal{G}\boldsymbol{y})(\mathcal{G}\boldsymbol{y})^*,
\end{aligned}$$

where $\boldsymbol{C}$ is a diagonal matrix which corresponds to the diagonal part of $(\mathcal{G}\boldsymbol{y})(\mathcal{G}\boldsymbol{y})^*$. Therefore

$$\left\| \mathbb{E}\left(\sum_{k=1}^{m} \boldsymbol{Z}_{a_k}\boldsymbol{Z}_{a_k}^* \right) \right\| \leq \frac{N}{m}\|\boldsymbol{C}\| \leq \frac{N}{m}\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty}^2,$$

where $\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty}$ denotes the maximum row ℓ_2 norm of $\mathcal{G}\boldsymbol{y}$. Similarly we can get

$$\left\|\mathbb{E}\left(\sum_{k=1}^{m}\boldsymbol{Z}_{a_k}^{*}\boldsymbol{Z}_{a_k}\right)\right\| \le \frac{N}{m}\left\|(\mathcal{G}\boldsymbol{y})^{*}\right\|_{2\to\infty}^{2}.$$

The definition of $\boldsymbol{H}_a$ implies $\|\boldsymbol{H}_a\| \le \frac{1}{\sqrt{w_a}}$. So

$$\|\boldsymbol{Z}_{a_k}\| \le \frac{N}{m}|y_{a_k}|\,\|\boldsymbol{H}_{a_k}\| + \frac{1}{m}\sum_{a=0}^{N-1}|y_a|\,\|\boldsymbol{H}_a\| \le \frac{2N}{m}\left\|\mathcal{D}^{-1}\boldsymbol{y}\right\|_{\infty}.$$

By matrix Bernstein inequality in Lem. 7, one can show that there exists a universal constant $C>0$ such that

$$\left\|\sum_{k=1}^{m}\boldsymbol{Z}_{a_k}\right\| \le C\left(\sqrt{\frac{N\log(N)}{m}}\max\left\{\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty},\|(\mathcal{G}\boldsymbol{y})^{*}\|_{2\to\infty}\right\} + \frac{N\log(N)}{m}\left\|\mathcal{D}^{-1}\boldsymbol{y}\right\|_{\infty}\right)$$

with probability at least $1-N^{-2}$. Consequently on the same event we have

$$\begin{aligned}\|\boldsymbol{L}_0-\mathcal{G}\boldsymbol{y}\| &\le \left\|\boldsymbol{L}_0-p^{-1}\mathcal{G}\mathcal{P}_{\Omega}(\boldsymbol{y})\right\| + \left\|p^{-1}\mathcal{G}\mathcal{P}_{\Omega}(\boldsymbol{y})-\mathcal{G}\boldsymbol{y}\right\| \le 2\left\|p^{-1}\mathcal{G}\mathcal{P}_{\Omega}(\boldsymbol{y})-\mathcal{G}\boldsymbol{y}\right\| \\ &\le C\left(\sqrt{\frac{N\log(N)}{m}}\max\left\{\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty},\|(\mathcal{G}\boldsymbol{y})^{*}\|_{2\to\infty}\right\} + \frac{N\log(N)}{m}\left\|\mathcal{D}^{-1}\boldsymbol{y}\right\|_{\infty}\right). \qquad (13)\end{aligned}$$

Thus it only remains to bound $\max\left\{\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty},\|(\mathcal{G}\boldsymbol{y})^{*}\|_{2\to\infty}\right\}$ and $\left\|\mathcal{D}^{-1}\boldsymbol{y}\right\|_{\infty}$ in terms of $\|\mathcal{G}\boldsymbol{y}\|$. From $\mathcal{G}\boldsymbol{y}=\mathcal{H}\boldsymbol{x}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{*}=\boldsymbol{E}_L\boldsymbol{D}\boldsymbol{E}_R^{T}$, we get

$$\begin{aligned}\|\mathcal{G}\boldsymbol{y}\|_{2\to\infty}^{2} &= \max_i\|\boldsymbol{e}_i^{*}(\mathcal{G}\boldsymbol{y})\|^2 = \max_i\|\boldsymbol{e}_i^{*}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{*}\|^2 \le \max_i\|\boldsymbol{e}_i^{*}\boldsymbol{U}\|^2\|\boldsymbol{\Sigma}\|^2 \\ &= \max_i\left\|\boldsymbol{U}^{(i,:)}\right\|^2\|\mathcal{G}\boldsymbol{y}\|_2^2 \le \frac{\mu_0 c_s r}{N}\|\mathcal{G}\boldsymbol{y}\|_2^2, \qquad (14)\end{aligned}$$

where the last inequality follows from Lem. 1. Similarly we also have

$$\|(\mathcal{G}\boldsymbol{y})^{*}\|_{2\to\infty}^{2} \le \frac{\mu_0 c_s r}{N}\|\mathcal{G}\boldsymbol{y}\|_2^2. \qquad (15)$$

The infinity norm of $\mathcal{D}^{-1}\boldsymbol{y}$ can be bounded as follows

$$\begin{aligned}\left\|\mathcal{D}^{-1}\boldsymbol{y}\right\|_{\infty} &= \|\mathcal{G}\boldsymbol{y}\|_{\infty} = \max_{i,j}|\boldsymbol{e}_i^{*}(\mathcal{G}\boldsymbol{y})\boldsymbol{e}_j| \le \max_{i,j}\|\boldsymbol{e}_i^{*}\boldsymbol{E}_L\|\,\|\boldsymbol{D}\|\,\left\|\boldsymbol{E}_R^{T}\boldsymbol{e}_j\right\| \\ &\le r\|\boldsymbol{D}\| \le r\left\|\boldsymbol{E}_L^{\dagger}\right\|\|\mathcal{G}\boldsymbol{y}\|\left\|(\boldsymbol{E}_R^{T})^{\dagger}\right\| \le \frac{\mu_0 c_s r}{N}\|\mathcal{G}\boldsymbol{y}\|, \qquad (16)\end{aligned}$$

where the last inequality follows from the μ_0-incoherence of $\mathcal{G}\boldsymbol{y}$.

Finally inserting (14), (15) and (16) into (13) gives

$$\|\boldsymbol{L}_0-\mathcal{G}\boldsymbol{y}\| \le C\sqrt{\frac{\mu_0 c_s r\log(N)}{m}}\|\mathcal{G}\boldsymbol{y}\|$$

provided $m \ge \mu_0 c_s r\log(N)$. □

Proof of Theorem 1. Following from (5), we only need to verify when the three conditions in Thm. 3 are satisfied. Lemma 4 implies (7) holds with probability at least $1 - N^{-2}$. Lemmas 1 and 5 guarantees (8) is true with probability at least $1 - N^{-2}$ if $m \geq C\varepsilon_0^{-2}\mu_0 c_s r \log(N)$ for a sufficiently large numerical constant $C > 0$. Similarly (9) can be satisfied with probability at least $1 - N^{-2}$ if $m \geq C(1+\varepsilon_0)\varepsilon_0^{-1}\mu_0^{1/2}c_s^{1/2}\kappa r N^{1/2}\log^{3/2}(N)$ following Lem. 2 and the fact $\|\boldsymbol{L}_0 - \mathcal{G}\boldsymbol{y}\|_F \leq \sqrt{2r}\,\|\boldsymbol{L}_0 - \mathcal{G}\boldsymbol{y}\|$, where κ denotes the condition number of $\mathcal{G}\boldsymbol{y}$. Taking an upper bound on the number of measurements completes the proof of Thm. 1. □

2.3 Proofs of Lemma 3 and Theorem 2

The proof of Lem. 3 relies on the following estimation of $\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\left(\widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}} - \mathcal{I}\right)\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\right\|$, which is a generalization of the asymmetric restricted isometry property [6] from matrix completion to low rank Hankel matrix completion.

Lemma 9. *Assume there exists a numerical constant μ such that*

$$\|\mathcal{P}_{\widehat{\boldsymbol{U}}_l}\boldsymbol{H}_a\|_F^2 \leq \frac{\mu c_s r}{N}, \quad \|\mathcal{P}_{\widehat{\boldsymbol{V}}_l}\boldsymbol{H}_a\|_F^2 \leq \frac{\mu c_s r}{N}, \tag{17}$$

and

$$\|\mathcal{P}_{\boldsymbol{U}}\boldsymbol{H}_a\|_F^2 \leq \frac{\mu c_s r}{N}, \quad \|\mathcal{P}_{\boldsymbol{V}}\boldsymbol{H}_a\|_F^2 \leq \frac{\mu c_s r}{N}. \tag{18}$$

for all $0 \leq a \leq N-1$. Let $\widehat{\Omega}_{l+1} = \{a_k \mid k = 1, \cdots, \widehat{m}\}$ be a set of indices sampled with replacement. If $\mathcal{P}_{\widehat{\Omega}_{l+1}}$ is independent of $\boldsymbol{U}$, $\boldsymbol{V}$, $\widehat{\boldsymbol{U}}_l$ and $\widehat{\boldsymbol{V}}_l$, then

$$\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\left(\mathcal{I} - \widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}}\right)\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\right\| \leq \sqrt{\frac{160\mu c_s r\log(N)}{\widehat{m}}}$$

with probability at least $1 - N^{-2}$ provided

$$\widehat{m} \geq \frac{125}{18}\mu c_s r\log(N).$$

Proof. Since for any $\boldsymbol{Z} \in \mathbb{C}^{(p_1p_2p_3)\times(q_1q_2q_3)}$

$$\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)(\boldsymbol{Z}) = \sum_{k=1}^{\widehat{m}}\left\langle \boldsymbol{Z}, \left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)(\boldsymbol{H}_{a_k})\right\rangle \mathcal{P}_{\mathcal{S}_l}(\boldsymbol{H}_{a_k}),$$

we can rewrite $\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)$ as

$$\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right) = \sum_{k=1}^{\widehat{m}}\mathcal{P}_{\mathcal{S}_l}(\boldsymbol{H}_{a_k}) \otimes \left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)(\boldsymbol{H}_{a_k}).$$

Define the random operator

$$\mathcal{R}_{a_k} = \mathcal{P}_{\widehat{\mathcal{S}}_l}(\boldsymbol{H}_{a_k}) \otimes \left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)(\boldsymbol{H}_{a_k}) - \frac{1}{N}\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right).$$

Then it is easy to see that $\mathbb{E}\left(\mathcal{R}_{a_k}\right) = 0$. By assumption, for any $0 \le a \le N-1$,

$$\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_a\right)\|_F^2 \le \|\mathcal{P}_{\widehat{U}_l}\left(\boldsymbol{H}_a\right)\|_F^2 + \|\mathcal{P}_{\widehat{V}_l}\left(\boldsymbol{H}_a\right)\|_F^2 \le \frac{2\mu c_s r}{N}.$$

So

$$\|\mathcal{R}_{a_k}\| \le \left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right)\right\|_F \left\|\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\left(\boldsymbol{H}_{a_k}\right)\right\|_F + \frac{1}{N}\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\right\| \le \frac{5\mu c_s r}{N}.$$

Next let us bound $\left\|\mathbb{E}(\mathcal{R}_{a_k}\mathcal{R}_{a_k}^*)\right\|$ as follows

$$\begin{aligned}
\left\|\mathbb{E}(\mathcal{R}_{a_k}\mathcal{R}_{a_k}^*)\right\| &= \left\|\mathbb{E}\left(\left\|\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\left(\boldsymbol{H}_{a_k}\right)\right\|_F^2 \mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right) \otimes \mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right)\right) - \frac{1}{N^2}\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)^2\mathcal{G}\mathcal{G}^*\mathcal{P}_{\widehat{\mathcal{S}}_l}\right\| \\
&\le \left\|\mathbb{E}\left(\left\|\left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\left(\boldsymbol{H}_{a_k}\right)\right\|_F^2 \mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right) \otimes \mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right)\right)\right\| + \frac{4}{N^2} \\
&\le \frac{4\mu c_s r}{N}\left\|\mathbb{E}\left(\mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right) \otimes \mathcal{P}_{\widehat{\mathcal{S}}_l}\left(\boldsymbol{H}_{a_k}\right)\right)\right\| + \frac{4}{N^2} \\
&= \frac{4\mu c_s r}{N^2}\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\widehat{\mathcal{S}}_l}\right\| + \frac{4}{N^2} \\
&\le \frac{8\mu c_s r}{N^2}.
\end{aligned}$$

This implies

$$\left\|\mathbb{E}\left(\sum_{k=1}^{\widehat{m}}\mathcal{R}_{a_k}\mathcal{R}_{a_k}^*\right)\right\| \le \sum_{k=1}^{\widehat{m}}\left\|\mathbb{E}(\mathcal{R}_{a_k}\mathcal{R}_{a_k}^*)\right\| \le \frac{8\mu c_s r\widehat{m}}{N^2}.$$

We can similarly obtain

$$\left\|\mathbb{E}\left(\sum_{k=1}^{\widehat{m}}\mathcal{R}_{a_k}^*\mathcal{R}_{a_k}\right)\right\| \le \frac{12\mu c_s r\widehat{m}}{N^2}.$$

So the application of the matrix Bernstein inequality in Lem. 7 gives

$$\mathbb{P}\left\{\left\|\sum_{k=1}^{\widehat{m}}\mathcal{R}_{a_k}\right\| \ge t\right\} \le 2(p_1p_2p_3)(q_1q_2q_3)\cdot\exp\left(\frac{-t^2/2}{\frac{12\mu c_s\widehat{m}r}{N^2} + \frac{5\mu c_s r}{N}t/3}\right).$$

If $t \le \frac{24\widehat{m}}{5N}$, then

$$\mathbb{P}\left\{\left\|\sum_{k=1}^{\widehat{m}}\mathcal{R}_{a_k}\right\| \ge t\right\} \le 2(p_1p_2p_3)(q_1q_2q_3)\cdot\exp\left(\frac{-t^2/2}{\frac{20\mu c_s\widehat{m}r}{N^2}}\right) \le N^2\exp\left(\frac{-t^2/2}{\frac{20\mu c_s\widehat{m}r}{N^2}}\right).$$

Setting $t = \sqrt{\frac{160\mu c_s\widehat{m}r\log(N)}{N^2}}$ gives

$$\mathbb{P}\left\{\left\|\sum_{k=1}^{\widehat{m}}\mathcal{R}_{a_k}\right\| \ge t\right\} \le N^{-2}.$$

The condition $t \leq \frac{24\widehat{m}}{5N}$ implies $\widehat{m} \geq \frac{125}{18}\mu c_s r \log(N)$. The proof is complete because

$$\frac{N}{\widehat{m}} \sum_{k=1}^{\widehat{m}} \mathcal{R}_{a_k} = \mathcal{P}_{\widehat{\mathcal{S}}_l} \mathcal{G} \left(\widehat{p}^{-1} \mathcal{P}_{\widehat{\Omega}_{l+1}} - \mathcal{I} \right) \mathcal{G}^* \left(\mathcal{P}_{\boldsymbol{U}} - \mathcal{P}_{\widehat{\boldsymbol{U}}_l} \right).$$

□

The following lemma from [6] will also be used in the proof of Lem. 3.

Lemma 10. *Let $\widetilde{\boldsymbol{L}}_l = \widetilde{\boldsymbol{U}}_l \widetilde{\boldsymbol{\Sigma}}_l \widetilde{\boldsymbol{V}}_l^*$ and $\mathcal{G}\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ be two rank r matrices which satisfy*

$$\left\| \widetilde{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \leq \frac{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}{10\sqrt{2}}.$$

Assume $\left\|\boldsymbol{U}^{(i,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N}$ and $\left\|\boldsymbol{V}^{(j,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N}$. Then the matrix $\widehat{\boldsymbol{L}}_l = \mathrm{Trim}_{\mu_0}(\widetilde{\boldsymbol{L}}_l) = \widehat{\boldsymbol{U}}_l \widehat{\boldsymbol{\Sigma}}_l \widehat{\boldsymbol{V}}_l^*$ returned by Alg. 4 satisfies

$$\left\| \widehat{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \leq 8\kappa \left\| \widetilde{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \quad \textit{and} \quad \max\left\{ \left\| \widehat{\boldsymbol{U}}^{(i,:)} \right\|^2, \left\| \widehat{\boldsymbol{V}}^{(j,:)} \right\|^2 \right\} \leq \frac{100\mu_0 c_s r}{81N},$$

where κ denotes the condition number of $\mathcal{G}\boldsymbol{y}$.

Proof of Lemma 3. Let us first assume that

$$\left\| \widetilde{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \leq \frac{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}{256\kappa^2}. \tag{19}$$

Then the application of Lem. 10 implies that

$$\left\| \widehat{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \leq 8\kappa \left\| \widetilde{\boldsymbol{L}}_l - \mathcal{G}\boldsymbol{y} \right\|_F \quad \text{and} \quad \max\left\{ \left\| \widehat{\boldsymbol{U}}^{(i,:)} \right\|^2, \left\| \widehat{\boldsymbol{V}}^{(j,:)} \right\|^2 \right\} \leq \frac{100\mu_0 c_s r}{81N} \tag{20}$$

by noting that $\left\|\boldsymbol{U}^{(i,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N}$ and $\left\|\boldsymbol{V}^{(j,:)}\right\|^2 \leq \frac{\mu_0 c_s r}{N}$ following from Lem. 1. Moreover, direct calculation gives

$$\left\| \mathcal{P}_{\widehat{\boldsymbol{U}}_l} \boldsymbol{H}_a \right\|_F^2 = \left\| \widehat{\boldsymbol{U}}_l^* \boldsymbol{H}_a \right\|_F^2 = \frac{1}{|\Gamma_a|} \sum_{i \in \Gamma_a} \left\| \left(\widehat{\boldsymbol{U}}_l \right)^{(i,:)} \right\|_2^2 \leq \frac{100\mu_0 c_s r}{81N}, \tag{21}$$

where Γ_a is the set of row indices for non-zero entries in $\boldsymbol{H}_a$ with cardinality $|\Gamma_a| = w_a$. Similarly,

$$\left\| \mathcal{P}_{\widehat{\boldsymbol{V}}_l} \boldsymbol{H}_a \right\|_F^2 \leq \frac{100\mu_0 c_s r}{81N}. \tag{22}$$

Recall that $\boldsymbol{y} = \mathcal{D}\boldsymbol{x}$ and $\mathcal{G}\boldsymbol{y} = \mathcal{H}\boldsymbol{x}$. Define $\widehat{\boldsymbol{y}}_l = \mathcal{D}\widehat{\boldsymbol{x}}_l$. Then $\widehat{\boldsymbol{y}}_l = \mathcal{G}^* \widehat{\boldsymbol{L}}_l$ and

$$\mathcal{P}_{\widehat{\mathcal{S}}_l} \mathcal{H} \left(\widehat{\boldsymbol{x}}_l + \widehat{p}^{-1} \mathcal{P}_{\Omega_{l+1}} \left(\boldsymbol{x} - \widehat{\boldsymbol{x}}_l \right) \right) = \mathcal{P}_{\mathcal{S}_l} \mathcal{G} \left(\widehat{\boldsymbol{y}}_l + \widehat{p}^{-1} \mathcal{P}_{\Omega_{l+1}} \left(\boldsymbol{y} - \widehat{\boldsymbol{y}}_l \right) \right).$$

Consequently,

$$
\begin{aligned}
\|\widetilde{\boldsymbol{L}}_{l+1}-\mathcal{G}\boldsymbol{y}\|_F &\le 2\left\|\mathcal{P}_{\mathcal{S}_l}\mathcal{G}\left(\widehat{\boldsymbol{y}}_l+\widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}}(\boldsymbol{y}-\widehat{\boldsymbol{y}}_l)\right)-\mathcal{G}\boldsymbol{y}\right\|_F\\
&\le 2\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\boldsymbol{y}-\mathcal{G}\boldsymbol{y}\right\|_F+2\left\|\left(\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}-\widehat{p}^{-1}\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\right)(\widehat{\boldsymbol{y}}_l-\boldsymbol{y})\right\|_F\\
&= 2\left\|\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\mathcal{G}\boldsymbol{y}\right\|_F+2\left\|\left(\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*-\widehat{p}^{-1}\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\mathcal{G}^*\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)\right\|_F\\
&\le 2\left\|\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)\right\|_F+2\left\|\left(\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\widehat{\mathcal{S}}_l}-\widehat{p}^{-1}\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\mathcal{P}_{\widehat{\Omega}_{l+1}}\mathcal{G}^*\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)\right\|_F\\
&+2\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\left(\mathcal{I}-\widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}}\right)\mathcal{G}^*\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y})\right\|_F\\
&:= I_5+I_6+I_7.
\end{aligned}
$$

The first item I_5 can be bounded as

$$
I_5\le\frac{2\left\|\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F^2}{\sigma_{\min}(\mathcal{G}\boldsymbol{y})}\le\frac{1}{2}\left\|\widetilde{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F,
$$

which follows from Lem. 6, the left inequality of (20) and the assumption (19). The application of Lem. 5 together with (21) and (22) implies

$$
I_6\le 2\sqrt{\frac{3200\mu_0c_sr\log(N)}{81\widehat{m}}}\left\|\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F\le 16\kappa\sqrt{\frac{3200\mu_0c_sr\log(N)}{81\widehat{m}}}\left\|\widetilde{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F
$$

with probability at least $1-N^2$. To bound I_7, first note that

$$
\begin{aligned}
\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)&=\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)(-\mathcal{G}\boldsymbol{y})=\left(\boldsymbol{I}-\widehat{\boldsymbol{U}}_l\widehat{\boldsymbol{U}}_l^*\right)(-\mathcal{G}\boldsymbol{y})\left(\boldsymbol{I}-\widehat{\boldsymbol{V}}_l\widehat{\boldsymbol{V}}_l^*\right)\\
&=\left(\boldsymbol{U}\boldsymbol{U}^*-\widehat{\boldsymbol{U}}_l\widehat{\boldsymbol{U}}_l^*\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)\left(\boldsymbol{I}-\widehat{\boldsymbol{V}}_l\widehat{\boldsymbol{V}}_l^*\right)\\
&=\left(\mathcal{P}_{\boldsymbol{U}}-\mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\left(\mathcal{I}-\mathcal{P}_{\widehat{\boldsymbol{V}}_l}\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
I_7&=2\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\left(\mathcal{I}-\widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}}\right)\mathcal{G}^*\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\left(\mathcal{P}_{\boldsymbol{U}}-\mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\left(\mathcal{I}-\mathcal{P}_{\widehat{\boldsymbol{V}}_l}\right)\left(\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right)\right\|_F\\
&\le 2\left\|\mathcal{P}_{\widehat{\mathcal{S}}_l}\mathcal{G}\left(\mathcal{I}-\widehat{p}^{-1}\mathcal{P}_{\widehat{\Omega}_{l+1}}\right)\mathcal{G}^*\left(\mathcal{I}-\mathcal{P}_{\widehat{\mathcal{S}}_l}\right)\left(\mathcal{P}_{\boldsymbol{U}}-\mathcal{P}_{\widehat{\boldsymbol{U}}_l}\right)\right\|\left\|\widehat{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F\\
&\le 16\kappa\sqrt{\frac{16000\mu_0c_sr\log(N)}{81\widehat{m}}}\left\|\widetilde{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F
\end{aligned}
$$

with probability at least $1-N^2$, where the last inequality follows from Lem. 9 and the left inequality of (20). Putting the bounds for I_5, I_6 and I_7 together gives

$$
\|\widetilde{\boldsymbol{L}}_{l+1}-\mathcal{G}\boldsymbol{y}\|_F\le\left(\frac{1}{2}+326\kappa\sqrt{\frac{\mu_0c_sr\log(N)}{\widehat{m}}}\right)\left\|\widetilde{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F\le\frac{5}{6}\left\|\widetilde{\boldsymbol{L}}_l-\mathcal{G}\boldsymbol{y}\right\|_F
$$

with probability at least $1-2N^{-2}$ provided $\widehat{m}\ge C\mu_0c_s\kappa^2r\log(N)$ for a sufficiently large universal constant C. Clearly on the same event, (19) also holds for the $(l+1)$-th iteration.

Since $\widetilde{\boldsymbol{L}}_0 = \mathcal{T}_r\left(\widehat{p}^{-1}\mathcal{H}\mathcal{P}_{\Omega_0}(\boldsymbol{x})\right)$, (19) is valid for $l = 0$ with probability at least $1 - N^2$ provides

$$\widehat{m} \geq C\mu_0 c_s \kappa^6 r^2 \log(N)$$

for some numerical constant $C > 0$. Taking the upper bound on the number of measurements completes the proof of Lem. 3 by noting $\mathcal{H}\boldsymbol{x} = \mathcal{G}\boldsymbol{y}$. □

Proof of Theorem 2. The third condition (9) in Thm. 3 can be satisfied with probability at least $1 - (2L+1)N^{-2}$ if we take $L = \left\lceil 6\log\left(\frac{\sqrt{N}\log(N)}{16\varepsilon_0}\right)\right\rceil$. So the theorem can be proved by combining this result together with Lems. 4 and 5. □

References

[1] J. F. Cai, T. Wang, and K. Wei, *Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion*, arXiv preprint arXiv:1606.01567, (2016).

[2] Y. Chen and Y. Chi, *Robust spectral compressed sensing via structured matrix completion*, Information Theory, IEEE Transactions on, 60 (2014), pp. 6576–6601.

[3] W. Liao, *MUSIC for multidimensional spectral estimation: Stability and super-resolution*, Signal Processing, IEEE Transactions on, 63 (2015), pp. 6395–6406.

[4] B. Recht, *A simpler approach to matrix completion*, The Journal of Machine Learning Research, 12 (2011), pp. 3413–3430.

[5] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Foundations of computational mathematics, 12 (2012), pp. 389–434.

[6] K. Wei, J. F. Cai, T. F. Chan, and S. Leung, *Guarantees of Riemannian optimization for low rank matrix completion*, arXiv preprint arXiv:1603.06610, (2016).