

# STRUCTURED GRADIENT DESCENT FOR FAST ROBUST LOW-RANK HANKEL MATRIX COMPLETION\*

HANQIN CAI<sup>†</sup>, JIAN-FENG CAI<sup>‡</sup>, AND JUNTAO YOU<sup>§</sup>

**Abstract.** We study the robust matrix completion problem for the low-rank Hankel matrix, which detects the sparse corruptions caused by extreme outliers while we try to recover the original Hankel matrix from partial observation. In this paper, we explore the convenient Hankel structure and propose a novel nonconvex algorithm, coined Hankel structured gradient descent (HSGD), for large-scale robust Hankel matrix completion problems. HSGD is highly computing- and sample-efficient compared to the state of the art. The recovery guarantee with a linear convergence rate has been established for HSGD under some mild assumptions. The empirical advantages of HSGD are verified on both synthetic datasets and real-world nuclear magnetic resonance signals.

**Key words.** low-rank Hankel matrix, robust matrix completion, structured gradient descent, outliers detection, nuclear magnetic resonance

**MSC codes.** 15A29, 15A83, 41A29, 47B35, 90C17, 90C26

**DOI.** 10.1137/22M1491009

**1. Introduction.** Recently, the problems of the Hankel matrix have received much attention. A complex-valued Hankel matrix has identical values on every anti-diagonal:

$$(1) \quad \mathcal{H}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n_2} \\ x_2 & x_3 & \cdots & x_{n_2+1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_1} & x_{n_1+1} & \cdots & x_n \end{bmatrix} \in \mathbb{C}^{n_1 \times n_2},$$

where  $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top \in \mathbb{C}^n$  is a vector consisting of the distinct entries of the Hankel matrix, and the linear operator  $\mathcal{H} : \mathbb{C}^n \rightarrow \mathbb{C}^{n_1 \times n_2}$  is called Hankel mapping. Note that we always have  $n = n_1 + n_2 - 1$  regardless of the shape of the Hankel matrix. Hence, a Hankel matrix can always be efficiently mapped from a much smaller vector.

The low-rank Hankel matrix has arisen in various applications, for instance, nuclear magnetic resonance (NMR) spectroscopy [21, 30, 32], medical imaging [22, 23], seismic imaging [15, 31], and autoregression [25, 27]. In real-world applications, there are often two major challenges for data analysis:

1. *Missing data.* Due to hardware limitations or time constraints, only partial data can be obtained [35]. By the structure of the Hankel matrix, we naturally

\* Submitted to the journal's Methods and Algorithms for Scientific Computing section April 18, 2022; accepted for publication (in revised form) November 28, 2022; published electronically June 8, 2023.

<https://doi.org/10.1137/22M1491009>

**Funding:** The work of the first author is partially supported by NSF DMS 2304489. The work of the second author is partially supported by HKRGC GRF grants 16309518, 16309219, 16310620, and 16306821.

<sup>†</sup> Department of Statistics and Data Science and Department of Computer Science, University of Central Florida, Orlando, FL 32816 USA (hqcai@ucf.edu).

<sup>‡</sup> Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China (jfc@ust.hk).

<sup>§</sup> Corresponding author Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China (jyouab@connect.ust.hk).

know all the values on an antidiagonal once we observe one entry on it. In this sense, the partial observation of a Hankel matrix must also be Hankel structured.

2. *Extreme outliers.* Due to hardware malfunction, the recorded data are often sparsely corrupted by extreme outliers [38]. Note that the outliers are usually Hankel structured as well due to the nature of these applications. In fact, the outliers who corrupt only a few entries on an antidiagonal can be easily detected by comparing all values on that antidiagonal. Thus, we are only interested in the Hankel structured outliers that destroy some antidiagonals completely.

In the NMR spectroscopy example, observing the full signal is very time-consuming and can take up to months. Thus, the researchers often deal with the partially observed signal due to time constraints. On the other hand, impulse noise often causes sparse corruptions on the observed signal, which can be viewed as additive outliers.

In this paper, we aim to handle both these challenges simultaneously by solving the robust low-rank Hankel matrix completion (RHC) problem: Given the sparsely corrupted partial observations on the Hankel matrix, we want to recover the original low-rank Hankel matrix. Note that the observations are Hankel structured since observing one entry gives the information for an entire antidiagonal. In addition, we are only interested in the sparse corruptions with Hankel structured support since (i) unstructured outliers can be easily removed when we sample multiple entries on the same antidiagonal,<sup>1</sup> or (ii) corrupting very few entries is equivalent to corrupting the entire antidiagonal if only very few entries were sampled. Moreover, in many applications, e.g., NMR spectroscopy, only one entry (viz. one copy) may be observed for each antidiagonal. Thus, adding outliers to such observations is the same as adding Hankel structured outliers.

With all the participants in the Hankel structure, we can write the RHC problem in the following form: Consider

$$\mathcal{H}\mathbf{y} = \mathcal{H}\mathbf{x}^{\natural} + \mathcal{H}\mathbf{w}^{\natural},$$

where  $\mathcal{H}\mathbf{x}^{\natural} \in \mathbb{C}^{n_1 \times n_2}$  is the underlying low-rank Hankel matrix and  $\mathcal{H}\mathbf{w}^{\natural} \in \mathbb{C}^{n_1 \times n_2}$  is the Hankel structured sparse corruption matrix. The problem is to recover  $\mathcal{H}\mathbf{x}^{\natural}$  (or equivalently  $\mathbf{x}^{\natural} \in \mathbb{C}^n$ ) from the sparsely corrupted partial observation

$$(2) \quad \mathcal{H}\Pi_{\Omega}\mathbf{y} = \mathcal{H}\Pi_{\Omega}(\mathbf{x}^{\natural} + \mathbf{w}^{\natural}),$$

where  $\Omega \subseteq [n]$  is the index set of the observed entries in corresponding vector form.  $\Pi_{\Omega} : \mathbb{C}^n \rightarrow \mathbb{C}^n$  is the sampling operator defined by  $[\Pi_{\Omega}\mathbf{y}]_i = y_i$  if  $i \in \Omega$ , or otherwise  $[\Pi_{\Omega}\mathbf{y}]_i = 0$ . Naturally, we consider the loss function:

$$(3) \quad \frac{1}{2p} \|\mathcal{H}\Pi_{\Omega}\mathbf{y} - \mathcal{H}\Pi_{\Omega}(\mathbf{x} + \mathbf{w})\|_F^2,$$

where  $p$  is the probability of an entry being observed. One can see that the RHC problem can be efficiently described in a more straightforward vector form  $\Pi_{\Omega}\mathbf{y} = \Pi_{\Omega}(\mathbf{x}^{\natural} + \mathbf{w}^{\natural})$ . However, the corresponding vector loss function  $\frac{1}{2}\|\Pi_{\Omega}\mathbf{y} - \Pi_{\Omega}(\mathbf{x} + \mathbf{w})\|_2^2$  is not equivalent to the original Hankel matrix loss function (3), because antidiagonals contain different numbers of repeated entries. Thus, we must adjust the weights of the entries in the vector loss function. To this end, we introduce the reweighting operator

<sup>1</sup>Applying a simple median filter will remove the outlier(s) as long as less than half of the observations on this antidiagonal were corrupted.

$\mathcal{D} : \mathbb{C}^n \rightarrow \mathbb{C}^n$  defined as  $[\mathcal{D}\mathbf{x}]_i = \sqrt{\varsigma_i}x_i$ , where  $\varsigma_i$  is the number of entries on the  $i$ th antidiagonal of an  $n_1 \times n_2$  matrix, so that

$$(4) \quad \frac{1}{2p} \|\Pi_\Omega \mathcal{D}\mathbf{y} - \Pi_\Omega \mathcal{D}(\mathbf{x} + \mathbf{w})\|_2^2$$

is equivalent to (3). For the interested reader, we extend further discussion on the necessity of the reweighting operator in the supplementary material (Supplementary\_Materials.pdf [local/web 606KB]). For ease of presentation, we introduce the reweighted notation:

$$(5) \quad \mathbf{f} := \mathcal{D}\mathbf{y}, \quad \mathbf{z} := \mathcal{D}\mathbf{x}, \quad \mathbf{s} := \mathcal{D}\mathbf{w}, \quad \mathcal{G} := \mathcal{H}\mathcal{D}^{-1}.$$

Note that  $\mathcal{G}\mathbf{f} = \mathcal{H}\mathbf{y}$ ,  $\mathcal{G}\mathbf{z} = \mathcal{H}\mathbf{x}$ , and  $\mathcal{G}\mathbf{s} = \mathcal{H}\mathbf{w}$ . Combined with the low-rank constraints on  $\mathbf{x}$  and the sparse constraints on  $\mathbf{w}$ , we have the vector formula that is equivalent to the original Hankel matrix formula:

$$(6) \quad \begin{aligned} & \underset{\mathbf{z}, \mathbf{s} \in \mathbb{C}^n}{\text{minimize}} \quad \frac{1}{2p} \|\Pi_\Omega \mathbf{f} - \Pi_\Omega(\mathbf{z} + \mathbf{s})\|_2^2 \\ & \text{subject to} \quad \text{rank}(\mathcal{G}\mathbf{z}) = r, \quad \Pi_\Omega \mathbf{s} \text{ is } \alpha p\text{-sparse,} \end{aligned}$$

where  $\alpha p$ -sparsity will be formally defined as Assumption 3 later. Note that  $\mathcal{D}^2 = \mathcal{H}^* \mathcal{H}$  and  $\mathcal{G}^* \mathcal{G} = \mathcal{I}$  where  $(\cdot)^*$  denotes the adjoint operator. Hence, RHC can be viewed as robust matrix completion on an orthonormal basis.

**1.1. Assumptions and notation.** In this subsection, we present the problem assumptions and notation that will be used for the rest of the paper. We start with the sampling model for the observations.

*Assumption 1* (Bernoulli sampling). The set of sampling index  $\Omega$  is drawn by the Bernoulli model with probability  $p$ . That is, the  $i$ th entry of  $\mathbf{y}$  is observed with probability  $p$  independent of all others.

Note that the entries on an antidiagonal are repeated. Thus, for sampling efficiency, we will sample no more than once on each antidiagonal for sampling efficiency. Hence, our sampling model is presented on the vector that consists of the distinct elements of the Hankel matrix. In practice, the actual probability  $p$  is usually unknown, and it is common to take  $p = |\Omega|/n$ . For the reader's interest, we highlight that a similar method and theorem can also be developed for the uniform sampling model with no extra difficulty.

*Assumption 2* ( $\mu$ -incoherence). The rank- $r$  Hankel matrix  $\mathcal{H}\mathbf{x}^\natural = \mathcal{G}\mathbf{z}^\natural \in \mathbb{C}^{n_1 \times n_2}$  is  $\mu$ -incoherent, i.e., there exists a constant  $\mu$  such that

$$\|\mathbf{U}^\natural\|_{2,\infty} \leq \sqrt{\mu c_s r / n} \quad \text{and} \quad \|\mathbf{V}^\natural\|_{2,\infty} \leq \sqrt{\mu c_s r / n},$$

where  $\mathbf{U}^\natural \boldsymbol{\Sigma}^\natural \mathbf{V}^{\natural*}$  is the compact SVD of  $\mathcal{H}\mathbf{x}^\natural$  and  $c_s = \max\{n/n_1, n/n_2\}$ .

The assumption of incoherence was first introduced in [13] and has been widely used in robust principal component analysis and matrix completion studies. The parameter  $\mu$  describes how evenly the information is distributed among the entries of the matrix, and its value is small in a well-conditioned dataset. Empirically, many applications related to low-rank matrices, such as face recognition [37] and video background subtraction [5], satisfy the incoherence condition with small  $\mu$ . In this paper, Assumption 2 is a Hankel variation of the standard incoherence. In the applications

of the Hankel matrix, the parameter  $\mu$  is often very small. For example, in the application of a spectrally sparse signal,  $\mu = \mathcal{O}(1)$  if the signal is well separated in the frequency domain [26, Theorem 2]. Note that the constant  $c_s$  describes how square the Hankel matrix is. Ideally, when  $n$  is fixed,  $n_1 \approx n_2 \approx n/2$  gives a relatively easier problem.

*Assumption 3* ( $\alpha p$ -sparsity). Under the Bernoulli model with probability  $p$ , i.e., Assumption 1, the vector of observed outliers is  $\alpha p$ -sparse. That is,

$$\|\Pi_\Omega \mathbf{s}^\natural\|_0 = \|\Pi_\Omega \mathbf{w}^\natural\|_0 \leq \alpha p n.$$

Following the same argument for Assumption 1, the observed outliers are also presented in the vector form. Note that Assumption 3 holds with high probability provided  $\mathbf{s}^\natural$  is  $(\alpha/2)$ -sparse [8, Proposition 4.5]. This assumption also implies that  $\mathcal{H}\Pi_\Omega \mathbf{s}^\natural$  has no more than  $\alpha p n$  nonzero entries in each row and column.

In the rest of this subsection, we describe the notation that we will use throughout the paper. We use regular lowercase letters for scalars, bold lowercase letters for vectors, bold capital letters for matrices, and calligraphic letters for operators. For a vector  $\mathbf{v}$ , let  $\|\mathbf{v}\|_0$ ,  $\|\mathbf{v}\|_2$ , and  $\|\mathbf{v}\|_\infty$  denote  $\ell_0$ -,  $\ell_2$ -, and  $\ell_\infty$ -norms, respectively. For a matrix  $\mathbf{M}$ ,  $\|\mathbf{M}\|_{2,\infty}$  denotes the largest  $\ell_2$ -norm of the rows,  $\|\mathbf{M}\|_\infty$  denotes the largest magnitude in the entries,  $\sigma_i(\mathbf{M})$  denotes the  $i$ th singular value,  $\|\mathbf{M}\|_2$  denotes its spectral norm, and  $\|\mathbf{M}\|_F$  denotes its Frobenius norm.  $\|\mathcal{P}\|$  denotes the operator norm of the linear map  $\mathcal{P}$ . For both vectors and matrices,  $\overline{(\cdot)}$ ,  $(\cdot)^\top$ ,  $(\cdot)^*$ , and  $\langle \cdot, \cdot \rangle$  denote conjugate, transpose, conjugate transpose, and inner product, respectively. Moreover,  $\sigma_i^\natural$  denotes the  $i$ th singular value of the underlying Hankel matrix  $\mathcal{H}\mathbf{x}^\natural$  and  $\kappa = \sigma_1^\natural/\sigma_r^\natural$  denotes its condition number. Later in Appendix A, we introduce some additional notation used only in the analysis.

**1.2. Related work and main contributions.** The decomposition problem for generic low-rank and sparse matrices is known as robust principal component analysis (full observation) or robust matrix completion (partial observation), which has been widely studied in both theoretical and empirical aspects [5, 6, 7, 8, 9, 12, 14, 17, 18, 20, 29, 39]. However, without utilizing the convenient Hankel structure, the generic matrix approaches are suboptimal on RHC problems, in terms of robustness, sample complexity, and computational efficiency.

Recently, dedicated methods have been proposed for low-rank Hankel matrix problems. Robust enhanced matrix completion (Robust-EMaC) [16] relaxes the nonconvex RHC problem with a convex formula where the nuclear norm and  $\ell_1$ -norm are used to enforce low-rank and sparsity constraints, respectively. While Robust-EMaC has the state-of-the-art sampling complexity, requiring merely  $\mathcal{O}(c_s^2 \mu^2 r^2 \log n)$  samples, it does not provide an efficient numerical solver.<sup>2</sup> Note that Robust-EMaC tolerates a small constant portion of outliers if the support of outliers is randomly distributed, which is more restrictive than Assumption 3. Other convex approaches [1, 15, 33] have similar computational challenges when the problem scale is large. Later, more provable nonconvex methods with linear convergence are introduced. [10, 11] propose two fast algorithms that both solve a Hankel matrix completion problem in  $\mathcal{O}(r^2 n + rn \log n)$  flops per iteration. Unfortunately, they are not designed to handle extreme outliers. Structured alternating projection (SAP) [40] is an alternating

<sup>2</sup>The original paper uses semidefinite programming to solve the convex model, which is as expensive as  $\mathcal{O}(n^6)$ . The first-order method can improve this to  $\mathcal{O}(n^3)$  flops per iteration, which is still too expensive.

projection-based algorithm that efficiently removes outliers from the Hankel matrix, but the theoretical guarantee is only established under full observation. The computational complexity of SAP is  $\mathcal{O}(r^2 n \log n)$  per iteration, which is later improved to  $\mathcal{O}(r^2 n + rn \log n)$  by its accelerated version, ASAP [4]; however, ASAP only focuses on the fully observed cases. In the follow-up work [41], SAP is modified for exploring the setting of partial observation, which we call PartialSAP. PartialSAP has the same computational complexity as SAP, requiring  $\mathcal{O}(c_s^2 \mu^2 r^3 \log^2 n \log \varepsilon^{-1})$  samples due to its iterative resampling requirement in theory. PartialSAP can tolerate  $\alpha \lesssim \mathcal{O}(1/(c_s \mu r))$  fraction outliers,<sup>3</sup> under the same Assumption 3.

In this work, we propose a novel nonconvex algorithm, coined Hankel structured gradient descent (HSGD), for solving large-scale RHC problems. Our main contributions are as follows:

1. HSGD is computing-efficient. Its computational complexity is  $\mathcal{O}(r^2 n + rn \log n)$  flops per iteration—better than the current state-of-the-art PartialSAP.
2. HSGD is sample-efficient. Without the requirement of iterative resampling, its overall sample complexity is  $\mathcal{O}(\max\{c_s^2 \mu^2 r^2 \log n, c_s \mu \kappa^3 r^2 \log n\})$ —tied with the state-of-the-art Robust-EMaC if the problem is well-conditioned.
3. The recovery guarantee has been established. Under some mild assumptions, we show HSGD converges linearly to the ground truth. In particular, the theoretical outlier tolerance is  $\alpha \lesssim \mathcal{O}(1/\max\{c_s \mu \kappa^{3/2} r^{3/2}, c_s \mu r \kappa^2\})$ .
4. The empirical advantages of HSGD are verified by both synthetic datasets and real-world NMR signals. We observe that HSGD outperforms Robust-EMaC and PartialSAP for both speed and recoverability.

**2. Proposed method.** In this section, we propose the HSGD algorithm, a highly efficient nonconvex algorithm for the RHC problem (6). The first major challenge in algorithm design is how to enforce the low-rank constraint on  $\mathcal{G}(\mathbf{z})$ . We reformulate the objective function so that the low-rank constraint can be avoided. Following [39], we rewrite the rank- $r$  matrix on factorized space:  $\mathcal{G}(\mathbf{z}) = \mathbf{L}\mathbf{R}^*$  where  $\mathbf{L} \in \mathbb{C}^{n_1 \times r}$  and  $\mathbf{R} \in \mathbb{C}^{n_2 \times r}$ . Thus, the low-rank constraint is automatically coded by the shapes of  $\mathbf{L}$  and  $\mathbf{R}$ . Moreover, we add a penalty term  $\frac{1}{2} \|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*)\|_{\text{F}}^2$  to make sure  $\mathbf{L}\mathbf{R}^*$  is a Hankel matrix, because  $\mathbf{L}\mathbf{R}^*$  is a Hankel matrix if and only if  $(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*) = \mathbf{0}$ . By replacing  $\mathbf{z} = \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$ , we have

$$(7) \quad \psi := \psi(\mathbf{L}, \mathbf{R}; \mathbf{s}) := \frac{1}{2p} \|\Pi_{\Omega}(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f})\|_2^2 + \frac{1}{2} \|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*)\|_{\text{F}}^2.$$

We also add another balance regularization,

$$(8) \quad \phi := \phi(\mathbf{L}, \mathbf{R}) := \frac{1}{4} \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_{\text{F}}^2,$$

to encourage that  $\mathbf{L}$  and  $\mathbf{R}$  have the same scale, which is a common technique for factorized gradient descent under the asymmetric setting. [28] suggests that the balance regularization  $\phi$  may be removed in the matrix sensing problem if we have a very good initialization. However, we decided to keep this term since the initialization is usually more challenging in RHC problems. Putting all the pieces together, we have the nonconvex loss function:

$$(9) \quad \ell := \ell(\mathbf{L}, \mathbf{R}; \mathbf{s}) := \psi(\mathbf{L}, \mathbf{R}; \mathbf{s}) + \lambda \phi(\mathbf{L}, \mathbf{R}),$$

where  $\lambda > 0$  is a regularization parameter.

<sup>3</sup> $a \lesssim \mathcal{O}(b)$  means there exists an absolute constant  $C > 0$  such that  $a \leq C \cdot b$ .

Based on the reformulated loss function (9), we detail the proposed HSGD method in three steps: (a) initialization, (b) iterative updates on outliers, and (c) iterative updates on the low-rank Hankel matrix. For ease of presentation, we start the discussion with iterative updates, followed by initialization.

*Iterative updates on outliers.* With the sparsity assumption on the outlier vector  $\mathbf{s}$ , we design a *sparsification operator* to filter large-magnitude entries:

$$[\Gamma_{\tilde{\alpha}}(\mathbf{v})]_i = \begin{cases} v_i & |v_i| \text{ is one of the largest } \tilde{\alpha}n \text{ element in } |\mathbf{v}|, \\ 0 & \text{otherwise} \end{cases}$$

for any vector  $\mathbf{v} \in \mathbb{C}^n$  and  $\tilde{\alpha} \in [0, 1)$ . At the  $(k+1)$ th iteration, we first compute the residues over the observed entries  $\Pi_{\Omega}(\mathbf{f} - \mathbf{z}^{(k)})$  where  $\mathbf{z}^{(k)} = \mathcal{G}^*(\mathbf{L}^{(k)} \mathbf{R}^{(k)*})$  is the reweighted vector representing the estimated low-rank Hankel matrix from the previous iteration. Then, we keep the largest  $\gamma_k \alpha p$ -fraction of the residue as the outliers, i.e.,

$$(10) \quad \mathbf{s}^{(k+1)} = \Gamma_{\gamma_k \alpha p} \left( \Pi_{\Omega} \left( \mathbf{f} - \mathbf{z}^{(k)} \right) \right),$$

where  $\gamma_k > 1$  is a parameter that allows us to overestimate the outlier density a bit.

Recall that  $\mathcal{G}^*$  is the adjoint operator of  $\mathcal{G} = \mathcal{H}\mathcal{D}^{-1}$ . For the  $t$ th entry of  $\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$ ,

$$[\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_t = \sum_{j=1}^r [\mathcal{G}^*(\mathbf{L}_{:,j} \mathbf{R}_{:,j}^*)]_t = \sum_{j=1}^r \frac{1}{\sqrt{\varsigma_t}} \sum_{i_1+i_2=t+1} L_{i_1,j} \bar{R}_{i_2,j},$$

where  $\varsigma_t$  is the number of entries on the  $t$ th antidiagonal of the  $n_1 \times n_2$  Hankel matrix. This suggests that  $\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$  can be computed via  $r$  fast convolutions (i.e., via FFT). Thus, (10) costs merely  $\mathcal{O}(rn \log n)$  flops.

*Iterative updates on low-rank Hankel matrix.* After removing the estimated outliers, we simultaneously update the factors  $\mathbf{L}$  and  $\mathbf{R}$  of the low-rank Hankel matrix via one step of *structured gradient descent*. The gradients with respect to  $\mathbf{L}$  and  $\mathbf{R}$  are calculated based on the loss function (9). Moreover, according to Assumption 2, we project the updates of  $\mathbf{L}$  and  $\mathbf{R}$  onto the convex sets

$$(11) \quad \mathcal{L} = \left\{ \mathbf{L} \mid \|\mathbf{L}\|_{2,\infty}^2 \leq \frac{2\mu r c_s}{n} \|\tilde{\mathbf{L}}^{(0)}\|_2^2 \right\} \text{ and } \mathcal{R} = \left\{ \mathbf{R} \mid \|\mathbf{R}\|_{2,\infty}^2 \leq \frac{2\mu r c_s}{n} \|\tilde{\mathbf{R}}^{(0)}\|_2^2 \right\},$$

respectively. The projection step ensures that the estimated low-rank matrix is always incoherent. Note that the ideal  $\mathcal{L}$  and  $\mathcal{R}$  should be defined with  $\|\mathcal{G}(\mathbf{z})\|_2^{1/2}$ . However, this information is usually unavailable to the user, so instead we use the initial estimations of  $\mathbf{L}$  and  $\mathbf{R}$ . It also requires an estimation of incoherence parameter  $\mu$ . Empirically,  $\mu$  can be well estimated by one step of the Cadzow algorithm [3, 4] if no prior knowledge is available. In summary, we have the projected gradient descent (PGD) for updating the low-rank Hankel matrix:

$$(12) \quad \begin{aligned} \mathbf{L}^{(k+1)} &= \Pi_{\mathcal{L}} \left( \mathbf{L}^{(k)} - \eta \nabla_{\mathbf{L}} \ell \left( \mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{s}^{(k+1)} \right) \right), \\ \mathbf{R}^{(k+1)} &= \Pi_{\mathcal{R}} \left( \mathbf{R}^{(k)} - \eta \nabla_{\mathbf{R}} \ell \left( \mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{s}^{(k+1)} \right) \right), \end{aligned}$$

where  $\eta > 0$  is the step size. The complexity of (12) is dominated by computing the gradients. Notice that

$$\nabla_{\mathbf{L}} \ell = \mathcal{G} \left( p^{-1} \Pi_{\Omega} (\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f}) - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) \right) \mathbf{R} + \mathbf{L} (\lambda \mathbf{L}^* \mathbf{L} + (1 - \lambda) \mathbf{R}^* \mathbf{R}).$$

As discussed, computing  $\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$  costs  $\mathcal{O}(rn \log n)$  flops, and so does computing the vector  $\mathbf{a} := p^{-1}\Pi_{\Omega}(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f}) - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$ . Next,  $\mathcal{G}(\mathbf{a})\mathbf{R}$  can be computed via another  $r$  fast convolutions since a Hankel matrix can be viewed as a convolution operator. Thus, the first term in  $\nabla_{\mathbf{L}}\ell(\mathbf{L}, \mathbf{R}; \mathbf{s})$  costs total  $\mathcal{O}(rn \log n)$  flops. While the second term costs  $\mathcal{O}(r^2n)$  flops, computing  $\nabla_{\mathbf{L}}\ell(\mathbf{L}, \mathbf{R}; \mathbf{s})$  costs merely  $\mathcal{O}(rn \log n + r^2n)$  flops. The same argument applies to  $\nabla_{\mathbf{R}}\ell(\mathbf{L}, \mathbf{R}; \mathbf{s})$ . Therefore, the update of the low-rank Hankel matrix is computationally efficient, in the complexity of  $\mathcal{O}(rn \log n + r^2n)$ .

*Initialization.* For a good initial guess, we modify the widely used spectral method [19, section VIII]. The first step is to detect the obvious (i.e., large) outliers from the observations:

$$\mathbf{s}^{(0)} = \mathcal{D}\Gamma_{\alpha p}(\Pi_{\Omega}\mathbf{y}) = \mathcal{D}\Gamma_{\alpha p}(\mathcal{D}^{-1}\Pi_{\Omega}\mathbf{f}),$$

where  $\alpha p$  comes from Assumption 3 and  $\mathcal{D}$  from (5). Next, we initialize

$$\tilde{\mathbf{L}}^{(0)} = \mathbf{U}^{(0)} \left( \boldsymbol{\Sigma}^{(0)} \right)^{1/2} \quad \text{and} \quad \tilde{\mathbf{R}}^{(0)} = \mathbf{V}^{(0)} \left( \boldsymbol{\Sigma}^{(0)} \right)^{1/2},$$

where  $\mathbf{U}^{(0)}\boldsymbol{\Sigma}^{(0)}\mathbf{V}^{(0)*}$  is the rank- $r$  truncated SVD of  $p^{-1}\mathcal{G}(\Pi_{\Omega}\mathbf{f} - \mathbf{s}^{(0)})$ . Then, the convex sets  $\mathcal{L}$  and  $\mathcal{R}$  can be defined based the spectral norms of  $\tilde{\mathbf{L}}^{(0)}$  and  $\tilde{\mathbf{R}}^{(0)}$  (see (11)). We immediately project  $\tilde{\mathbf{L}}^{(0)}$  and  $\tilde{\mathbf{R}}^{(0)}$  onto  $\mathcal{L}$  and  $\mathcal{R}$  to ensure their incoherence:

$$\mathbf{L}^{(0)} = \Pi_{\mathcal{L}}\tilde{\mathbf{L}}^{(0)} \quad \text{and} \quad \mathbf{R}^{(0)} = \Pi_{\mathcal{R}}\tilde{\mathbf{R}}^{(0)}.$$

This finishes the initialization.

The complexity of the initialization remains  $\mathcal{O}(rn \log n)$ , which is dominated by the step of truncated SVD. Although a typical truncated SVD costs  $\mathcal{O}(n^2r)$  flops, it costs only  $\mathcal{O}(rn \log n)$  flops on a Hankel matrix since the involved Hankel matrix-vector multiplications can be computed via fast convolutions.

We summarize the proposed HSGD as Algorithm 1. Although we are solving a matrix problem, HSGD never has to form the whole matrix due to the convenient Hankel structure—we only need to track the  $n$  distinct entries in the reweighted vector form. If the user needs the recovered result in Hankel matrix form, simply apply  $\mathcal{G}(\mathbf{z}^{(K)})$  on the output vector  $\mathbf{z}^{(K)}$ . Therefore, we conclude HSGD is both computationally and memory efficient. In particular, the overall computational complexity is as low as  $\mathcal{O}(rn \log n + r^2n)$ , as we discussed.

**2.1. Recovery guarantee.** In this section, we present the recovery guarantee of the proposed HSGD. Denote  $\mathbf{L}^{\natural} := \mathbf{U}^{\natural}\boldsymbol{\Sigma}^{\natural\frac{1}{2}}$  and  $\mathbf{R}^{\natural} := \mathbf{V}^{\natural}\boldsymbol{\Sigma}^{\natural\frac{1}{2}}$  where  $\mathbf{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\mathbf{V}^{\natural*}$  is the compact SVD of the underlying Hankel matrix  $\mathcal{G}\mathbf{z}^{\natural}$ . Consider the error that measures the distance between  $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$  and  $(\mathbf{L}^{\natural}, \mathbf{R}^{\natural})$ :

$$(13) \quad d_k := \text{dist} \left( \mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{L}^{\natural}, \mathbf{R}^{\natural} \right) := \min_{\mathbf{Q} \in \mathbb{Q}_r} \sqrt{\|\mathbf{L}^{(k)} - \mathbf{L}^{\natural}\mathbf{Q}\|_{\text{F}}^2 + \|\mathbf{R}^{(k)} - \mathbf{R}^{\natural}\mathbf{Q}\|_{\text{F}}^2}.$$

Therein,  $\mathbb{Q}_r$  denotes the set of  $r \times r$  rotation matrices and the best rotation matrix  $\mathbf{Q}$  is used to align  $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$  and  $(\mathbf{L}^{\natural}, \mathbf{R}^{\natural})$  since the matrix factorization is not unique. Note that the standard reconstruction error (i.e., in Frobenius norm) is controlled by

$$(14) \quad \|\mathcal{H}\mathbf{x}^{\natural} - \mathbf{L}^{(k)}\mathbf{R}^{(k)*}\|_{\text{F}}^2 \leq \sigma_1^{\natural} d_k^2,$$

provided  $d_k^2 \leq \sigma_1^{\natural}$  [39]. Thus, it is sufficient to bound  $d_k$  in our analysis.

We are ready to present our main results now. First, we present the local linear convergence of HSGD as Theorem 4, provided we have a sufficiently close initial guess.

**Algorithm 1** Hankel structured gradient descent (HSGD)

---

1: **Input:**  $\Pi_\Omega \mathbf{f}$ : partial observation on the corrupted Hankel matrix in reweighted vector form;  $r$ : the rank of underlying Hankel matrix;  $p$ : observation rate;  $\alpha$ : outlier density;  $\{\gamma_k\}$ : parameters for sparsification operator.  
2:  $\mathbf{s}^{(0)} = \mathcal{D}\Gamma_{\alpha p}(\mathcal{D}^{-1}\Pi_\Omega \mathbf{f})$   
3:  $[\mathbf{U}^{(0)}, \mathbf{\Sigma}^{(0)}, \mathbf{V}^{(0)}] = \text{SVD}_r(p^{-1}\mathcal{G}(\Pi_\Omega \mathbf{f} - \mathbf{s}^{(0)}))$   
4:  $\tilde{\mathbf{L}}^{(0)} = \mathbf{U}^{(0)}(\mathbf{\Sigma}^{(0)})^{1/2}$ ,  $\tilde{\mathbf{R}}^{(0)} = \mathbf{V}^{(0)}(\mathbf{\Sigma}^{(0)})^{1/2}$   
5: Define  $\mathcal{L}$  and  $\mathcal{R}$  by (11)  
6:  $\mathbf{L}^{(0)} = \Pi_{\mathcal{L}}\tilde{\mathbf{L}}^{(0)}$ ,  $\mathbf{R}^{(0)} = \Pi_{\mathcal{R}}\tilde{\mathbf{R}}^{(0)}$   
7: **for**  $k = 0, 1, \dots, K-1$  **do**  
8:    $\mathbf{z}^{(k)} = \mathcal{G}^*(\mathbf{L}^{(k)}\mathbf{R}^{(k)*})$   
9:    $\mathbf{s}^{(k+1)} = \Gamma_{\gamma_k \alpha p}(\Pi_\Omega(\mathbf{f} - \mathbf{z}^{(k)}))$   
10:    $\mathbf{L}^{(k+1)} = \Pi_{\mathcal{L}}(\mathbf{L}^{(k)} - \eta \nabla_{\mathbf{L}} \ell(\mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{s}^{(k+1)}))$   
11:    $\mathbf{R}^{(k+1)} = \Pi_{\mathcal{R}}(\mathbf{R}^{(k)} - \eta \nabla_{\mathbf{R}} \ell(\mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{s}^{(k+1)}))$   
12: **end for**  
13: **Output:**  $\mathbf{z}^{(K)} = \mathcal{G}^*(\mathbf{L}^{(K)}\mathbf{R}^{(K)*})$ : the recovered low-rank Hankel matrix in reweighted vector form.

---

**THEOREM 4.** Suppose Assumptions 1 to 3 hold with  $p \gtrsim \mathcal{O}((c_s^2 \mu^2 r^2 \log n)/n)$  and  $\alpha \lesssim \mathcal{O}(1/(c_s \mu r \kappa^2))$ . Choose the parameters  $\lambda = 1/16$ ,  $\gamma_k \in [1 + 1/b_0, 2]$  with some fixed  $b_0 \geq 1$ , and  $\eta = \tilde{c}/\sigma_r^{\frac{1}{2}}$  with some sufficiently small  $\tilde{c}$ . If the initialization satisfies  $d_0 \lesssim \mathcal{O}(\sqrt{\sigma_r^{\frac{1}{2}}/\kappa})$ , then with probability at least  $1 - 6n^{-2}$ , the iterations of HSGD satisfy

$$d_{k+1}^2 \leq \left(1 - \frac{\eta \sigma_r^{\frac{1}{2}}}{64}\right)^k d_0^2.$$

*Proof.* The proof of this theorem is deferred to Appendix A.3.  $\square$

Next, we present the initialization guarantee as Theorem 5. Therein, we show the initial guess  $(\mathbf{L}^{(0)}, \mathbf{R}^{(0)})$  falls in the basin of attraction specified in Theorem 4.

**THEOREM 5.** Suppose Assumptions 1 to 3 hold with  $p \geq (\varepsilon_0^{-2} \kappa^3 c_s \mu r^2 \log n)/n$  and  $\alpha \leq 1/(32c_s \mu r \kappa)$  where  $\varepsilon_0 \in (0, \frac{\sqrt{\kappa T}}{8c_0})$  with some constant  $c_0$ . Then with probability at least  $1 - 2n^{-2}$ , the initialization step of HSGD satisfies

$$d_0 \leq 26\alpha c_s \kappa \mu r \sqrt{r} \sqrt{\sigma_r^{\frac{1}{2}} + 7c_0 \varepsilon_0} \sqrt{\sigma_r^{\frac{1}{2}}/\kappa}.$$

*Proof.* The proof of this theorem is deferred to Appendix A.2.  $\square$

Note that the probabilities in Theorems 4 and 5 come from some analogous restricted isometry properties (see Lemmas 9 to 11), which hold uniformly for our results. Thus, all our theorems hold uniformly with a probability at least  $1 - 6n^{-2}$ . By directly combining (14) and Theorems 4 and 5, we show HSGD has global convergence to the ground truth with high probability, provided sufficiently many samples and sufficiently sparse outliers.

**COROLLARY 6.** Suppose Assumptions 1 to 3 hold with

$$p \gtrsim \mathcal{O}\left(\frac{\max\{c_s^2 \mu^2 r^2 \log n, c_s \mu \kappa^3 r^2 \log n\}}{n}\right) \text{ and } \alpha \lesssim \mathcal{O}\left(\frac{1}{\max\{c_s \mu \kappa^{3/2} r^{3/2}, c_s \mu r \kappa^2\}}\right).$$



Then, HSGD finds an  $\varepsilon$ -optimal solution, i.e.,  $\|\mathcal{H}\mathbf{x}^\natural - \mathbf{L}^{(K)}\mathbf{R}^{(K)*}\|_F / \|\mathcal{H}\mathbf{x}^\natural\|_F \leq \varepsilon$ , in  $K = \mathcal{O}(\kappa \log \varepsilon^{-1})$  iterations with probability at least  $1 - 6n^{-2}$ .

*Remark 7.* When outliers are not appearing (i.e.,  $\alpha = 0$ ), the RHC problem reduces to the vanilla Hankel matrix completion problem, and HSGD becomes the PGD introduced in [10]. In the original paper, PGD theoretically requires  $\mathcal{O}(\kappa^2 \log \varepsilon^{-1})$  iterations to find an  $\varepsilon$ -optimal solution. With the improved proof techniques, we show that running  $\mathcal{O}(\kappa \log \varepsilon^{-1})$  iterations is sufficient for HSGD. Thus, as a special case of HSGD, we also theoretically improve the convergence speed of PGD to  $\mathcal{O}(\kappa \log \varepsilon^{-1})$  iterations.

*Remark 8.* The convergence rate of HSGD suggests that the proposed algorithm runs faster on well-conditioned problems, i.e.,  $\kappa = \mathcal{O}(1)$ . In many applications, the condition number of the underlying Hankel matrix is indeed good. For example, in an NMR spectroscopy problem,  $\kappa$  depends on the ratio between the largest and smallest magnitudes of the complex amplitudes [11, Remark 1], which typically is modest.

**3. Numerical experiments.** In this section, we compare the proposed HSGD against the state-of-the-art RHC approaches, PartialSAP [41] and RobustEMaC [16]. We demonstrate the empirical advantages of HSGD on both synthetic and real datasets. We hand tuned the parameters for their best performance. In particular, we use an iterative decaying  $\gamma_k = 1.05 + 0.45 \cdot 0.95^k$  for HSGD, so it starts with  $\gamma_0 = 1.5$  and  $\gamma_k \rightarrow 1.05$  as  $k \rightarrow \infty$ . By Theorem 4, any  $\gamma_k \in [1.05, 2]$  (i.e.,  $b_0 = 20$ ) will work, and we find the iterative decaying  $\gamma_k$  provides the best empirical performance for HSGD. The reason behind this parameter choice is that HSGD gets better outlier estimations in the latter iterations, so fewer false-positive outliers will be taken. For the sake of fairness, incoherence parameter  $\mu$  is estimated by one step of the Cadzow algorithm [3] through all experiments. Further discussion about the input parameters is given in the supplementary material (Supplementary\_Materials.pdf [local/web 606KB]). All numerical experiments were performed in MATLAB on a Windows laptop equipped with an Intel i7-8750H CPU and 32 GB RAM. For a fair comparison, PROPACK [24] was used for fast truncated SVD in all tested algorithms. The MATLAB implementation of HSGD is available online at <https://github.com/caesarcai/HSGD>.

**3.1. Synthetic examples.** We generate the rank- $r$  Hankel matrices via two steps: (i) generate a vector  $\mathbf{x}^\natural \in \mathbb{C}^n$  that is sparse in Fourier space with exact  $r$  active frequencies;<sup>4</sup> then (ii) generate the corresponding Hankel matrix  $\mathcal{H}(\mathbf{x}^\natural) \in \mathbb{C}^{n_1 \times n_2}$  with  $n_1 \approx n_2 \approx n/2$ .<sup>5</sup> Such a Hankel matrix must be rank- $r$  [26]. For the suitable algorithms, we also generate the reweighted vector  $\mathbf{z}^\natural = \mathcal{D}\mathbf{x}^\natural$  so that  $\mathcal{H}(\mathbf{x}^\natural) = \mathcal{G}(\mathbf{z}^\natural)$ . We uniformly (without replacement) observe  $m := pn$  entries from  $\mathbf{z}^\natural$ , then we uniformly choose  $\alpha m$  entries among the observed ones to be corrupted. The corruption is done by adding complex outliers whose real parts and imaginary parts are drawn independent and identically distributed from the uniform distribution over the intervals  $[-10\mathbb{E}(|\operatorname{Re}(z_i^\natural)|), 10\mathbb{E}(|\operatorname{Re}(z_i^\natural)|)]$  and  $[-10\mathbb{E}(|\operatorname{Im}(z_i^\natural)|), 10\mathbb{E}(|\operatorname{Im}(z_i^\natural)|)]$ , respectively. In the experiments, we use a uniform sampling model instead of the Bernoulli sampling model since with the former it is easier to control the number of samples and outliers. We emphasize that the empirical behaviors are not much different between these two sampling models.

<sup>4</sup>We follow the same method used in [4, section III.A] to generate such vectors. In our tests, we ensure the active frequencies are well separated in the generated vectors.

<sup>5</sup>If  $n$  is odd, we use  $n_1 = n_2 = (n + 1)/2$ . If  $n$  is even, we use  $n_1 = n_2 - 1 = n/2$ .

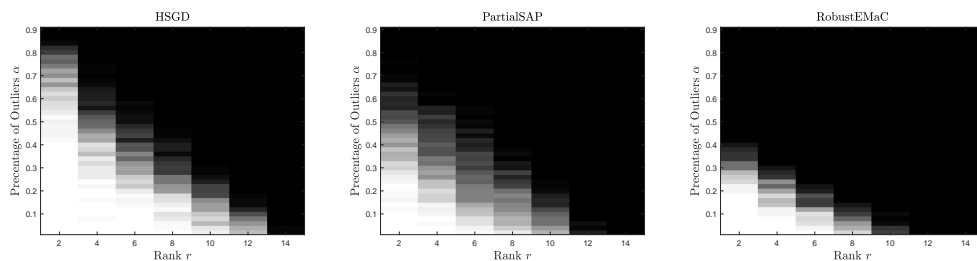


FIG. 1. Empirical phase transition for HSGD, PartialSAP, and RobustEMaC: Rank versus rate of outliers. 50 entries are sampled in all testing problems.

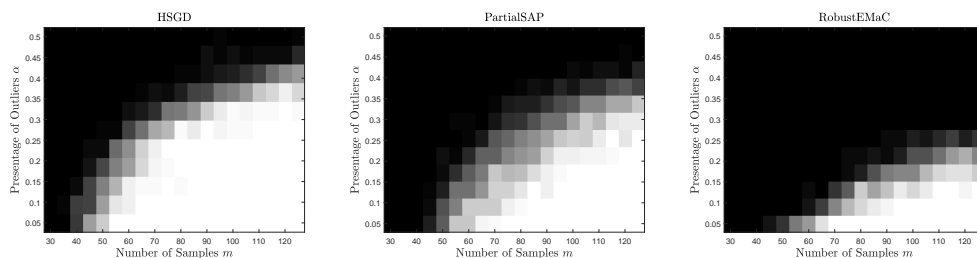


FIG. 2. Empirical phase transition for HSGD, PartialSAP, and RobustEMaC: Number of samples versus rate of outliers. All testing problems have rank 10.

*Empirical phase transition.* In this section, we present the recoverability of the tested algorithms under various settings. We fix  $n = 125$  for all experiments in the section. The pixels on the phase transition plots represent different problem settings. For each pixel, we conduct 50 testing problems; a white pixel means all 50 cases were recovered and a black pixel means all 50 cases were failed. Specifically speaking, the output of a testing problem is considered a successful recovery if  $\|\mathcal{G}\mathbf{z}^{(K)} - \mathcal{G}\mathbf{z}^{\dagger}\|_{\mathbf{F}} / \|\mathcal{G}\mathbf{z}^{\dagger}\|_{\mathbf{F}} \leq 10^{-3}$ , while the stopping criteria is  $\|\mathcal{G}\Pi_{\Omega}\mathbf{z}^{(k)} + \mathcal{G}\Pi_{\Omega}\mathbf{s}^{(k)} - \mathcal{G}\Pi_{\Omega}\mathbf{f}\|_{\mathbf{F}} / \|\mathcal{G}\Pi_{\Omega}\mathbf{f}\|_{\mathbf{F}} \leq 10^{-5}$ .<sup>6</sup>

In Figure 1, we fix the number of samples  $m = 50$  and study the empirical phase transition with varying rank  $r$  and outlier sparsity  $\alpha$ . In Figure 2, we fix rank  $r = 10$  and study the empirical phase transition with varying outlier sparsity  $\alpha$  and number of samples  $m$ . In Figure 3, we fix the outlier sparsity  $\alpha = 0.1$  and study the empirical phase transition with varying rank  $r$  and number of samples  $m$ . In all three comparisons, we find HSGD has the best recoverability and the other nonconvex algorithm, i.e., PartialSAP, is competitive.

*Computational efficiency.* We demonstrate the computational efficiency of the tested algorithms. All testing problems in this experiment have rank  $r = 10$ , observation rate  $p = 40\%$ , and outlier sparsity  $\alpha = 10\%$ . The reported runtime is averaged over 20 trials. Since the large-scale problems are prohibitive for the convex method Robust-EMaC, we only compare HSGD against PartialSAP in this section. While both nonconvex algorithms have similar complexity orders, the leading constant for HSGD is expected to be much smaller. In the left subfigure of Figure 4, we exponentially increase the problem dimension  $n$  and record the runtime. We observe that HSGD is  $10\times$  faster than PartialSAP when the problem dimension is large. In the middle subfigure of Figure 4, we run the same dimension versus runtime experiment with only HSGD and even larger dimensions. In this plot, we match the logarithmic

<sup>6</sup>When suitable, we calculate the empirical residue in the equivalent vector form to save runtime.

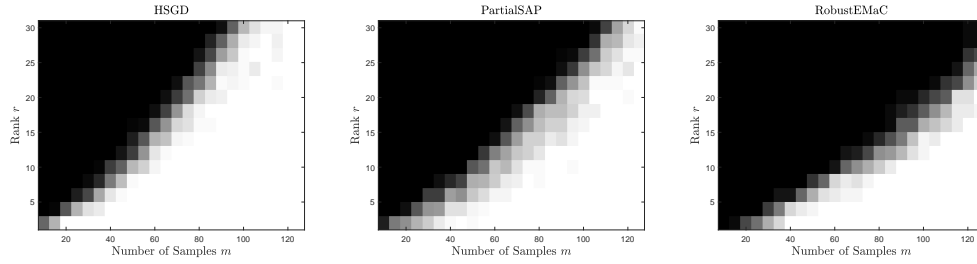


FIG. 3. Empirical phase transition for HSGD, PartialSAP, and RobustEMaC: Number of samples versus rank. 10% of samples are corrupted by outliers in all testing problems.

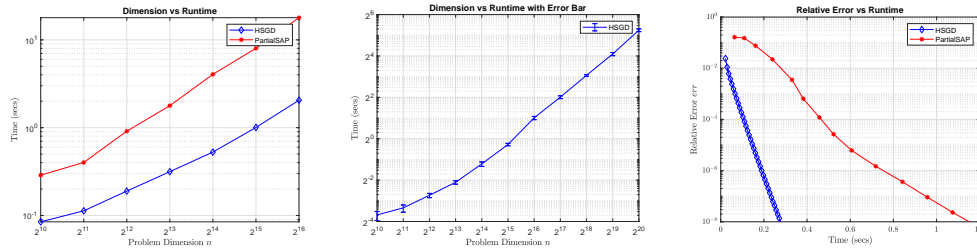


FIG. 4. Experimental results for speed tests between HSGD and PartialSAP. Left: Dimension versus runtime. Middle: Dimension versus runtime with error bar for HSGD only. Right: Relative error versus runtime.

base for the  $x$ -axis and the  $y$ -axis in the log-log plot and also include error bars. The slope of this log-log plot is approximately 1 when  $n$  is large, which verifies the claimed computational complexity for HSGD, i.e., the dependence on problem dimension is merely  $\mathcal{O}(n \log n)$ . Moreover, the narrow error bar in the plot shows the runtime of HSGD is stable. In the right subfigure of Figure 4, we present the convergence behavior of the tested algorithms where we fix  $n = 2^{12}$ . One can see both algorithms have linear convergence as the theorems indicated, and HSGD has a sharper convergence rate with respect to runtime. Overall, we conclude HSGD is a highly efficient algorithm, compared to the state-of-the-art RHC approaches.

**3.2. Nuclear magnetic resonance spectroscopy.** As we discussed in section 1, NMR signal recovery is a widely used real-world benchmark for the problems of the low-rank Hankel matrix: Given a clear one-dimensional NMR signal  $\mathbf{x}^b$ , the corresponding Hankel matrix  $\mathcal{H}\mathbf{x}^b$  is rank- $r$ , where  $r$  is determined by the number of bars in the power spectrum of the signal. In this section, we apply RHC algorithms to complete the partially observed NMR signal and remove the impulse corruptions, simultaneously. The testing NMR signal has the dimension  $n = 32,768$  and rank  $r \approx 40$ , which is a prohibitively large size for RobustEMaC. We test HSGD and PartialSAP for recovering this signal under various observation rate  $p$  and outlier sparsity  $m$ . The runtime comparison results are summarized as Table 1, where tested algorithms recover the desired signal in all cases. One can see that HSGD maintains its speed advantage in this real-world application, under each of the settings. Moreover, in Figure 5, we demonstrate the power spectrum of the signal recovered by HSGD. Therein, we not only successfully recover the NMR signal but also clear the small noise in the original data. Although it is not theoretically verified, the empirical results suggest that HSGD can also denoise small white noise when it detects extreme outliers.

TABLE 1

Runtime comparison between HSGD and PartialSAP for NMR signal recovery under various observation rate  $p$  and outlier sparsity  $\alpha$ .

$(p, \alpha)$	(0.3, 0.1)	(0.3, 0.2)	(0.3, 0.3)	(0.4, 0.4)	(0.5, 0.5)
PartialSAP	335.09 s	346.328 s	352.01 s	357.75 s	369.32 s
HSGD	31.142 s	33.67 s	45.82 s	53.69 s	64.11 s

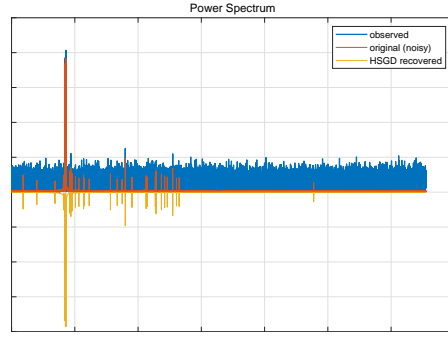


FIG. 5. Power spectrum of the noisy original NMR signal, the observed signal ( $p = 30\%$  and  $\alpha = 30\%$ ), and HSGD recovered signal (upside down). Note that the observed signal in the picture is rescaled by  $1/p$ , which is a common method to offset the energy loss due to partial observation.

**4. Concluding remarks.** In this work, we proposed a novel nonconvex algorithm, coined Hankel structured gradient descent (HSGD), for robust Hankel matrix completion problems. HSGD is highly computing- and sample-efficient. In particular, HSGD costs merely  $\mathcal{O}(r^2n + rn \log n)$  flops per iteration, while it requires as few as  $\mathcal{O}(\max\{c_s^2 \mu^2 r^2 \log n, c_s \mu \kappa^3 r^2 \log n\})$  samples. HSGD is also robust and tolerates  $\mathcal{O}(1/\max\{c_s \mu \kappa^{3/2} r^{3/2}, c_s \mu r \kappa^2\})$ -fraction outliers. Theoretical recovery guarantees have been established for HSGD, along with a provable linear convergence rate. The superior performance of HSGD, in terms of efficiency and robustness, is verified by numerical experiments on both synthetic and real datasets.

**Appendix A. Proofs of theoretical results.** In this section, we provide the analysis for the claimed theoretical results. All the proofs are under Assumptions 1 to 3. We start with introducing some additional notation used in the analysis. We denote the tangent space of the rank- $r$  matrix manifold at  $\mathcal{G}\mathbf{z}^\natural$  by

$$T := \left\{ \mathbf{X} \mid \mathbf{X} = \mathbf{U}^\natural \mathbf{C}^* + \mathbf{D} \mathbf{V}^{\natural*}, \text{ where } \mathbf{C} \in \mathbb{C}^{n_1 \times r}, \mathbf{D} \in \mathbb{C}^{n_2 \times r} \right\}.$$

For any  $\mathbf{Q} \in \mathbb{Q}_r$ ,  $(\mathbf{L}^\natural \mathbf{Q}, \mathbf{R}^\natural \mathbf{Q})$  is an equivalent solution to  $(\mathbf{L}^\natural, \mathbf{R}^\natural)$ . Thus, we define the solution set to be

$$\mathcal{E}(\mathbf{L}^\natural, \mathbf{R}^\natural) := \{(\mathbf{L}, \mathbf{R}) \in \mathbb{C}^{n_1 \times r} \times \mathbb{C}^{n_2 \times r} \mid \mathbf{L} = \mathbf{L}^\natural \mathbf{Q}, \mathbf{R} = \mathbf{R}^\natural \mathbf{Q}, \text{ where } \mathbf{Q} \in \mathbb{Q}_r\}.$$

The sequence  $\{\tilde{\mathbf{d}}_k\}_{k \geq 0}$  is defined as

$$\tilde{\mathbf{d}}_k := \text{dist}(\tilde{\mathbf{L}}^{(k)}, \tilde{\mathbf{R}}^{(k)}; \mathbf{L}^\natural, \mathbf{R}^\natural),$$

where we use  $\tilde{\mathbf{L}}^{(k)} := \mathbf{L}^{(k-1)} - \eta \nabla_{\mathbf{L}} \ell(\mathbf{L}^{(k-1)}, \mathbf{R}^{(k-1)}; \mathbf{s}^{(k)})$  and  $\tilde{\mathbf{R}}^{(k)} := \mathbf{R}^{(k-1)} - \eta \nabla_{\mathbf{R}} \ell(\mathbf{L}^{(k-1)}, \mathbf{R}^{(k-1)}; \mathbf{s}^{(k)})$  to denote the middle step of (12) before the projection onto

$\mathcal{L}$  and  $\mathcal{R}$ , respectively. With a slight abuse of notation, we let  $\mathcal{P}_\Omega$  be the projection operator onto the space which can be represented by an orthonormal basis of Hankel matrices. That is, for any matrix  $\mathbf{Z} \in \mathbb{C}^{n_1 \times n_2}$ ,

$$\mathcal{P}_\Omega(\mathbf{Z}) = \sum_{a \in \Omega} \langle \mathbf{Z}, \mathbf{H}_a \rangle \mathbf{H}_a,$$

where  $\{\mathbf{H}_a\}_{a=1}^n$  is the orthonormal basis of Hankel matrices, defined by

$$\mathbf{H}_a := \frac{1}{\sqrt{\zeta_a}} \mathcal{H} \mathbf{e}_a, \quad \text{where } \mathbf{e}_a \text{ is the } a\text{th standard basis vector of } \mathbb{R}^n.$$

By this definition, for any Hankel matrix  $\mathcal{G}\mathbf{z}$ , we have  $\mathcal{P}_\Omega \mathcal{G}\mathbf{z} = \mathcal{G}\Pi_\Omega \mathbf{z}$ .  $\delta_a$  is defined as

$$(15) \quad \delta_a = \begin{cases} 1 & \text{with probability } p; \\ 0 & \text{otherwise} \end{cases}$$

for all  $a \in [n]$ .

**A.1. Technical lemmas.** We prove some technical lemmas that will be used in the convergence analysis in this subsection.

LEMMA 9. *There exists some universal constant  $c_0 > 0$  such that*

$$(16) \quad \|(p^{-1} \mathcal{G}\Pi_\Omega - \mathcal{G}) \mathbf{z}^\natural\|_2 \leq c_0 \sqrt{(p^{-1} \mu c_s r \log n)/n} \|\mathcal{G}\mathbf{z}^\natural\|_2$$

*holds with probability at least  $1 - 2n^{-2}$  provided  $p \geq (\mu c_s r \log n)/n$ .*

*Proof.* By the definition of  $\delta_a$  in (15), first we have

$$p^{-1} \mathcal{G}\Pi_\Omega \mathbf{z}^\natural - \mathcal{G}\mathbf{z}^\natural = p^{-1} (\delta_a - p) \mathbf{z}_a^\natural \mathbf{H}_a.$$

Denote  $\mathbf{Z}_a := p^{-1} (\delta_a - p) \mathbf{z}_a^\natural \mathbf{H}_a$ . Thus,  $\mathbb{E}[\mathbf{Z}_a] = 0$ . By  $\|\mathbf{H}_a\|_2 \leq \frac{1}{\sqrt{\zeta_a}}$ , we have

$$\|\mathbf{Z}_a\|_2 \leq p^{-1} |\mathbf{z}_a^\natural| \|\mathbf{H}_a\|_2 \leq p^{-1} \frac{|\mathbf{z}_a^\natural|}{\sqrt{\zeta_a}} \leq p^{-1} \|\mathcal{D}^{-1} \mathbf{z}^\natural\|_\infty.$$

Note that  $\mathbf{Z}_a \mathbf{Z}_a^* = (\frac{\delta_a}{p} - 1)^2 |\mathbf{z}_a^\natural|^2 \mathbf{H}_a \mathbf{H}_a^*$ . Thus,

$$\left\| \mathbb{E} \sum_a \mathbf{Z}_a \mathbf{Z}_a^* \right\|_2 \leq \frac{1}{p} \left\| \sum_a |\mathbf{z}_a^\natural|^2 \mathbf{H}_a \mathbf{H}_a^* \right\|_2 \leq \frac{1}{p} \left\| \text{diag}(\mathcal{G}\mathbf{z}^\natural (\mathcal{G}\mathbf{z}^\natural)^*) \right\|_2 \leq \frac{1}{p} \|\mathcal{G}\mathbf{z}^\natural\|_{2,\infty}^2.$$

Similarly, we have  $\|\mathbb{E}(\sum_a \mathbf{Z}_a^* \mathbf{Z}_a)\|_2 \leq \frac{1}{p} \|(\mathcal{G}\mathbf{z}^\natural)^*\|_{2,\infty}^2$ . Moreover, by the  $\mu$ -incoherence condition, we have

$$\|\mathcal{G}\mathbf{z}^\natural\|_{2,\infty}^2 = \max_i \left\| \mathbf{e}_i^* \mathbf{U}^\natural \boldsymbol{\Sigma}^\natural \mathbf{V}^{\natural*} \right\|_2^2 \leq \|\mathcal{G}\mathbf{z}^\natural\|_2^2 \max_i \left\| \mathbf{e}_i \mathbf{U}^\natural \right\| \leq \frac{c_s \mu r}{n} \|\mathcal{G}\mathbf{z}^\natural\|_2^2.$$

Similarly, we also have  $\|(\mathcal{G}\mathbf{z}^\natural)^*\|_{2,\infty}^2 \leq \frac{c_s \mu r}{n} \|\mathcal{G}\mathbf{z}^\natural\|_2^2$ . Note that

$$\|\mathcal{D}^{-1} \mathbf{z}^\natural\|_\infty = \|\mathcal{G}\mathbf{z}^\natural\|_\infty = \max_{i,j} |\mathbf{e}_i^* (\mathcal{G}\mathbf{z}^\natural) \mathbf{e}_j| \leq \frac{c_s \mu r}{n} \|\mathcal{G}\mathbf{z}^\natural\|_2.$$

Using the bounds of  $\|\mathcal{G}\mathbf{z}^\natural\|_{2,\infty}$ ,  $\|(\mathcal{G}\mathbf{z}^\natural)^*\|_{2,\infty}$ , and  $\|\mathcal{D}^{-1} \mathbf{z}^\natural\|_\infty$ , the Bernstein's inequality [34, Theorem 1.6] then yields

$$(17) \quad \mathbb{P} \left( \left\| \sum_a \mathbf{Z}_a \right\|_2 > t \right) \leq (n_1 + n_2) \exp \left( \frac{-n p t^2 / 2}{\|\mathcal{G}\mathbf{z}^\natural\|_2^2 c_s \mu r + t \|\mathcal{G}\mathbf{z}^\natural\|_2 c_s \mu r / 3} \right).$$

Let  $t = c_0 \sqrt{(\mu c_s r \log n)/(pn)} \|\mathcal{G}z^\natural\|_2$  and  $p \geq (\mu c_s r \log n)/n$ . We then have

$$\mathbb{P}\left(\left\|p^{-1}\mathcal{G}\Pi_\Omega z^\natural - \mathcal{G}z^\natural\right\|_2 > c_0 \sqrt{p^{-1}\mu c_s r \log n/n} \|\mathcal{G}z^\natural\|_2\right) \leq 2n^{-2}$$

for any constant  $c_0 \geq 4$ .  $\square$

LEMMA 10. *There exists a constant  $c_4$  such that if  $p \geq (\varepsilon_0^{-2} c_4 \mu r \log n)/n$ , it holds that*

$$(18) \quad \left\|p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{G}\mathcal{G}^*\mathcal{P}_T - \mathcal{P}_T\mathcal{G}\mathcal{G}^*\mathcal{P}_T\right\| \leq \varepsilon_0$$

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

*Proof.* Notice  $\mathcal{G}\mathcal{G}^*\mathbf{H}_a = \mathbf{H}_a$ . For any matrix  $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ , we have

$$\begin{aligned} \left(\frac{1}{p}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{G}\mathcal{G}^*\mathcal{P}_T - \mathcal{P}_T\mathcal{G}\mathcal{G}^*\mathcal{P}_T\right)(\mathbf{X}) &= \mathcal{P}_T\left(\frac{1}{p}\sum_a \delta_a \langle \mathcal{P}_T\mathbf{X}, \mathcal{G}\mathcal{G}^*\mathbf{H}_a \rangle \mathbf{H}_a - \mathcal{G}\mathcal{G}^*\mathcal{P}_T\mathbf{X}\right) \\ &= \sum_a (p^{-1}\delta_a - 1) \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a. \end{aligned}$$

Denote  $\mathcal{Y}_a : \mathbf{X} \mapsto (p^{-1}\delta_a - 1) \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a$ . Thus  $\mathbb{E}[\mathcal{Y}_a] = 0$  and

$$\begin{aligned} \|\mathcal{Y}_a\| &= \sup_{\|\mathbf{X}\|_F=1} \left\| (p^{-1}\delta_a - 1) \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \right\|_F \\ &\leq p^{-1} \sup_{\|\mathbf{X}\|_F=1} \|\mathbf{X}\|_F \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \leq 2p^{-1}c_s\mu r/n, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \|\mathcal{P}_T\mathbf{H}_a\|_F^2 &\leq \|\mathcal{P}_{U^\natural}\mathbf{H}_a\|_F^2 + \|\mathcal{P}_{V^\natural}\mathbf{H}_a\|_F^2 = \left\|U^\natural U^{\natural*} \mathbf{H}_a\right\|_F^2 + \left\|\mathbf{H}_a V^\natural V^{\natural*}\right\|_F^2 \\ &= \left\|U^{\natural*} \mathbf{H}_a\right\|_F^2 + \left\|\mathbf{H}_a V^\natural\right\|_F^2 \leq \left\|U^\natural\right\|_{2,\infty}^2 + \left\|V^\natural\right\|_{2,\infty}^2 \leq 2c_s\mu r/n. \end{aligned}$$

Moreover, we notice that  $\mathcal{Y}_a$  is self-adjoint and it holds that

$$\begin{aligned} \mathcal{Y}_a^2(\mathbf{X}) &= \mathcal{Y}_a \left( (p^{-1}\delta_a - 1) \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \right) \\ &= (p^{-1}\delta_a - 1)^2 \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \langle \mathcal{P}_T\mathbf{H}_a, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \\ &= (p^{-1}\delta_a - 1)^2 \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \left\| \mathbb{E} \left( \sum_a \mathcal{Y}_a^2 \right) \right\| &= \sup_{\|\mathbf{X}\|_F=1} \left\| \mathbb{E} \left[ \sum_a (p^{-1}\delta_a - 1)^2 \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \right] \right\|_F \\ &\leq p^{-1} \sup_{\|\mathbf{X}\|_F=1} \left\| \left( \sum_a \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \right) \right\|_F \\ &\leq p^{-1} \max_a \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \sup_{\|\mathbf{X}\|_F=1} \left\| \left( \sum_a \langle \mathbf{X}, \mathcal{P}_T\mathbf{H}_a \rangle \mathcal{P}_T\mathbf{H}_a \right) \right\|_F \\ &= p^{-1} \max_a \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \sup_{\|\mathbf{X}\|_F=1} \|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{X})\|_F \\ &\leq p^{-1} \max_a \|\mathcal{P}_T\mathbf{H}_a\|_F^2 \leq 2p^{-1}c_s\mu r/n, \end{aligned}$$

where the last inequality follows from  $\|\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X})\|_F^2 = \langle \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X}), \mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X}) \rangle \leq \|\mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X})\|_F \|\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X})\|_F$  and thus  $\|\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X})\|_F \leq \|\mathcal{P}_\Omega \mathcal{P}_T(\mathbf{X})\|_F \leq \|\mathbf{X}\|_F$ . Then, by the Bernstein's inequality [34, Theorem 1.6], we have

$$\mathbb{P}\left(\left\|\sum_a \mathbf{y}_a\right\| > t\right) \leq (n_1 + n_2) \exp\left(\frac{-pnt^2/2}{2c_s\mu r + 2c_s\mu rt/3}\right).$$

For any  $\varepsilon_0 > 0$ , let  $t = \varepsilon_0 > 0$ , and  $p \geq c_4(\varepsilon_0^{-2}c_1\mu r \log n)/n$  for some universal constant  $c_4 > 12 + 12\varepsilon_0$ . The above inequality then implies  $\mathbb{P}(\|\sum_a \mathbf{y}_a\| > \varepsilon_0) \leq 2n^{-2}$ .  $\square$

LEMMA 11. For any  $\mathbf{U} \in \mathbb{C}^{n_1 \times r}$  and  $\mathbf{V} \in \mathbb{C}^{n_2 \times r}$ , if  $p \geq (\log n)/n$ , then it holds that

$$(19) \quad p^{-1} \|\mathcal{P}_\Omega(\mathbf{UV}^*)\|_F^2 \leq \|\mathbf{UV}^*\|_F^2 + \sqrt{8p^{-1}n \log n} \|\mathbf{U}\|_F \|\mathbf{V}\|_F \|\mathbf{U}\|_{2,\infty} \|\mathbf{V}\|_{2,\infty}$$

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

*Proof.* By the supporting Lemma 19 in Appendix B, we have

$$\begin{aligned} p^{-1} \|\mathcal{P}_\Omega(\mathbf{UV}^*)\|_F^2 &= p^{-1} \sum_a \left| \sum_{i+j=a+1} \delta_a \langle \mathbf{UV}^*, \mathbf{e}_i \mathbf{e}_j^\top \rangle \right|^2 \\ &\leq p^{-1} \sum_a \sum_{i+j=a+1} \delta_a \|\mathbf{U}_{(i,:)}\|_2^2 \|\mathbf{V}_{(j,:)}\|_2^2 \\ &\leq \|\mathbf{UV}^*\|_F^2 + \sqrt{8p^{-1}n \log n} \sqrt{\sum_i \|\mathbf{U}_{(i,:)}\|_2^4} \sqrt{\sum_j \|\mathbf{V}_{(j,:)}\|_2^4} \\ &\leq \|\mathbf{UV}^*\|_F^2 + \sqrt{p^{-1}8n \log n} \|\mathbf{U}\|_F \|\mathbf{U}\|_{2,\infty} \|\mathbf{V}\|_F \|\mathbf{V}\|_{2,\infty}. \quad \square \end{aligned}$$

LEMMA 12. For any matrix  $\mathbf{A} \in T$ , under event (18), it holds that

$$\begin{aligned} p(1 - \varepsilon_0) \|\mathbf{A}\|_F^2 &\leq \|\mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathbf{A}\|_F^2 \leq p(1 + \varepsilon_0) \|\mathbf{A}\|_F^2, \\ \|\mathcal{G}(p^{-1}\Pi_\Omega - \mathcal{I})\mathcal{G}^* \mathbf{A}\|_F^2 &\leq \varepsilon_0 \|\mathbf{A}\|_F^2. \end{aligned}$$

*Proof.* For  $\mathbf{A} \in T$ , we have  $\mathcal{P}_T \mathbf{A} = \mathbf{A}$ . Thus, by Lemma 10 we have

$$\|\mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathbf{A}\|_F^2 = \langle \mathcal{P}_T \mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathcal{P}_T \mathbf{A}, \mathbf{A} \rangle \leq \|\mathcal{P}_T \mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathcal{P}_T \mathbf{A}\|_F \|\mathbf{A}\|_F \leq p(1 + \varepsilon_0) \|\mathbf{A}\|_F^2.$$

Similarly,  $p(1 - \varepsilon_0) \|\mathbf{A}\|_F^2 \leq \|\mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathbf{A}\|_F^2$ . Using Lemma 10 and the fact that  $\|\mathcal{P}_\Omega \mathcal{G} \mathcal{G}^* \mathbf{X}\|_F = \|\mathcal{G} \mathcal{G}^* \mathcal{P}_\Omega \mathbf{X}\|_F$  for  $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ , we further have

$$\left\| \mathcal{G} \left( \frac{1}{p} \Pi_\Omega - \mathcal{I} \right) \mathcal{G}^* \mathbf{A} \right\|_F^2 = \left\langle \mathcal{P}_T \left( \frac{1}{p} \mathcal{P}_\Omega - \mathcal{I} \right) \mathcal{G} \mathcal{G}^* \mathcal{P}_T \mathbf{A}, \mathbf{A} \right\rangle \leq \varepsilon_0 \|\mathbf{A}\|_F^2. \quad \square$$

## A.2. Proof of Theorem 5 (guaranteed initialization).

*Proof of Theorem 5.* We will finish the proof in three steps under event (16). For simplicity, following the notation in Algorithm 1, we denote

$$(20) \quad \mathbf{M}^{(0)} := p^{-1} \mathcal{G} \left( \Pi_\Omega \mathbf{f} - \mathbf{s}^{(0)} \right), \quad \text{and } \mathbf{M}_r^{(0)} \text{ is the top-}r \text{ SVD of } \mathbf{M}^{(0)}.$$

*Step 1.* We first bound  $\|\mathbf{M}^{(0)} - \mathcal{G}\mathbf{z}^\natural\|_2$ . By the triangle inequality we have

$$(21) \quad \left\| \mathbf{M}^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 \leq \left\| \mathbf{M}^{(0)} - p^{-1}\mathcal{G}\Pi_\Omega \mathbf{z}^\natural \right\|_2 + \left\| p^{-1}\mathcal{G}\Pi_\Omega \mathbf{z}^\natural - \mathcal{G}\mathbf{z}^\natural \right\|_2.$$

The definition in (20) yields

$$\left\| \mathbf{M}^{(0)} - \frac{1}{p}\mathcal{G}\Pi_\Omega \mathbf{z}^\natural \right\|_2 = \left\| \frac{1}{p}\mathcal{G} \left( \Pi_\Omega \mathbf{f} - \mathbf{s}^{(0)} \right) - \frac{1}{p}\mathcal{G}\Pi_\Omega \mathbf{z}^\natural \right\|_2 = \left\| \frac{1}{p}\mathcal{G} \left( \Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)} \right) \right\|_2.$$

Denote the support of  $\Pi_\Omega \mathbf{s}^\natural$  and  $\mathbf{s}^{(0)}$  by  $\Omega_s^\natural$  and  $\Omega_s^{(0)}$ , respectively. Notice that  $\Omega_s^\natural \subseteq \Omega$ ,  $\Omega_s^{(0)} \subseteq \Omega$ . By the definition of  $\mathbf{s}^{(0)}$  in Algorithm 1, we have

$$\left[ \mathcal{D}^{-1}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}) \right]_i = -(\mathcal{D}^{-1} \mathbf{z}^\natural)_i \quad \text{for } i \in \Omega_s^{(0)}.$$

Recall  $\Pi_\Omega \mathbf{f} = \Pi_\Omega \mathbf{z}^\natural + \Pi_\Omega \mathbf{s}^\natural$  and  $\|\Pi_\Omega \mathbf{s}^\natural\|_0 \leq \alpha p n$ . We see there are no more than  $\alpha p n$  elements in  $\mathcal{D}^{-1}\Pi_\Omega \mathbf{f}$  such that  $|(\mathcal{D}^{-1} \mathbf{f})_i| > \|\mathcal{D}^{-1}\Pi_\Omega \mathbf{z}^\natural\|_\infty$ . Also, by the definition of the operator  $\Gamma_{\alpha p}$ , we know  $|(\mathcal{D}^{-1} \mathbf{f})_i| \leq \|\mathcal{D}^{-1}\Pi_\Omega \mathbf{z}^\natural\|_\infty$  for all  $i \in \Omega_s^\natural \setminus \Omega_s^{(0)}$ . Using these facts, we obtain

$$\left[ \mathcal{D}^{-1}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}) \right]_i = (\mathcal{D}^{-1} \mathbf{s}^\natural)_i \leq |(\mathcal{D}^{-1} \mathbf{f})_i| + |(\mathcal{D}^{-1} \mathbf{z}^\natural)_i| \leq 2 \|\mathcal{D}^{-1}\Pi_\Omega \mathbf{z}^\natural\|_\infty$$

for  $i \in \Omega_s^\natural \setminus \Omega_s^{(0)}$ . Combining all the pieces for  $\mathcal{D}^{-1}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)})$ , we obtain

$$(22) \quad \left\| \mathcal{D}^{-1}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}) \right\|_\infty \leq 2 \|\mathcal{D}^{-1}\Pi_\Omega \mathbf{z}^\natural\|_\infty.$$

Since  $\mathcal{G} = \mathcal{H}\mathcal{D}^{-1}$ , by the definition of  $\mathcal{H}$  we know  $\|\mathcal{G}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)})\|_\infty \leq 2\|\mathcal{G}\Pi_\Omega \mathbf{z}^\natural\|_\infty$ . It then yields that

$$(23) \quad \left\| \mathcal{G}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}) \right\|_2 \leq 2\alpha p n \left\| \mathcal{G}(\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}) \right\|_\infty \leq 4\alpha p n \|\mathcal{G}\mathbf{z}^\natural\|_\infty \leq 4\alpha p c_s \mu r \sigma_1^\natural,$$

where the first inequality follows from  $\|\Pi_\Omega \mathbf{s}^\natural - \mathbf{s}^{(0)}\|_0 \leq 2\alpha p n$  and Lemma 18. And the last inequality follows from  $\|\mathcal{G}\mathbf{z}^\natural\|_\infty \leq \|\mathbf{L}^\natural\|_{2,\infty} \|\mathcal{G}\mathbf{z}^\natural\|_2 \|\mathbf{R}^\natural\|_{2,\infty} \leq \mu \frac{c_s r}{n} \|\mathcal{G}\mathbf{z}^\natural\|_2$ . Finally, under event (16), we combine (21) and (23) to obtain

$$(24) \quad \left\| \mathbf{M}^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 \leq 4\alpha c_s \mu r \sigma_1^\natural + c_0 \sqrt{(p^{-1} \mu c_s r \log n)/n} \sigma_1^\natural \leq 4\alpha c_s \mu r \kappa \sigma_r^\natural + \frac{c_0 \varepsilon_0 \sigma_r^\natural}{\sqrt{\kappa r}},$$

provided  $p \geq (\varepsilon_0^{-2} \kappa^3 c_s \mu r^2 \log n)/n$

*Step 2.* The second step is to bound  $\tilde{d}_0$ . First we have

$$(25) \quad \begin{aligned} \left\| \mathbf{M}_r^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 &\leq \left\| \mathbf{M}_r^{(0)} - \mathbf{M}^{(0)} \right\|_2 + \left\| \mathbf{M}^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 \\ &\leq 2 \left\| \mathbf{M}^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 \leq 8\alpha c_s \mu r \kappa \sigma_r^\natural + 2c_0 \varepsilon_0 \sigma_r^\natural (\kappa r)^{-1/2}, \end{aligned}$$

where the second inequality follows from the definition of  $\mathbf{M}_r^{(0)}$  and the Eckart–Young–Mirsky theorem, and the third inequality follows from (24). Hence, for any  $\varepsilon_0 \in (0, \frac{\sqrt{\kappa r}}{8c_0})$ , as long as  $\alpha \leq \frac{1}{32c_s \mu r \kappa}$  one has

$$(26) \quad \left\| \mathbf{M}_r^{(0)} - \mathcal{G}\mathbf{z}^\natural \right\|_2 \leq \frac{1}{2} \sigma_r^\natural.$$



Then, by the definition of  $\tilde{d}_0$  and the inequality in [36, Lemma 5.14], we have

$$(27) \quad \tilde{d}_0^2 \leq \frac{2}{(\sqrt{2}-1)\sigma_r^{\natural}} \left\| \mathbf{M}_r^{(0)} - \mathcal{G}\mathbf{z}^{\natural} \right\|_{\mathbf{F}}^2 \leq \frac{10r}{\sigma_r^{\natural}} \left\| \mathbf{M}_r^{(0)} - \mathcal{G}\mathbf{z}^{\natural} \right\|_2^2,$$

which together with (25) reveals

$$\tilde{d}_0 \leq 26\alpha c_s \kappa \mu r \sqrt{r} \sqrt{\sigma_r^{\natural}} + 7c_0 \varepsilon_0 \sqrt{\sigma_r^{\natural}/\kappa}.$$

*Step 3.* Now we present the final step, which is to show  $d_0 \leq \tilde{d}_0$ . By (26) and Weyl's theorem [2], we obtain

$$(28) \quad \sqrt{\sigma_1^{\natural}/2} \leq \left\| \tilde{\mathbf{L}}^{(0)} \right\|_2 \leq \sqrt{3\sigma_1^{\natural}/2}, \quad \sqrt{\sigma_1^{\natural}/2} \leq \left\| \tilde{\mathbf{R}}^{(0)} \right\|_2 \leq \sqrt{3\sigma_1^{\natural}/2}.$$

Together with the  $\mu$ -incoherence of  $\mathcal{G}\mathbf{z}^{\natural}$ , it yields  $(\mathbf{L}^{\natural}, \mathbf{R}^{\natural}) \in \mathcal{L} \times \mathcal{R}$ . Let  $\tilde{\mathbf{Q}}$  be the best aligned matrix between  $(\tilde{\mathbf{L}}^{(0)}, \tilde{\mathbf{R}}^{(0)})$  and  $(\mathbf{L}^{\natural}, \mathbf{R}^{\natural})$ . We then have

$$(29) \quad \begin{aligned} \text{dist}^2 \left( \mathbf{L}^{(0)}, \mathbf{R}^{(0)}; \mathbf{L}^{\natural}, \mathbf{R}^{\natural} \right) &\leq \left\| \Pi_{\mathcal{L}} \left( \tilde{\mathbf{L}}^{(0)} \right) - \mathbf{L}^{\natural} \tilde{\mathbf{Q}} \right\|_{\mathbf{F}}^2 + \left\| \Pi_{\mathcal{R}} \left( \tilde{\mathbf{R}}^{(0)} \right) - \mathbf{Z} \mathbf{V}^{\natural} \tilde{\mathbf{Q}} \right\|_{\mathbf{F}}^2 \\ &\leq \left\| \tilde{\mathbf{L}}^{(0)} - \mathbf{L}^{\natural} \tilde{\mathbf{Q}} \right\|_{\mathbf{F}}^2 + \left\| \tilde{\mathbf{R}}^{(0)} - \mathbf{R}^{\natural} \tilde{\mathbf{Q}} \right\|_{\mathbf{F}}^2 \\ &= \text{dist}^2 \left( \tilde{\mathbf{L}}^{(0)}, \tilde{\mathbf{R}}^{(0)}; \mathbf{L}^{\natural}, \mathbf{R}^{\natural} \right), \end{aligned}$$

where the second inequality comes from the nonexpansion property of projections onto the convex sets  $\mathcal{L}$  and  $\mathcal{R}$ . Finally, (29) implies  $d_0 \leq \tilde{d}_0$ .  $\square$

**A.3. Proof of Theorem 4 (local convergence).** First, we present some key lemmas that are essential for the proof of local convergence. Some parts of the proofs in this section follow similar techniques introduced in [10, 39]. We begin with the following definition.

**DEFINITION 13.** Let  $(\mathbf{L}, \mathbf{R})$  be arbitrary matrices in the space  $(\mathcal{L} \times \mathcal{R}) \cap \mathcal{B}(\sqrt{\sigma_1^{\natural}})$ .<sup>7</sup> Define  $\mathbf{s}$  as

$$\mathbf{s} := \Gamma_{\gamma\alpha p} \left( \Pi_{\Omega} \left( \mathbf{f} - \mathcal{G}^* (\mathbf{L}\mathbf{R}^*) \right) \right).$$

Define  $(\mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}})$  as the solution set which satisfies

$$(\mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}}) \in \arg \min_{(\hat{\mathbf{L}}, \hat{\mathbf{R}}) \in \mathcal{E}(\mathbf{L}^{\natural}, \mathbf{R}^{\natural})} \left\| \hat{\mathbf{L}} - \mathbf{L} \right\|_{\mathbf{F}}^2 + \left\| \hat{\mathbf{R}} - \mathbf{R} \right\|_{\mathbf{F}}^2.$$

Define  $\Delta_{\mathbf{L}}$ ,  $\Delta_{\mathbf{R}}$ , and  $\Delta$  as

$$\Delta_{\mathbf{L}} := \mathbf{L} - \mathbf{L}_{\mathcal{E}}, \quad \Delta_{\mathbf{R}} := \mathbf{R} - \mathbf{R}_{\mathcal{E}}, \quad \text{and } \Delta := \left\| \Delta_{\mathbf{L}} \right\|_{\mathbf{F}}^2 + \left\| \Delta_{\mathbf{R}} \right\|_{\mathbf{F}}^2.$$

Actually, in Definition 13,  $(\mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}})$  is aligned with some  $\mathbf{Q} \in \mathbb{Q}_r$  to be the solution that matches  $(\mathbf{L}, \mathbf{R})$  best. So the error between  $(\mathbf{L}, \mathbf{R})$  and the solution set is then defined by  $\Delta_{\mathbf{L}}$ ,  $\Delta_{\mathbf{R}}$ , and  $\Delta$ . Let the iteration sequence  $\{\mathbf{L}^{(k)}, \mathbf{R}^{(k)}\}_{k \geq 1}$  be generated by the gradient descent strategy described in (12). Denote  $\nabla \ell^{(k)} := \nabla \ell(\mathbf{L}^{(k)}, \mathbf{R}^{(k)}; \mathbf{s}^{(k+1)})$ . In fact, if there is a proper lower bound for the term

<sup>7</sup> $\mathcal{B}(\cdot)$  is the ball with the center  $(\mathbf{L}^{\natural}, \mathbf{R}^{\natural})$  and a radius defined by the distance in (13).

$\text{Re}(\langle \nabla_{\mathbf{L}} \ell^{(k)}, \Delta_{\mathbf{L}}^{(k)} \rangle + \langle \nabla_{\mathbf{R}} \ell^{(k)}, \Delta_{\mathbf{R}}^{(k)} \rangle)$  and also a proper upper bound for the term  $\|\nabla_{\mathbf{L}} \ell^{(k)}\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}} \ell^{(k)}\|_{\mathbb{F}}^2$ , we can then show the local convergence of HSGD (see (43) for details).

In the rest of this section, we give bounds for the above terms in Lemmas 16 and 17, and the proof of Theorem 4 follows. Some crucial lemmas have to be shown first for the bounds.

LEMMA 14. Let  $\Delta_{\mathbf{L}}, \Delta_{\mathbf{R}}, \mathbf{L}, \mathbf{R}, \mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}}$ , and  $\Delta$  be defined in Definition 13. Then, if provided  $p \geq (\varepsilon_0^{-2} c_4 c_s^2 \mu^2 r^2 \log n)/n$ ,  $\alpha \leq \frac{1}{32 c_s \mu r \kappa}$ ,  $\varepsilon_0 \in (0, \frac{\sqrt{\kappa r}}{8 c_0})$ , under events (16), (18), and (19) it holds that

$$(30) \quad \begin{aligned} \|\mathcal{P}_{\Omega}(\Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*)\|_{\mathbb{F}}^2 &\leq \frac{p}{4} \Delta^2 + 18 \varepsilon_0 p \sigma_1^{\natural} \Delta, \\ \|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 &\leq 4(1 + 10 \varepsilon_0) p \sigma_1^{\natural} \Delta + \frac{p}{2} \Delta^2. \end{aligned}$$

*Proof.* By Lemma 11, we have

$$\begin{aligned} &\|\mathcal{P}_{\Omega}(\Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*)\|_{\mathbb{F}}^2 \\ &\leq p \|\Delta_{\mathbf{L}}\|_{\mathbb{F}}^2 \|\Delta_{\mathbf{R}}\|_{\mathbb{F}}^2 + \sqrt{8 p n \log n} \|\Delta_{\mathbf{L}}\|_{\mathbb{F}} \|\Delta_{\mathbf{R}}\|_{\mathbb{F}} \|\Delta_{\mathbf{L}}\|_{2, \infty} \|\Delta_{\mathbf{R}}\|_{2, \infty} \\ &\leq p \|\Delta_{\mathbf{L}}\|_{\mathbb{F}}^2 \|\Delta_{\mathbf{R}}\|_{\mathbb{F}}^2 + 36 \varepsilon_0 p \sigma_1^{\natural} \|\Delta_{\mathbf{L}}\|_{\mathbb{F}} \|\Delta_{\mathbf{R}}\|_{\mathbb{F}} \leq \frac{p}{4} \Delta^2 + 18 \varepsilon_0 p \sigma_1^{\natural} \Delta, \end{aligned}$$

where the second inequality follows from  $p \geq (\varepsilon_0^{-2} c_s^2 \mu^2 r^2 \log n)/n$  and (by (28))

$$(31) \quad \|\Delta_{\mathbf{L}}\|_{2, \infty} \leq \|\mathbf{L}\|_{2, \infty} + \|\mathbf{L}_{\mathcal{E}}\|_{2, \infty} \leq 2 \sqrt{2 \mu r c_s n^{-1}} \|\tilde{\mathbf{L}}^{(0)}\|_2 \leq 2 \sqrt{3 \mu r c_s \sigma_1^{\natural} n^{-1}}.$$

Then, we have

$$\begin{aligned} &\|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 \\ &\leq 2 \|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* + \Delta_{\mathbf{L}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 + 2 \|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*)\|_{\mathbb{F}}^2 \\ &\leq 2(1 + \varepsilon_0) p \|\mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* + \Delta_{\mathbf{L}} \mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + \frac{p}{2} \Delta^2 + 36 \varepsilon_0 p \sigma_1^{\natural} \Delta \\ &\leq 4(1 + \varepsilon_0) p \left( \|\mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^*\|_{\mathbb{F}}^2 + \|\Delta_{\mathbf{L}} \mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 \right) + \frac{p}{2} \Delta^2 + 36 \varepsilon_0 p \sigma_1^{\natural} \Delta \\ (32) \quad &\leq 4(1 + 10 \varepsilon_0) p \sigma_1^{\natural} \Delta + \frac{p}{2} \Delta^2, \end{aligned}$$

where the second inequality follows from Lemma 12.  $\square$

LEMMA 15. Let  $\mathbf{L}, \mathbf{R}, \mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}}$ , and  $\Delta$  be defined in Definition 13. Then, under the same condition to Lemma 14, if  $|\Omega| \leq \alpha n$ , we have

$$\|\mathcal{P}_{\Omega}(\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 \leq 18 \alpha c_s \mu r \sigma_1^{\natural} \Delta.$$

*Proof.* By Definition 13 we have

$$\begin{aligned} |[\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*]_{i,j}| &= |[\mathbf{L}_{\mathcal{E}}(\mathbf{R}^* - \mathbf{R}_{\mathcal{E}}^*) + (\mathbf{L} - \mathbf{L}_{\mathcal{E}}) \mathbf{R}_{\mathcal{E}}^* + (\mathbf{L} - \mathbf{L}_{\mathcal{E}})(\mathbf{R}^* - \mathbf{R}_{\mathcal{E}}^*)]_{i,j}| \\ &\leq \|\mathbf{L}_{\mathcal{E}}\|_{2, \infty} \|\Delta_{\mathbf{R}}(j, :)\|_2 + \|\mathbf{R}_{\mathcal{E}}\|_{2, \infty} \|\Delta_{\mathbf{L}}(i, :)\|_2 \\ &\quad + \frac{1}{2} \left( \|\Delta_{\mathbf{L}}\|_{2, \infty} \|\Delta_{\mathbf{R}}(j, :)\|_2 + \|\Delta_{\mathbf{R}}\|_{2, \infty} \|\Delta_{\mathbf{L}}(i, :)\|_2 \right) \\ &\leq (1 + \sqrt{3}) \sqrt{c_s \mu r \sigma_1^{\natural} n^{-1}} \|\Delta_{\mathbf{R}}(j, :)\|_2 \\ &\quad + (1 + \sqrt{3}) \sqrt{c_s \mu r \sigma_1^{\natural} n^{-1}} \|\Delta_{\mathbf{L}}(i, :)\|_2, \end{aligned}$$

where the last inequality follows from (31). Denote  $\Phi := \{(i, j) : (i + j - 1) \in \Omega\}$ . The Hankel structure yields  $\text{card}(\Phi(i, :)) \leq \alpha n$ ,  $\text{card}(\Phi(:, j)) \leq \alpha n$ . Hence, we use the inequality  $(a + b)^2 \leq 2a^2 + 2b^2$  and the fact  $\sum_{(i,j) \in \Phi} \|\Delta_{\mathbf{R}}(j, :)\|_2^2 = \sum_{i \in \Phi(:, j)} \sum_j \|\Delta_{\mathbf{R}}(j, :)\|_2^2$  to obtain

$$\begin{aligned} \|\mathcal{P}_{\Omega}(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*)\|_{\text{F}}^2 &\leq \sum_{(i,j) \in \Phi} |[\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*]_{i,j}^2| \\ &\leq \sum_{(i,j) \in \Phi} 18c_s \mu r \sigma_1^{\natural} n^{-1} \left( \|\Delta_{\mathbf{R}}(j, :)\|_2^2 + \|\Delta_{\mathbf{L}}(i, :)\|_2^2 \right) \leq 18\alpha c_s \mu r \sigma_1^{\natural} \Delta. \quad \square \end{aligned}$$

With Lemmas 12, 14, and 15 in hand, we can give the bounds in the following Lemmas 16 and 17 that lead to the local descent property of  $\ell$ .

LEMMA 16. Let  $\Delta_{\mathbf{L}}, \Delta_{\mathbf{R}}, \mathbf{L}, \mathbf{R}, \mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}}$ , and  $\Delta$  be defined in Definition 13. Set  $\lambda = \frac{1}{16}$ . Let  $\gamma \in [1 + \frac{1}{b_0}, 2]$  with any given  $1 \leq b_0 < \infty$ . If provided  $p \geq (\varepsilon_0^{-2} c_4 c_s^2 \mu^2 r^2 \log n)/n$ ,  $\alpha \leq \frac{1}{32c_s \mu r \kappa}$ , then for any  $\varepsilon_0 \in (0, \frac{1}{10}) \cap (0, \frac{\sqrt{\kappa r}}{8c_0})$ , under events (16), (18), and (19) we have

$$\begin{aligned} &\text{Re}(\langle \nabla_{\mathbf{L}} \ell, \mathbf{L} - \mathbf{L}_{\mathcal{E}} \rangle + \langle \nabla_{\mathbf{R}} \ell, \mathbf{R} - \mathbf{R}_{\mathcal{E}} \rangle) \\ &\geq \frac{7}{8} \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\text{F}}^2 + \left( \frac{1}{64} \sigma_r^{\natural} - \nu \sigma_1^{\natural} \right) \Delta - \left( 2 + \frac{\sqrt{2b_0}}{4} + \frac{b_0}{4\beta} \right) \Delta^2 \\ &\quad - \left( \sqrt{3b_0} + 6\sqrt{c_s \alpha \mu r} + \frac{\sqrt{2}}{16} \right) \sqrt{\sigma_1^{\natural} \Delta^3} + \frac{1}{64} \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_{\text{F}}^2, \end{aligned}$$

where  $\nu := (54 + 9\beta)\alpha \mu r + 4b_0\beta^{-1} + \varepsilon_1$  with  $\varepsilon_1 := \sqrt{\varepsilon_0}(12\sqrt{b_0} + 44\sqrt{c_s \alpha \mu r}) + 41\varepsilon_0$  and arbitrary  $\beta > 0$ .

*Proof.* The proof is divided into two main parts. In the first part of the proof, we establish a lower bound of  $\text{Re}\langle \nabla_{\mathbf{L}} \psi, \mathbf{L} - \mathbf{L}_{\mathcal{E}} \rangle + \text{Re}\langle \nabla_{\mathbf{R}} \psi, \mathbf{R} - \mathbf{R}_{\mathcal{E}} \rangle$ . In fact,

$$\begin{aligned} \text{Re}\langle \nabla_{\mathbf{L}} \psi, \mathbf{L} - \mathbf{L}_{\mathcal{E}} \rangle &= \text{Re}\langle p^{-1} \mathcal{G} \Pi_{\Omega} (\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f}) + (\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*), \Delta_{\mathbf{L}}\mathbf{R}^* \rangle, \\ \text{Re}\langle \nabla_{\mathbf{R}} \psi, \mathbf{R} - \mathbf{R}_{\mathcal{E}} \rangle &= \text{Re}\langle p^{-1} \mathcal{G} \Pi_{\Omega} (\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f}) + (\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*), \mathbf{L}\Delta_{\mathbf{R}}^* \rangle. \end{aligned}$$

Notice  $(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*) = 0$  and  $\mathbf{z}^{\natural} = \mathcal{G}^*(\mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*)$ . Rearrangement gives

$$\begin{aligned} &\text{Re}(\langle \nabla_{\mathbf{L}} \psi, \mathbf{L} - \mathbf{L}_{\mathcal{E}} \rangle + \langle \nabla_{\mathbf{R}} \psi, \mathbf{R} - \mathbf{R}_{\mathcal{E}} \rangle) \\ &= \underbrace{\text{Re}\langle \mathcal{G}(p^{-1} \Pi_{\Omega} - \mathcal{I}) \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*), \Delta_{\mathbf{L}}\mathbf{R}^* + \mathbf{L}\Delta_{\mathbf{R}}^* \rangle}_{T_1} \\ &\quad + \underbrace{\text{Re}\langle \mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*, \Delta_{\mathbf{L}}\mathbf{R}^* + \mathbf{L}\Delta_{\mathbf{R}}^* \rangle}_{T_2} + \underbrace{\text{Re}\langle p^{-1} \mathcal{G} \Pi_{\Omega} (\mathbf{s} - \mathbf{s}^{\natural}), \Delta_{\mathbf{L}}\mathbf{R}^* + \mathbf{L}\Delta_{\mathbf{R}}^* \rangle}_{T_3}. \end{aligned}$$

Now we estimate the bounds of  $T_1, T_2, T_3$ . By direct calculation, we have

$$(33) \quad \Delta_{\mathbf{L}}\mathbf{R}^* + \mathbf{L}\Delta_{\mathbf{R}}^* = \mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^* + \Delta_{\mathbf{L}}\Delta_{\mathbf{R}}^* = \Delta_{\mathbf{L}}\mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}}\Delta_{\mathbf{R}}^* + 2\Delta_{\mathbf{L}}\Delta_{\mathbf{R}}^*.$$

Then, for  $T_1$  we have

$$\begin{aligned}
 |T_1| &= \left| \operatorname{Re} \left\langle \mathcal{G} \left( p^{-1} \Pi_{\Omega} - \mathcal{I} \right) \mathcal{G}^* \left( \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* + \Delta_L \Delta_{\mathbf{R}}^* \right), \right. \right. \\
 &\quad \left. \left. \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* + 2 \Delta_L \Delta_{\mathbf{R}}^* \right\rangle \right| \\
 &\leq \left| \operatorname{Re} \left\langle \mathcal{G} \left( p^{-1} \Pi_{\Omega} - \mathcal{I} \right) \mathcal{G}^* \left( \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right), \left( \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right) \right\rangle \right| \\
 &\quad + 2 \left| \operatorname{Re} \left\langle \mathcal{G} \left( p^{-1} \Pi_{\Omega} - \mathcal{I} \right) \mathcal{G}^* \left( \Delta_L \Delta_{\mathbf{R}}^* \right), \Delta_L \Delta_{\mathbf{R}}^* \right\rangle \right| \\
 &\quad + 3 \left\| \mathcal{G} \left( p^{-1} \Pi_{\Omega} - \mathcal{I} \right) \mathcal{G}^* \left( \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right) \right\|_{\mathbf{F}} \left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}.
 \end{aligned}$$

Notice that  $\mathcal{P}_T(\Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^*) = \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^*$ , thus

$$\begin{aligned}
 |T_1| &\leq \varepsilon_0 \left\| \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 + 2p^{-1} \left\| \mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^* \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 - 2 \left\| \mathcal{G} \mathcal{G}^* \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 \\
 &\quad + \frac{3\varepsilon_0}{2} \left\| \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 + \frac{3}{2} \left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 \\
 &\leq \frac{5\varepsilon_0}{2} \left\| \Delta_L \mathbf{R}_{\mathcal{E}}^* + \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 + \frac{1}{2} \Delta^2 + 36\varepsilon_0 \sigma_1^{\natural} \Delta + \frac{3}{2} \left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 \\
 (34) \quad &\leq 41\varepsilon_0 \sigma_1^{\natural} \Delta + \frac{7}{8} \Delta^2,
 \end{aligned}$$

where the first inequality follows from Lemma 12 and  $ab \leq \frac{a^2+b^2}{2}$ , and the second inequality follows from  $\left\| \mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^* \mathbf{X} \right\|_{\mathbf{F}} = \left\| \mathcal{G} \mathcal{G}^* \mathcal{P}_{\Omega} \mathbf{X} \right\|_{\mathbf{F}}$  for  $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ ; then apply Lemma 14, and the fact  $\mathcal{G} \mathcal{G}^*$  is a projection operator. The last inequality follows from  $\left\| \Delta_L \mathbf{R}_{\mathcal{E}}^* \right\|_{\mathbf{F}}^2 \leq \sigma_1^{\natural} \left\| \Delta_L \right\|_{\mathbf{F}}^2$ ,  $\left\| \mathbf{L}_{\mathcal{E}} \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 \leq \sigma_1^{\natural} \left\| \Delta_{\mathbf{R}} \right\|_{\mathbf{F}}^2$ , and the Cauchy-Schwarz inequality which yields  $\left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}} \leq \left\| \Delta_L \right\|_{\mathbf{F}} \left\| \Delta_{\mathbf{R}} \right\|_{\mathbf{F}} \leq \frac{\Delta}{2}$ .

About  $T_2$ , by  $\Delta_L \mathbf{R}^* + \mathbf{L} \Delta_{\mathbf{R}}^* = \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* + \Delta_L \Delta_{\mathbf{R}}^*$  we have

$$\begin{aligned}
 T_2 &\geq \left\| \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\|_{\mathbf{F}}^2 - \left\| \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\|_{\mathbf{F}} \left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}} \\
 (35) \quad &\geq \frac{15}{16} \left\| \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\|_{\mathbf{F}}^2 - 4 \left\| \Delta_L \Delta_{\mathbf{R}}^* \right\|_{\mathbf{F}}^2 \geq \frac{15}{16} \left\| \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\|_{\mathbf{F}}^2 - \Delta^2,
 \end{aligned}$$

where the second inequality follows from  $a^2 - ab \geq \frac{15}{16} a^2 - 4b^2$ .

About  $T_3$ , we denote the support of  $\mathbf{s}$  and  $\mathbf{s}^{\natural}$  by  $\Omega_s$  and  $\Omega_{\natural}$ , respectively. We have

$$\begin{aligned}
 p|T_3| &\leq \left| \operatorname{Re} \left\langle \mathcal{G} \Pi_{\Omega} (\mathbf{s} - \mathbf{s}^{\natural}), \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\rangle \right| + \left| \operatorname{Re} \left\langle \mathcal{G} \Pi_{\Omega} (\mathbf{s} - \mathbf{s}^{\natural}), \Delta_L \Delta_{\mathbf{R}}^* \right\rangle \right| \\
 &\leq \underbrace{\left| \operatorname{Re} \left\langle \mathcal{P}_{\Omega} \mathcal{G} \Pi_{\Omega_s} (\mathbf{s} - \mathbf{s}^{\natural}), \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\rangle \right|}_{T_4} \\
 &\quad + \underbrace{\left| \operatorname{Re} \left\langle \mathcal{P}_{\Omega} \mathcal{G} \Pi_{\Omega_{\natural} \setminus \Omega_s} (\mathbf{s} - \mathbf{s}^{\natural}), \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\rangle \right|}_{T_5} + \underbrace{\left| \operatorname{Re} \left\langle \mathcal{G} \Pi_{\Omega} (\mathbf{s} - \mathbf{s}^{\natural}), \Delta_L \Delta_{\mathbf{R}}^* \right\rangle \right|}_{T_6}.
 \end{aligned}$$

Noticing that for  $i \in \Omega \cap \Omega_s$ , we have  $s_i = z_i^{\natural} + s_i^{\natural} - [\mathcal{G}^*(\mathbf{L} \mathbf{R}^*)]_i$ , then

$$\begin{aligned}
 T_4 &= \left| \operatorname{Re} \left\langle \mathcal{P}_{\Omega} \mathcal{P}_{\Omega_s} [\mathcal{G} \mathbf{z}^{\natural} - \mathcal{G} \mathcal{G}^* (\mathbf{L} \mathbf{R}^*)], \mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* \right\rangle \right| \\
 &= \left| \operatorname{Re} \left\langle \mathcal{G} \mathcal{G}^* (\mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* - \mathbf{L} \mathbf{R}^*), \mathcal{P}_{\Omega_s} (\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*) \right\rangle \right| \\
 &= \left| \operatorname{Re} \left\langle (\mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* - \mathbf{L} \mathbf{R}^*), \mathcal{P}_{\Omega_s} \mathcal{G} \mathcal{G}^* \mathcal{P}_{\Omega_s} (\mathbf{L} \mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*) \right\rangle \right| \\
 &= \left\| \mathcal{G} \mathcal{G}^* \mathcal{P}_{\Omega_s} (\mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* - \mathbf{L} \mathbf{R}^*) \right\|_{\mathbf{F}}^2 \leq \left\| \mathcal{P}_{\Omega_s} (\mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* - \mathbf{L} \mathbf{R}^*) \right\|_{\mathbf{F}}^2,
 \end{aligned}$$

where the second equality follows from  $\Omega_s \subseteq \Omega$ . The third equality follows from  $\mathcal{P}_{\Omega_s} \mathcal{G} \mathcal{G}^* \mathcal{P}_{\Omega_s} X = \mathcal{G} \mathcal{G}^* \mathcal{P}_{\Omega_s} X$  for  $X \in \mathbb{C}^{n_1 \times n_2}$ , which comes from the definition of  $\mathcal{G}$  and  $\mathcal{G}^*$ . The inequality in the last line follows from the fact that  $\mathcal{G} \mathcal{G}^*$  is a projection operator. By the definition of  $\mathcal{P}_{\Omega_s}$  and  $\Omega_s$ , we know  $\mathcal{P}_{\Omega_s}(\mathbf{M})$  has at most  $\gamma \alpha p n$  nonzero elements each row and each column. Then by Lemma 15, we have

$$(36) \quad T_4 \leq 18c_s \gamma \alpha p \mu \sigma_1^{\frac{1}{2}} \Delta.$$

About  $T_5$ , by the definition of  $\Gamma_{\gamma \alpha p}$ , we know for  $i \in \Omega \cap (\Omega_{\mathfrak{h}} \setminus \Omega_s)$ , the term  $z_i^{\mathfrak{h}} + s_i^{\mathfrak{h}} - [\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i$  is smaller than the  $\gamma \alpha p n$ -th largest element in  $\Pi_{\Omega}(z^{\mathfrak{h}} + \mathbf{s}^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*))$ , then it is smaller than the  $(\gamma \alpha p n - \alpha p n)$ -th largest element in  $\Pi_{\Omega}(z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*))$  as  $\|\Pi_{\Omega} \mathbf{s}^{\mathfrak{h}}\|_0 \leq \alpha p n$ . It then implies for  $i \in \Omega \cap (\Omega_{\mathfrak{h}} \setminus \Omega_s)$  that we have

$$(37) \quad |z_i^{\mathfrak{h}} + s_i^{\mathfrak{h}} - [\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i|^2 \leq \frac{\|\Pi_{\Omega}(z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*))\|_2^2}{\gamma \alpha p n - \alpha p n}.$$

Thus, for  $i \in \Omega \cap (\Omega_{\mathfrak{h}} \setminus \Omega_s)$ , it holds that

$$\begin{aligned} |s_i^{\mathfrak{h}} \overline{[z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i}| &= |[\mathbf{s}^{\mathfrak{h}} + z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) - z^{\mathfrak{h}}]_i \overline{[z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i}| \\ &\leq |[\mathbf{s}^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i|^2 + |[\mathbf{s}^{\mathfrak{h}} + z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i| |[\overline{z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)}]_i| \\ &\leq \left(1 + \frac{\beta}{2}\right) |z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)|_i|^2 + \frac{1}{2\beta} |z_i^{\mathfrak{h}} + s_i^{\mathfrak{h}} - [\mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]_i|^2 \\ &\leq \left(1 + \frac{\beta}{2}\right) |z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)|_i|^2 + \frac{\|\Pi_{\Omega}[z^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]\|_2^2}{2\beta(\gamma - 1)\alpha p n} \end{aligned}$$

for any  $\beta > 0$ , where the second inequality follows from  $a^2 + ab \leq (\frac{\beta}{2} + 1)a^2 + \frac{b^2}{2\beta}$ . Denote  $\Omega' := \Omega_{\mathfrak{h}} \setminus \Omega_s$ . By the above inequality we obtain

$$\begin{aligned} T_5 &= \left| \operatorname{Re} \langle \Pi_{\Omega'}(\mathbf{0} - \mathbf{s}^{\mathfrak{h}}), \Pi_{\Omega}[\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) - z^{\mathfrak{h}}] \rangle \right| \\ &\leq \left(1 + \frac{\beta}{2}\right) \|\mathcal{P}_{\Omega} \mathcal{P}_{\Omega'} \mathcal{G} \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbf{F}}^2 + \sum_{i \in \Omega'} \frac{\|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbf{F}}^2}{2\beta(\gamma - 1)\alpha p n} \\ (38) \quad &\leq (18 + 9\beta) c_s \alpha p \mu \sigma_1^{\frac{1}{2}} \Delta + \frac{4(1 + 10\varepsilon_0) p \sigma_1^{\frac{1}{2}} \Delta + \frac{p}{2} \Delta^2}{2\beta(\gamma - 1)}, \end{aligned}$$

where the first inequality follows from  $\|\mathcal{G} \mathbf{z}\|_{\mathbf{F}}^2 = \|\sum_{i=1}^n z_i / \sqrt{\varsigma_i} \mathcal{H} e_i\|_{\mathbf{F}}^2 = \|\mathbf{z}\|_2^2$ ,  $\mathbf{z} \in \mathbb{C}^n$  and thus  $\|\mathcal{G}^* \mathbf{Z}\|_2 = \|\mathcal{G} \mathcal{G}^* \mathbf{Z}\|_{\mathbf{F}}$ ,  $\mathbf{Z} \in \mathbb{C}^{n_1 \times n_2}$ . The second inequality follows from the same argument to (36),  $|\Omega'| \leq \alpha p n$ , and Lemma 14.

To obtain the upper bound of  $T_6$ , the strategy is similar to  $T_4$  and  $T_5$ . We have

$$\begin{aligned} T_6 &\leq \left| \operatorname{Re} \langle \mathcal{P}_{\Omega} \mathcal{G} \Pi_{\Omega_s}(\mathbf{s} - \mathbf{s}^{\mathfrak{h}}), \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^* \rangle \right| + \left| \operatorname{Re} \langle \mathcal{P}_{\Omega} \mathcal{G} \Pi_{\Omega_{\mathfrak{h}} \setminus \Omega_s}(\mathbf{s} - \mathbf{s}^{\mathfrak{h}}), \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^* \rangle \right| \\ &\leq \left| \operatorname{Re} \langle \mathcal{G} \Pi_{\Omega_s}(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) - z^{\mathfrak{h}}), \mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^* \rangle \right| + \left| \operatorname{Re} \langle \mathcal{G} \Pi_{\Omega_{\mathfrak{h}} \setminus \Omega_s} \mathbf{s}^{\mathfrak{h}}, \mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^* \rangle \right| \\ &\leq \|\mathcal{P}_{\Omega_s} \mathcal{G} \mathcal{G}^*(\mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^* - \mathbf{L}\mathbf{R}^*)\|_{\mathbf{F}} \|\mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*\|_{\mathbf{F}} \\ &\quad + \|\Pi_{\Omega_{\mathfrak{h}} \setminus \Omega_s} [z^{\mathfrak{h}} + \mathbf{s}^{\mathfrak{h}} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)]\|_{\mathbf{F}} \|\mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*\|_{\mathbf{F}} \\ &\quad + \|\Pi_{\Omega_{\mathfrak{h}} \setminus \Omega_s} (\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) - z^{\mathfrak{h}})\|_{\mathbf{F}} \|\mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*\|_{\mathbf{F}} \\ &\leq \left( (\sqrt{\gamma} + 1) \sqrt{18c_s \alpha p \mu \sigma_1^{\frac{1}{2}} \Delta} + \frac{\|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}} \mathbf{R}_{\mathcal{E}}^*)\|_{\mathbf{F}}}{\sqrt{\gamma - 1}} \right) \|\mathcal{P}_{\Omega} \Delta_{\mathbf{L}} \Delta_{\mathbf{R}}^*\|_{\mathbf{F}} \end{aligned}$$

$$\leq \left( (\sqrt{\gamma} + 1) \sqrt{18c_s \alpha p \mu r \sigma_1^{\natural} \Delta} + \sqrt{\frac{4(1 + 10\varepsilon_0) p \sigma_1^{\natural} \Delta + \frac{p}{2} \Delta^2}{\gamma - 1}} \right) \sqrt{\frac{p}{4} \Delta^2 + 18\varepsilon_0 p \sigma_1^{\natural} \Delta},$$

where the fourth inequality follows from (36), (37), and the equality  $\|\mathcal{G}^* Z\|_2 = \|\mathcal{G}\mathcal{G}^* Z\|_F$ . The fifth inequality follows from Lemma 14. Let  $\gamma$  be some constant satisfying  $1 + \frac{1}{b_0} \leq \gamma \leq 2$ , where  $1 \leq b_0 < \infty$ . Then, for all  $\varepsilon_0 \in (0, \frac{1}{10})$ , one has the upper bound of  $T_6$  given by

$$p^{-1} T_6 \leq \frac{\sqrt{2b_0}}{4} \Delta^2 + \left( \sqrt{3b_0} + 6\sqrt{c_s \alpha \mu r} \right) \sqrt{\sigma_1^{\natural} \Delta^3} + \sqrt{\varepsilon_0} \left( 12\sqrt{b_0} + 44\sqrt{c_s \alpha \mu r} \right) \sigma_1^{\natural} \Delta,$$

and thus

$$(39) \quad T_3 \leq [(54 + 9\beta) c_s \alpha \mu r + 4b_0 \beta^{-1} + \varepsilon_2] \sigma_1^{\natural} \Delta + \left( \frac{\sqrt{2b_0}}{4} + \frac{b_0}{4\beta} \right) \Delta^2 + \left( \sqrt{3b_0} + 6\sqrt{c_s \alpha \mu r} \right) \sqrt{\sigma_1^{\natural} \Delta^3},$$

where  $\varepsilon_2 = \sqrt{\varepsilon_0} (12\sqrt{b_0} + 44\sqrt{c_s \alpha \mu r})$ , and thus  $\varepsilon_1$  can be any small number in  $(0, \theta)$  for some  $\theta > 0$  when  $\alpha \mu r$  is bounded. Combining all the pieces, we obtain

$$\begin{aligned} T_1 + T_2 + T_3 &\geq T_2 - |T_1| - |T_3| \\ &\geq \frac{15}{16} \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_F^2 - \left( 2 + \frac{\sqrt{2b_0}}{4} + \frac{b_0}{4\beta} \right) \Delta^2 \\ &\quad - \nu \sigma_1^{\natural} \Delta - \left( \sqrt{3b_0} + 6\sqrt{c_s \alpha \mu r} \right) \sqrt{\sigma_1^{\natural} \Delta^3}, \end{aligned}$$

where we let  $\nu = (54 + 9\beta) c_s \alpha \mu r + 4b_0 \beta^{-1} + \varepsilon_1$  and  $\varepsilon_1 = \varepsilon_2 + 41\varepsilon_0$ . Now in the second part of the proof, we give a lower bound of  $\langle \nabla \phi, \mathbf{Z} - \mathbf{Z}^{\natural} \rangle$ , where  $\phi(\mathbf{L}, \mathbf{R}) = \frac{1}{4} \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_F^2$  is actually a standard regularization term. Therefore, by directly applying the result in [39, Lemma 3], we have

$$(40) \quad \begin{aligned} &\text{Re}(\langle \nabla_{\mathbf{L}} \phi, \Delta_{\mathbf{L}} \rangle + \langle \nabla_{\mathbf{R}} \phi, \Delta_{\mathbf{R}} \rangle) \\ &\geq \frac{1}{4} \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_F^2 + \frac{1}{4} \sigma_r^{\natural} \Delta - \sqrt{2\sigma_1^{\natural} \Delta^3} - \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_F^2, \end{aligned}$$

provided  $(\mathbf{L}, \mathbf{R}) \in \mathcal{B}(\sqrt{\sigma_r^{\natural}})$ . Finally, taking  $\lambda = \frac{1}{16}$ , combining (40) and the lower bound for  $T_1 + T_2 + T_3$ , we have

$$\begin{aligned} &\text{Re}(\langle \nabla_{\mathbf{L}} \ell, \Delta_{\mathbf{L}} \rangle + \langle \nabla_{\mathbf{R}} \ell, \Delta_{\mathbf{R}} \rangle) \\ &\geq \frac{7}{8} \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_F^2 + \frac{1}{64} \sigma_r^{\natural} \Delta - \nu \sigma_1^{\natural} \Delta - \left( 2 + \frac{\sqrt{2b_0}}{4} + \frac{b_0}{4\beta} \right) \Delta^2 \\ &\quad - \left( \sqrt{3b_0} + 6\sqrt{c_s \alpha \mu r} + \frac{\sqrt{2}}{16} \right) \sqrt{\sigma_1^{\natural} \Delta^3} + \frac{1}{64} \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_F^2. \quad \square \end{aligned}$$

LEMMA 17. Let  $\Delta_{\mathbf{L}}, \Delta_{\mathbf{R}}, \mathbf{L}, \mathbf{R}, \mathbf{L}_{\mathcal{E}}, \mathbf{R}_{\mathcal{E}}$ , and  $\Delta$  be defined in Definition 13. Set  $\lambda = \frac{1}{16}$ . Let  $\gamma \in [1 + \frac{1}{b_0}, 2]$ . Then under events (16), (18), and (19), for any  $\varepsilon_0 \in$

$(0, \frac{1}{10}) \cap (0, \frac{\sqrt{\kappa r}}{8c_0})$ , if  $p \geq (\varepsilon_0^{-2} c_4 c_s^2 \mu^2 r^2 \log n)/n$ ,  $\alpha \leq \frac{1}{32c_s \mu r \kappa}$ , it holds that

$$\begin{aligned} & \|\nabla_{\mathbf{L}} \ell(\mathbf{L}, \mathbf{R}; \mathbf{s})\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}} \ell(\mathbf{L}, \mathbf{R}; \mathbf{s})\|_{\mathbb{F}}^2 \\ & \leq 32\sigma_1^{\frac{1}{2}} \left( (9b_0 + 10) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + 72(b_0 + 1)\varepsilon_0\sigma_1^{\frac{1}{2}}\Delta + \frac{11}{2}(b_0 + 1)\Delta^2 \right) \\ & \quad + \frac{\sigma_1^{\frac{1}{2}}}{16} \|\mathbf{L}^*\mathbf{L} - \mathbf{R}^*\mathbf{R}\|_{\mathbb{F}}^2. \end{aligned}$$

*Proof.* To bound  $\|\nabla_{\mathbf{L}} \ell\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}} \ell\|_{\mathbb{F}}^2 = \|\nabla_{\mathbf{L}} \psi + \lambda \nabla_{\mathbf{L}} \phi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}} \psi + \lambda \nabla_{\mathbf{R}} \phi\|_{\mathbb{F}}^2$ , we first bound  $\|\nabla_{\mathbf{L}} \psi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}} \psi\|_{\mathbb{F}}^2$ . And we have

$$\begin{aligned} \|\nabla_{\mathbf{L}} \psi\|_{\mathbb{F}}^2 &= \left\| \left( \frac{1}{p} \mathcal{G} \Pi_{\Omega} [\mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*) - \mathbf{z}^{\natural} + \mathbf{s} - \mathbf{s}^{\natural}] + (\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*) \right) \mathbf{R} \right\|_{\mathbb{F}}^2 \\ &\leq 2 \|\mathbf{R}\|_2^2 \left( \underbrace{\|p^{-1} \mathcal{G} \Pi_{\Omega} [\mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*) - \mathbf{z}^{\natural} + \mathbf{s} - \mathbf{s}^{\natural}]\|_{\mathbb{F}}^2}_{T_7} + \underbrace{\|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2}_{T_8} \right). \end{aligned}$$

Then we bound  $T_7$  and  $T_8$ . Let  $\Omega_s$  and  $\Omega_s^{\natural}$  be the support of  $\mathbf{s}$  and  $\Pi_{\Omega} \mathbf{s}^{\natural}$ ; then we have

$$\text{for } i \in \Omega \cap \Omega_s, \quad \mathbf{s}_i = [\mathbf{z}^{\natural} + \mathbf{s}^{\natural} - \mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*)]_i \Rightarrow [\mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*) - \mathbf{z}^{\natural} + \mathbf{s} - \mathbf{s}^{\natural}]_i = 0,$$

$$\text{for } i \in \Omega \cap (\Omega_s^{\natural} \setminus \Omega_s), \quad [\mathbf{z}^{\natural} + \mathbf{s}^{\natural} - \mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*)]_i^2 \leq \frac{\|\Pi_{\Omega}(\mathbf{z}^{\natural} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*))\|_2^2}{\gamma \alpha p n - \alpha p n},$$

$$\text{for other } i \text{ in } \Omega, \text{ we have } [\mathbf{z}^{\natural} + \mathbf{s}^{\natural} - \mathbf{s} - \mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*)]_i = [\mathbf{z}^{\natural} - \mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*)]_i,$$

where the equality in the first line follows from the definition of  $\mathbf{s}$  in Definition 13, and the second line follows from (37). Thus we have

$$\begin{aligned} T_7 &\leq 2 \left\| p^{-1} \Pi_{\Omega \cap (\Omega_s^{\natural} \setminus \Omega_s)} [\mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*) - \mathbf{z}^{\natural} + \mathbf{s} - \mathbf{s}^{\natural}] \right\|_2^2 + 2 \left\| p^{-1} \mathcal{G} \Pi_{\Omega} [\mathcal{G}^*(\mathbf{L}\mathbf{R}_{\mathcal{E}}^*) - \mathbf{z}^{\natural}] \right\|_{\mathbb{F}}^2 \\ &\leq 2\alpha p n \frac{\|p^{-1} \Pi_{\Omega}(\mathbf{z}^{\natural} - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*))\|_2^2}{\gamma \alpha p n - \alpha p n} + 2p^{-1} \|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 \\ &\leq p^{-1} \frac{2\gamma}{\gamma - 1} \|\mathcal{P}_{\Omega} \mathcal{G} \mathcal{G}^*(\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*)\|_{\mathbb{F}}^2 \\ &\leq \frac{4\gamma}{\gamma - 1} (1 + \varepsilon_0) \|\mathbf{L}_{\mathcal{E}} \mathbf{\Delta} \mathbf{R}^* + \mathbf{\Delta} \mathbf{L} \mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + \frac{\gamma}{\gamma - 1} \Delta^2 + \frac{72\gamma}{\gamma - 1} \varepsilon_0 \sigma_1^{\frac{1}{2}} \Delta \\ &\leq \frac{8\gamma}{\gamma - 1} (1 + \varepsilon_0) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + \frac{(5 + 4\varepsilon_0)\gamma}{\gamma - 1} \Delta^2 + \frac{72\gamma}{\gamma - 1} \varepsilon_0 \sigma_1^{\frac{1}{2}} \Delta, \end{aligned}$$

where the first to third inequalities follow from  $\|\mathcal{G}\mathbf{z}\|_{\mathbb{F}}^2 = \|\mathbf{z}\|_2^2$ , the fourth inequality follows from the third line in (32), and the last inequality follows from (33). Since  $\mathcal{I} - \mathcal{G}\mathcal{G}^*$  is a projection, we have

$$T_8 \leq \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2.$$

Thus, as  $\gamma \in [1 + \frac{1}{b_0}, 2]$  with  $b_0 \geq 1$ , by setting  $\varepsilon_0 \in (0, \frac{1}{10})$ , combining  $T_7$  and  $T_8$  gives

$$\|\nabla_{\mathbf{L}} \psi\|_{\mathbb{F}}^2 \leq 2 \|\mathbf{R}\|_2^2 \left( (9b_0 + 10) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + 72(b_0 + 1)\varepsilon_0\sigma_1^{\frac{1}{2}}\Delta + \frac{11b_0 + 22}{2}\Delta^2 \right).$$

One can follow the same process to bound  $\|\nabla_{\mathbf{R}}\psi\|_{\mathbb{F}}^2$ , and thus we have

$$(41) \quad \begin{aligned} & \|\nabla_{\mathbf{L}}\psi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\psi\|_{\mathbb{F}}^2 \\ & \leq 16\sigma_1^{\frac{1}{2}} \left( (9b_0 + 10) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + 72(b_0 + 1)\varepsilon_0\sigma_1^{\frac{1}{2}}\Delta + \frac{11}{2}(b_0 + 1)\Delta^2 \right), \end{aligned}$$

where the inequality follows from  $\|\mathbf{L}\|_2 \leq 2\sqrt{\sigma_1^{\frac{1}{2}}}$ ,  $\|\mathbf{R}\|_2 \leq 2\sqrt{\sigma_1^{\frac{1}{2}}}$  as  $(\mathbf{L}, \mathbf{R}) \in \mathcal{B}(\sqrt{\sigma_1^{\frac{1}{2}}})$ . Now we bound  $\|\nabla_{\mathbf{L}}\phi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\phi\|_{\mathbb{F}}^2$ . In fact, we have

$$(42) \quad \begin{aligned} \|\nabla_{\mathbf{L}}\phi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\phi\|_{\mathbb{F}}^2 &= \|\mathbf{L}(\mathbf{L}^*\mathbf{L} - \mathbf{R}^*\mathbf{R})\|_{\mathbb{F}}^2 + \|\mathbf{R}(\mathbf{R}^*\mathbf{R} - \mathbf{L}^*\mathbf{L})\|_{\mathbb{F}}^2 \\ &\leq \left( \|\mathbf{L}\|_2^2 + \|\mathbf{R}\|_2^2 \right) \|\mathbf{L}^*\mathbf{L} - \mathbf{R}^*\mathbf{R}\|_{\mathbb{F}}^2 \leq 8\sigma_1^{\frac{1}{2}} \|\mathbf{L}^*\mathbf{L} - \mathbf{R}^*\mathbf{R}\|_{\mathbb{F}}^2. \end{aligned}$$

Finally, taking  $\lambda = \frac{1}{16}$  and combining (41) and (42) we obtain

$$\begin{aligned} & \|\nabla_{\mathbf{L}}\ell\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\ell\|_{\mathbb{F}}^2 \\ & \leq 2 \left( \|\nabla_{\mathbf{L}}\psi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\psi\|_{\mathbb{F}}^2 \right) + 2\lambda^2 \left( \|\nabla_{\mathbf{L}}\phi\|_{\mathbb{F}}^2 + \|\nabla_{\mathbf{R}}\phi\|_{\mathbb{F}}^2 \right) \\ & \leq 32\sigma_1^{\frac{1}{2}} \left( (9b_0 + 10) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 + 72(b_0 + 1)\varepsilon_0\sigma_1^{\frac{1}{2}}\Delta + \frac{11}{2}(b_0 + 1)\Delta^2 \right) \\ & \quad + \frac{\sigma_1^{\frac{1}{2}}}{16} \|\mathbf{L}^*\mathbf{L} - \mathbf{R}^*\mathbf{R}\|_{\mathbb{F}}^2. \end{aligned} \quad \square$$

As discussed above, with Lemmas 16 and 17, we are now ready to show the local descent property of the proposed method.

*Proof of Theorem 4.* The proof of the theorem is under events (18) and (19). For ease of notation, let  $(\mathbf{L}_{\mathcal{E}}^{(k)}, \mathbf{R}_{\mathcal{E}}^{(k)}) \in \mathcal{E}(\mathbf{L}^{\frac{1}{2}}, \mathbf{R}^{\frac{1}{2}})$  defined to be matrices aligned with  $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$ . Then, by the definition of  $d_k$  we have

$$(43) \quad \begin{aligned} d_{k+1}^2 &= \left\| \mathbf{L}^{(k+1)} - \mathbf{L}_{\mathcal{E}}^{(k+1)} \right\|_{\mathbb{F}}^2 + \left\| \mathbf{R}^{(k+1)} - \mathbf{R}_{\mathcal{E}}^{(k+1)} \right\|_{\mathbb{F}}^2 \\ &\leq \left\| \mathbf{L}^{(k+1)} - \mathbf{L}_{\mathcal{E}}^{(k)} \right\|_{\mathbb{F}}^2 + \left\| \mathbf{R}^{(k+1)} - \mathbf{R}_{\mathcal{E}}^{(k)} \right\|_{\mathbb{F}}^2 \\ &\leq \left\| \mathbf{L}^{(k)} - \eta \nabla_{\mathbf{L}}\ell^{(k)} - \mathbf{L}_{\mathcal{E}}^{(k)} \right\|_{\mathbb{F}}^2 + \left\| \mathbf{R}^{(k)} - \eta \nabla_{\mathbf{R}}\ell^{(k)} - \mathbf{R}_{\mathcal{E}}^{(k)} \right\|_{\mathbb{F}}^2 \\ &\leq d_k^2 - 2\eta \operatorname{Re} \left( \left\langle \nabla_{\mathbf{L}}\ell^{(k)}, \mathbf{L}^{(k)} - \mathbf{L}_{\mathcal{E}}^{(k)} \right\rangle + \left\langle \nabla_{\mathbf{R}}\ell^{(k)}, \mathbf{R}^{(k)} - \mathbf{R}_{\mathcal{E}}^{(k)} \right\rangle \right) \\ &\quad + \eta^2 \left( \left\| \nabla_{\mathbf{L}}\ell^{(k)} \right\|_{\mathbb{F}}^2 + \left\| \nabla_{\mathbf{R}}\ell^{(k)} \right\|_{\mathbb{F}}^2 \right), \end{aligned}$$

where the second inequality comes from the nonexpansion property of projection onto  $\mathcal{L}$  and  $\mathcal{R}$ . Let  $\Delta_k := d_k^2$ . Set  $\eta \leq \frac{\tilde{\theta}}{(b_0+1)\sigma_r^{\frac{1}{2}}}$  for sufficiently small constant  $\tilde{\theta}$  and  $\gamma_k \in [1 + \frac{1}{b_0}, 2]$  with  $b_0 \geq 1$ . By Lemmas 16 and 17, we then have

$$\begin{aligned} & -2\eta \operatorname{Re} \left( \left\langle \nabla_{\mathbf{L}}\ell^{(k)}, \mathbf{L}^{(k)} - \mathbf{L}_{\mathcal{E}}^{(k)} \right\rangle + \left\langle \nabla_{\mathbf{R}}\ell^{(k)}, \mathbf{R}^{(k)} - \mathbf{R}_{\mathcal{E}}^{(k)} \right\rangle \right) \\ & + \eta^2 \left( \left\| \nabla_{\mathbf{L}}\ell^{(k)} \right\|_{\mathbb{F}}^2 + \left\| \nabla_{\mathbf{R}}\ell^{(k)} \right\|_{\mathbb{F}}^2 \right) \\ & \leq \eta \left[ \left( -\frac{7}{4} + 10\tilde{\theta} \right) \|\mathbf{L}\mathbf{R}^* - \mathbf{L}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}^*\|_{\mathbb{F}}^2 - \left( \frac{1}{32}\sigma_r^{\frac{1}{2}} - 2\nu\sigma_1^{\frac{1}{2}} - 72\tilde{\theta}\varepsilon_0\sigma_1^{\frac{1}{2}} \right) \Delta_k \right] \end{aligned}$$



$$\begin{aligned}
& + \left( 4 + \frac{\sqrt{2b_0}}{2} + \frac{b_0}{2\beta} + \frac{11\tilde{\theta}}{2} \right) \Delta_k^2 + \left( \sqrt{12b_0} + 12\sqrt{c_s\alpha\mu r} + \frac{\sqrt{2}}{8} \right) \sqrt{\sigma_1^{\natural} \Delta_k^3} \\
& - \left( \frac{1}{32} - \frac{\tilde{\theta}}{64} \right) \|\mathbf{L}^* \mathbf{L} - \mathbf{R}^* \mathbf{R}\|_{\mathbb{F}}^2 \Big] \\
& \leq -\frac{1}{64} \eta \sigma_r^{\natural} \Delta_k,
\end{aligned}$$

where in the second inequality, we use  $\Delta_k \leq \frac{\theta_1}{b_0} \frac{\sigma_r^{\natural}}{\kappa}$  with  $\frac{\theta_1}{b_0} = c_1^2$  sufficiently small. Also, in the inequality we set  $\beta = \frac{\kappa}{\theta_2}$ ,  $\varepsilon_0 = \frac{\theta_3}{\kappa}$ ,  $\varepsilon_1 = \frac{\theta_4}{\kappa}$ , and assume  $\alpha \leq \frac{\theta_5}{c_s \mu r \kappa^2}$  for sufficiently small constants  $\theta_2, \theta_3, \theta_4, \theta_5$ . Then, it implies

$$d_{k+1}^2 \leq \left( 1 - \frac{\eta \sigma_r^{\natural}}{64} \right) d_k^2.$$

Note that under events (16), (18), and (19), the above inequality holds for all  $k \geq 0$ .  $\square$

## Appendix B. Supporting lemmas.

LEMMA 18 ([4, Lemma 6]). *For any  $\mathbf{z} \in \mathbb{C}^n$  such that  $\|\mathbf{z}\|_0 \leq \alpha n$ , it holds that*

$$\|\mathcal{H}\mathbf{z}\|_2 \leq \alpha n \|\mathcal{H}\mathbf{z}\|_{\infty} = \alpha n \|\mathbf{z}\|_{\infty}.$$

LEMMA 19. *For all  $\mathbf{u} \in \mathbb{R}^{n_1}$ ,  $\mathbf{v} \in \mathbb{R}^{n_2}$ , it holds that*

$$p^{-1} \sum_a \sum_{i+j=a+1} \delta_a \mathbf{u}_i \mathbf{v}_j \leq \|\mathbf{u}\|_1 \|\mathbf{v}\|_1 + \sqrt{8p^{-1}n \log n} \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$$

with probability at least  $1 - 2n^{-2}$ , provided  $p \geq (\log n)/n$ .

*Proof.* The result of this lemma is similar to [10, Lemma 5] except for the slightly different sampling models, and one can follow the same strategy to prove it.  $\square$

## REFERENCES

- [1] B. N. BHASKAR, G. TANG, AND B. RECHT, *Atomic norm denoising with applications to line spectral estimation*, IEEE Trans. Signal Process., 61 (2013), pp. 5987–5999.
- [2] R. BHATIA, *Matrix Analysis*, Grad. Texts Math. 169, Springer Science & Business Media, New York, 2013.
- [3] J. A. CADZOW, *Signal enhancement—a composite property mapping algorithm*, IEEE Trans. Acoustics Speech Signal Process., 36 (1988), pp. 49–62.
- [4] H. CAI, J.-F. CAI, T. WANG, AND G. YIN, *Accelerated structured alternating projections for robust spectrally sparse signal recovery*, IEEE Trans. Signal Process., 69 (2021), pp. 809–821.
- [5] H. CAI, J.-F. CAI, AND K. WEI, *Accelerated alternating projections for robust principal component analysis*, J. Mach. Learn. Res., 20 (2019), pp. 685–717.
- [6] H. CAI, Z. CHAO, L. HUANG, AND D. NEEDELL, *Fast robust tensor principal component analysis via fiber CUR decomposition*, in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2021, pp. 189–197.
- [7] H. CAI, K. HAMM, L. HUANG, J. LI, AND T. WANG, *Rapid robust principal component analysis: CUR accelerated inexact low rank estimation*, IEEE Signal Process. Lett., 28 (2021), pp. 116–120.
- [8] H. CAI, K. HAMM, L. HUANG, AND D. NEEDELL, *Robust CUR decomposition: Theory and imaging applications*, SIAM J. Imaging Sci., 14 (2021), pp. 1472–1503.
- [9] H. CAI, J. LIU, AND W. YIN, *Learned robust PCA: A scalable deep unfolding approach for high-dimensional outlier detection*, in Advances in Neural Information Processing Systems, Vol. 34, 2021, pp. 16977–16989.

- [10] J.-F. CAI, T. WANG, AND K. WEI, *Spectral compressed sensing via projected gradient descent*, SIAM J. Optim., 28 (2018), pp. 2625–2653.
- [11] J.-F. CAI, T. WANG, AND K. WEI, *Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion*, Appl. Comput. Harmon. Anal., 46 (2019), pp. 94–121.
- [12] E. J. CANDÈS, X. LI, Y. MA, AND J. WRIGHT, *Robust principal component analysis?*, J. ACM, 58 (2011), pp. 1–37.
- [13] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.
- [14] V. CHANDRASEKARAN, S. SANGHAVI, P. A. PARRILO, AND A. S. WILLSKY, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2011), pp. 572–596.
- [15] J. CHEN, W. GAO, AND K. WEI, *Exact matrix completion based on low rank Hankel structure in the Fourier domain*, Appl. Comput. Harmon. Anal., 55 (2021), pp. 149–184.
- [16] Y. CHEN AND Y. CHI, *Robust spectral compressed sensing via structured matrix completion*, IEEE Trans. Inform. Theory, 60 (2014), pp. 6576–6601.
- [17] Y. CHEN, A. JALALI, S. SANGHAVI, AND C. CARAMANIS, *Low-rank matrix recovery from errors and erasures*, IEEE Trans. Inform. Theory, 59 (2013), pp. 4324–4337.
- [18] Y. CHERAPANAMJERI, K. GUPTA, AND P. JAIN, *Nearly optimal robust matrix completion*, in Proceedings of the International Conference on Machine Learning, 2017, pp. 797–805.
- [19] Y. CHI, Y. M. LU, AND Y. CHEN, *Nonconvex optimization meets low-rank matrix factorization: An overview*, IEEE Trans. Signal Process., 67 (2019), pp. 5239–5269.
- [20] K. HAMM, M. MESKINI, AND H. CAI, *Riemannian CUR decompositions for robust principal component analysis*, in Topological, Algebraic and Geometric Learning Workshops 2022, PMLR, 2022, pp. 152–160.
- [21] D. J. HOLLAND, M. J. BOSTOCK, L. F. GLADDEN, AND D. NIETLISPACH, *Fast multidimensional NMR spectroscopy using compressed sensing*, Angew. Chem. Int. Ed., 50 (2011), pp. 6548–6551.
- [22] M. JACOB, M. P. MANI, AND J. C. YE, *Structured low-rank algorithms: Theory, magnetic resonance applications, and links to machine learning*, IEEE Signal Process. Mag., 37 (2020), pp. 54–68.
- [23] K. H. JIN, D. LEE, AND J. C. YE, *A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix*, IEEE Trans. Comput. Imaging, 2 (2016), pp. 480–495.
- [24] R. M. LARSEN, *PROPACK—Software for Large and Sparse SVD Calculations*. <http://sun.stanford.edu/~rmunk/PROPACK/>
- [25] B. LEE AND A. LAMPERSKI, *Non-asymptotic closed-loop system identification using autoregressive processes and Hankel model reduction*, in Proceedings of the 59th IEEE Conference on Decision and Control, 2020, pp. 3419–3424.
- [26] W. LIAO AND A. FANNJIANG, *Music for single-snapshot spectral estimation: Stability and super-resolution*, Appl. Comput. Harmon. Anal., 40 (2016), pp. 33–67.
- [27] R. A. LOBOS, R. M. LEAHY, AND J. P. HALDAR, *Autoregression and structured low-rank modeling of sinograms*, in Proceedings of the IEEE 17th International Symposium on Biomedical Imaging, 2020, pp. 1–4.
- [28] C. MA, Y. LI, AND Y. CHI, *Beyond Procrustes: Balancing-free gradient descent for asymmetric low-rank matrix sensing*, IEEE Trans. Signal Process., 69 (2021), pp. 867–877.
- [29] P. NETRAPALLI, U. NIRANJAN, S. SANGHAVI, A. ANANDKUMAR, AND P. JAIN, *Non-convex robust PCA*, in Advances in Neural Information Processing Systems, 2014, pp. 1107–1115.
- [30] H. M. NGUYEN, X. PENG, M. N. DO, AND Z.-P. LIANG, *Denoising MR spectroscopic imaging data with low-rank approximations*, IEEE Trans. Biomed. Eng., 60 (2013), pp. 78–89.
- [31] V. OROPEZA AND M. SACCHI, *Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis*, Geophysics, 76 (2011), pp. V25–V32.
- [32] X. QU, M. MAYZEL, J.-F. CAI, Z. CHEN, AND V. OREKHOV, *Accelerated NMR spectroscopy with low-rank reconstruction*, Angew. Chem. Int. Ed., 54 (2015), pp. 852–854.
- [33] G. TANG, B. N. BHASKAR, P. SHAH, AND B. RECHT, *Compressed sensing off the grid*, IEEE Trans. Inform. Theory, 59 (2013), pp. 7465–7490.
- [34] J. A. TROPP, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math., 12 (2012), pp. 389–434.
- [35] J. A. TROPP, J. N. LASKA, M. F. DUARTE, J. K. ROMBERG, AND R. G. BARANIUK, *Beyond Nyquist: Efficient sampling of sparse bandlimited signals*, IEEE Trans. Inform. Theory, 56 (2009), pp. 520–544.
- [36] S. TU, R. BOCZAR, M. SIMCHOWITZ, M. SOLTANOLKOTABI, AND B. RECHT, *Low-rank solutions of linear matrix equations via procrustes flow*, in Proceedings of the International Conference on Machine Learning, PMLR, 2016, pp. 964–973.

- [37] J. WRIGHT, A. Y. YANG, A. GANESH, S. S. SASTRY, AND Y. MA, *Robust face recognition via sparse representation*, IEEE Trans. Pattern Anal. Mach. Intell., 31 (2008), pp. 210–227.
- [38] Y. XI AND D. M. ROCKE, *Baseline correction for NMR spectroscopic metabolomics data analysis*, BMC Bioinform., 9 (2008), pp. 1–10.
- [39] X. YI, D. PARK, Y. CHEN, AND C. CARAMANIS, *Fast algorithms for robust PCA via gradient descent*, in Advances in Neural Information Processing Systems, 2016, pp. 4152–4160.
- [40] S. ZHANG AND M. WANG, *Correction of simultaneous bad measurements by exploiting the low-rank Hankel structure*, in Proceedings of the IEEE International Symposium on Information Theory, 2018, pp. 646–650.
- [41] S. ZHANG AND M. WANG, *Correction of corrupted columns through fast robust Hankel matrix completion*, IEEE Trans. Signal Process., 67 (2019), pp. 2580–2594.