

PAPER

Accelerating ill-conditioned Hankel matrix recovery via structured Newton-like descent



To cite this article: HanQin Cai *et al* 2025 *Inverse Problems* **41** 075015

View the [article online](#) for updates and enhancements.

You may also like

- [A fault feature extraction method for rotating shaft with multiple weak faults based on underdetermined blind source signal](#)
Hongchun Sun, Liang Fang and Jingzheng Guo
- [Fault diagnosis of rotating machinery in a noisy environment based on modified general normalized sparse filtering](#)
Lei Guo, Jinrui Wang, Shan Wang et al.
- [A new selection method of singular values using correlation coefficients for detecting weak exponential damped sinusoidal signals under strong noise](#)
Ting Zhang

Accelerating ill-conditioned Hankel matrix recovery via structured Newton-like descent

HanQin Cai¹ , Longxiu Huang² , Xiliang Lu³ 
and Juntao You^{4,5,*} 

¹ Department of Statistics and Data Science and Department of Computer Science, University of Central Florida, Orlando, FL 32816, United States of America

² Department of Computational Mathematics, Science and Engineering and Department of Mathematics, Michigan State University, East Lansing, MI 48823, United States of America

³ School of Mathematics and Statistics, and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, People's Republic of China

⁴ School of Artificial Intelligence, National Center for Applied Mathematics in Hubei, and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan, People's Republic of China

⁵ Institute for Advanced Study, Shenzhen University, Shenzhen 518000, People's Republic of China

E-mail: jyouab@connect.ust.hk, hqcai@ucf.edu, huangl3@msu.edu and xllv.math@whu.edu.cn

Received 17 December 2024; revised 2 June 2025

Accepted for publication 4 July 2025

Published 17 July 2025



CrossMark

Abstract

This paper studies the robust Hankel recovery problem, which simultaneously removes the sparse outliers and fulfills missing entries from the partial observation. We propose a novel non-convex algorithm, coined Hankel structured Newton-like descent (HSNLD), to tackle the robust Hankel recovery problem. HSNLD is highly efficient with linear convergence, and its convergence rate is independent of the condition number of the underlying Hankel matrix. The recovery guarantee has been established under some mild conditions. Numerical experiments on both synthetic and real datasets show the superior performance of HSNLD against state-of-the-art algorithms.

Keywords: ill-conditioned Hankel matrix, robust matrix completion, Newton-like descent, precondition, direction of arrival, nuclear magnetic resonance

* Author to whom any correspondence should be addressed.

1. Introduction

A Hankel matrix has identical complex values in each of its skew-diagonal, that is,

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n_2} \\ x_2 & x_3 & \cdots & x_{n_2+1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_1} & x_{n_1+1} & \cdots & x_n \end{bmatrix} \in \mathbb{C}^{n_1 \times n_2}.$$

Each column of the Hankel matrix can be viewed as a shifting window of size n_1 that moves from the beginning toward the end on a vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top \in \mathbb{C}^n$, where $n = n_1 + n_2 - 1$. Thus, the Hankel matrix can also be viewed as a convolution operator. Since it contains only n distinct values in total, one can simply keep track of the vector \mathbf{x} instead of the entire Hankel matrix for the sake of memory efficiency.

The low-rank Hankel matrix and its recovery problem have widely appeared in applications, such as nuclear magnetic resonance (NMR) spectroscopy [22, 23, 30], direction of arrival (DOA) estimation [32, 40], magnetic resonance imaging [24, 25], Fourier transform ion cyclotron resonance mass spectrometry [17], seismic analysis [11, 15, 37], system identification [18, 33], and time series forecasting [1, 12, 20, 34]. For instance, a Hankel matrix \mathbf{X} is rank- r if the corresponding vector \mathbf{x} has r -sparse representation in the Fourier domain:

$$x_t = \sum_{j=1}^r a_j e^{(2\pi i f_j - d_j)(t-1)}, \quad t = 1, 2, \dots, n, \quad (1)$$

where i is the imaginary unit, a_j , f_j and d_j are the complex amplitude, frequency and damping factor of the j th sinusoid, respectively. This low-rank property is guaranteed via an explicit Vandermonde decomposition [14, 28]:

$$\mathbf{X} = \mathbf{E}_L \mathbf{D} \mathbf{E}_R^\top,$$

where $\mathbf{D} = \text{diag}([a_1, \dots, a_r])$, $[\mathbf{E}_L]_{i_1 j} = (e^{2\pi i f_j - d_j})^{i_1 - 1}$ and $[\mathbf{E}_R]_{i_2 j} = (e^{2\pi i f_j - d_j})^{i_2 - 1}$ for $i_1 \in [n_1]$, $i_2 \in [n_2]$ and $j \in [r]$. Let us consider DOA signals and NMR signals as concrete examples, both of which are characterized by spectral sparsity. For DOA signals, suppose there are r narrowband far-field source signals $\{a_j\}_{j=1}^r \subseteq \mathbb{C}$ impinging on an array of omnidirectional sensors from distinct directions, as illustrated in figure 1. Denote the DOAs of these sources by $\{\theta_j\}_{j=1}^r \subseteq [-\frac{\pi}{2}, \frac{\pi}{2}]$. The noiseless received signal x_t at the t th sensor of a uniform linear array (ULA) with half-wavelength spacing between adjacent sensors is given by [40]:

$$x_t = \sum_{j=1}^r a_j e^{-\pi i (t-1) \sin \theta_j}, \quad t = 1, 2, \dots, n,$$

where $i = \sqrt{-1}$. This formulation aligns with (1) by setting $f_j = -\sin \theta_j$. Meanwhile, NMR signals exhibit spectral sparsity and conform to (1) (see figure 9 for an illustration). Consequently, problems such as DOA and NMR signal recovery can be formulated as low-rank Hankel matrix recovery problems, which are advantageous for handling off-grid signals and unknown damping factors [6, 28].

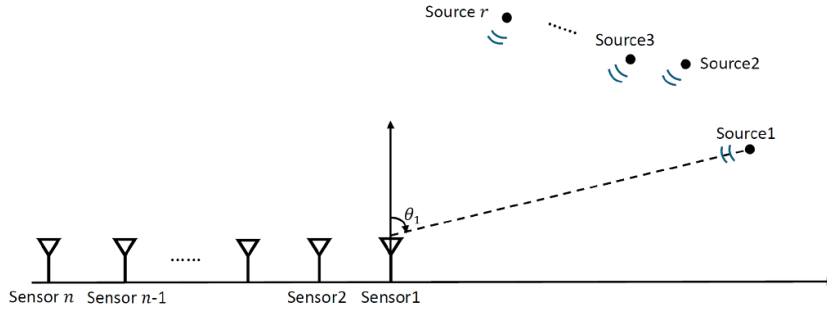


Figure 1. DOA signal collected by a ULA with n sensors and r far-field sources.

1.1. Problem formula and challenges

In practice, the low-rank Hankel recovery problem faces three major challenges: (i) *missing data*, (ii) *sparse outliers*, and (iii) *ill condition*. Missing data often result from hardware limitations or time constraints, leading to incomplete observations [36]. Sparse outliers typically arise due to hardware malfunctions, causing the recorded data to be sporadically corrupted by extreme values [39]. For instance, in DOA estimation, the availability of only a few sensors leads to missing data, while malfunctioning sensors introduce corrupted measurements. To address the first two challenges, we can formulate the Hankel matrix recovery problem as:

$$\begin{aligned} & \underset{\mathbf{X}, \mathbf{O} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \frac{1}{2p} \langle \mathcal{P}_{\mathbf{H}_\Omega} \mathbf{Y} - \mathcal{P}_{\mathbf{H}_\Omega} (\mathbf{X} + \mathbf{O}), \mathbf{Y} - (\mathbf{X} + \mathbf{O}) \rangle \\ & \text{subject to } \mathbf{X} \text{ is rank-}r \text{ Hankel matrix and } \mathbf{O} \text{ is sparse,} \end{aligned} \quad (2)$$

where p is the observation rate, \mathbf{O} is the sparse outlier matrix, \mathbf{Y} is the sparsely corrupted Hankel data matrix, and $\mathcal{P}_{\mathbf{H}_\Omega} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_2}$ is the matrix observation operator with the index set \mathbf{H}_Ω . However, as discussed in the literature [4], randomly positioned, non-structured missing entries and outliers can be easily recovered through the redundant skew-diagonal in the Hankel matrix. One can also Hankelize $\mathcal{P}_{\mathbf{H}_\Omega} \mathbf{Y}$ by averaging its skew-diagonal elements as \mathbf{X} is Hankel structured. Thus, we are only interested in the more challenging recovery problem with Hankel structured missing entries and outliers, i.e. some skew-diagonals are entirely missed or corrupted. Under this setting, all parties of (2) are Hankel structured including the observation pattern \mathbf{H}_Ω , thus we can rewrite the Hankel recovery problem in an equivalent vector form.

Instead of using the plain vector \mathbf{x} , we introduce a reweighed vector representation of \mathbf{X} :

$$\mathbf{z} = [x_1, \sqrt{2}x_2, \dots, \sqrt{c_i}x_i, \dots, \sqrt{2}x_{n-1}, x_n]^\top \in \mathbb{C}^n$$

such that $\|\mathbf{z}\|_2 = \|\mathbf{X}\|_F$ to avoid any unexpected bias in the vectorized formula. Moreover, define the reweighed Hankel mapping $\mathcal{G} : \mathbb{C}^n \rightarrow \mathbb{C}^{n_1 \times n_2}$ and reweighting operator $\mathcal{W} : \mathbb{C}^n \rightarrow \mathbb{C}^n$ as:

$$\mathcal{G}\mathbf{z} = \mathbf{X} \quad \text{and} \quad \mathcal{W}\mathbf{z} = \mathbf{x}. \quad (3)$$

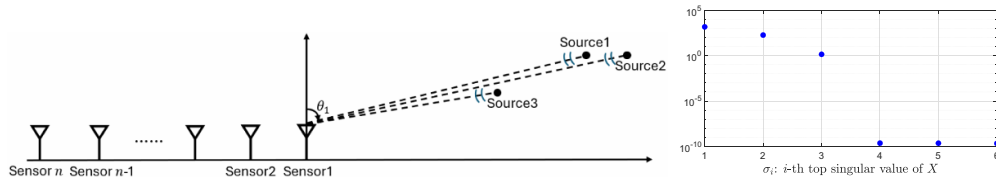


Figure 2. Left: DOA estimation for $r=3$ far-field sources from the directions $\theta = 87^\circ, 87.1^\circ, 87.3^\circ$, signal collected by a ULA with $n = 2^{12}$ sensors. Right: Top six singular values of the corresponding square Hankel matrix. For a rank-3 approximation, the condition number $\kappa = \sigma_1/\sigma_3 \approx 1083$. See section 4.2 for more details of the DOA experiment.

Note that $\mathcal{G}^*\mathcal{G} = \mathcal{I}$ where $(\cdot)^*$ denotes the adjoint operator. Thus, (2) can be rewritten as:

$$\begin{aligned} & \underset{\mathbf{z}, \mathbf{s} \in \mathbb{C}^n}{\text{minimize}} \quad \frac{1}{2p} \langle \Pi_\Omega \mathbf{f} - \Pi_\Omega (\mathbf{z} + \mathbf{s}), \mathbf{f} - (\mathbf{z} + \mathbf{s}) \rangle \\ & \text{subject to } \mathcal{G}\mathbf{z} \text{ is rank-}r \text{ and } \Pi_\Omega \mathbf{s} \text{ is } \alpha\text{-sparse,} \end{aligned} \quad (4)$$

where $\Pi_\Omega : \mathbb{C}^n \rightarrow \mathbb{C}^n$ is the observation operator on \mathbb{C}^n that corresponds to the Hankel structured sampling pattern, $\mathbf{f} = \mathcal{G}^*\mathbf{Y}$ and $\mathbf{s} = \mathcal{G}^*\mathbf{O}$ are observed data and outlier matrices in their equivalent vector form, and α is a sparsity parameter defined later in assumption 3. Since only the observed outliers interfere with the recovery, the sparsity constraint is only enforced on $\Pi_\Omega \mathbf{s}$.

Ill condition is one major challenge that remains unsolved for existing Hankel recovery literature, as the existing algorithms often struggle with convergence speed (detailed in section 1.4). Unfortunately, ill-conditioned Hankel matrices often appear in many applications, specifically, when the frequency of a spectrally sparse signal is inseparable, the corresponding Hankel matrix has a large condition number [26, 27]. An example is the DOA estimation mentioned above. When multiple sources arrive from close angles, the corresponding Hankel matrix of the DOA signal has a large condition number, see figure 2 for illustration. DOA and many other applications of Hankel recovery, such as traffic anomaly detection [12] and seismic monitoring [11], are very time-sensitive, thus developing highly efficient solvers is an eager task. This paper aims to study a provable non-convex algorithm for solving (4) that keeps high computational efficiency even if the condition number of the underlying Hankel matrix is large.

1.2. Assumptions

Denote the vector-form groundtruth as \mathbf{z}_\star and \mathbf{s}_\star for low-rank Hankel and sparse outlier respectively. To address the Hankel recovery problem (4), we introduce the following standard assumptions on data sampling, Hankel matrix incoherence, and outlier sparsity.

Assumption 1 (Uniform sampling with replacement). Let the observation index set $\Omega = \{a_k\}_{k=1}^m$ be sampled uniformly with replacement from $[n]$. The observation rate is $p := |\Omega|/n = m/n$.

In this context, the sampling operator is defined as $\Pi_\Omega \mathbf{v} = \sum_{i \in \Omega} v_i \mathbf{e}_i$ and allows repeated samples, thus $\langle \Pi_\Omega \mathbf{v}, \mathbf{v} \rangle \neq \|\Pi_\Omega \mathbf{v}\|_2^2$.

Assumption 2 (μ -incoherence). Let $U_\star \Sigma_\star V_\star^*$ be the compact singular value decomposition (SVD) of the underlying rank- r Hankel matrix $\mathcal{G}_{\mathbf{z}_\star} \in \mathbb{C}^{n_1 \times n_2}$. Then $\mathcal{G}_{\mathbf{z}_\star}$ is said μ -incoherent if there exists a positive constant μ such that

$$\|U_\star\|_{2,\infty} \leq \sqrt{\mu c_s r/n} \quad \text{and} \quad \|V_\star\|_{2,\infty} \leq \sqrt{\mu c_s r/n},$$

where $c_s = \max\{n/n_1, n/n_2\}$ measures how square the Hankel matrix is.

Assumption 3 (αp -sparsity). Given a sampling model (i.e. assumption 1), the observed outliers are at most αp -sparse. That is, $\|\Pi_\Omega \mathbf{s}_\star\|_0 \leq \alpha |\Omega| = \alpha p n$.

Assumptions 1–3 have been widely used in matrix completion [9, 31, 38], robust principal component analysis [8, 29, 43], and low-rank Hankel recovery literature [3, 6, 7, 14], ensuring that (4) is a well-posed problem [10]. Assumption 3 states that only a α proportion of observed data are corrupted, without specifying the randomness of the corruption pattern. This condition holds with high probability as long as \mathbf{s}_\star is $\frac{\alpha}{2}$ -sparse [5]. Moreover, the sparsity of \mathbf{s}_\star (resp. $\Pi_\Omega \mathbf{s}_\star$) implies the sparsity of $\mathcal{G}_{\mathbf{s}_\star}$ (resp. $\mathcal{G}\Pi_\Omega \mathbf{s}_\star$) in each of its rows and columns, provided $n_1 \approx n_2$.

1.3. Notation

Throughout the paper, we use regular lowercase letters for scalars (e.g. n), bold lowercase letters for vectors (e.g. \mathbf{v}), bold capital letters for matrices (e.g. \mathbf{M}), and calligraphic letters for operators (e.g. \mathcal{G}). $[n]$ denotes the set $\{1, 2, \dots, n\}$ for any positive integer n . Given a vector, $\|\mathbf{v}\|_0$, $\|\mathbf{v}\|_2$, and $\|\mathbf{v}\|_\infty$ denote its ℓ_0 -, ℓ_2 -, and ℓ_∞ -norms respectively. $|\mathbf{v}|$ is the entrywise absolute value of \mathbf{v} . Given a matrix \mathbf{M} , $\sigma_i(\mathbf{M})$ denotes the i th singular value, $\|\mathbf{M}\|_{2,\infty}$ denotes the largest row-wise ℓ_2 -norm, $\|\mathbf{M}\|_\infty$ denotes the largest entry-wise magnitude, $\|\mathbf{M}\|_2$ and $\|\mathbf{M}\|_F$ denote the spectral and Frobenius norms respectively. $\|\mathcal{G}\|$ denotes the operator norm of a linear operator. For all, $(\cdot)^*$, $(\cdot)^\top$, $(\cdot)^\dagger$, and $\langle \cdot, \cdot \rangle$ denote conjugate, transpose, conjugate transpose, and inner product respectively. Moreover, for ease of notation, we denote σ_i^* for the i th singular value of the underlying Hankel matrix $\mathcal{G}_{\mathbf{z}_\star}$ and $\kappa = \sigma_1^*/\sigma_r^*$ for its condition number. Finally, $a \lesssim \mathcal{O}(b)$ means there exists an absolute numerical constant $c > 0$ such that a can be upper bounded by cb .

1.4. Related work and contributions

The low-rank Hankel recovery problem has been extensively studied in the past decade. A naive approach is to directly apply robust matrix recovery algorithms [13, 16, 21, 41] on the large $n_1 \times n_2$ Hankel matrix which suffers from high computational complexities (i.e. $\mathcal{O}(n^2 r)$ to $\mathcal{O}(n^3)$ per iteration) and inferior recoverability since the Hankel structure is ignored. Lately, many dedicated Hankel recovery algorithms have been proposed. Robust-EMaC [14] proposes a convex relaxation for the non-convex problem (2) and solves it with semidefinite programming which is still computationally and memory too expensive. References [6, 7, 35] propose three algorithms with linear convergence that solve the non-convex Hankel completion problem in $\mathcal{O}(r^2 n + m \log n)$ flops per iteration; however, none of them handles outliers and [35] also lacks theoretical guarantee for Hankel recovery problem. SAP [42] proposes to effectively solve (4) via alternating projections. The recovery guarantee is established with $m \gtrsim \mathcal{O}(c_s^2 \mu^2 r^3 \log^2 n)$ and $\alpha \lesssim \mathcal{O}(1/(c_s \mu r))$. The overall computational complexity of SAP is $\mathcal{O}(r^2 n \log n \log(1/\varepsilon))$ with a large hidden constant due to a truncated SVD used in every iteration. The accelerated version of SAP, namely ASAP [3], improves the complexity to $\mathcal{O}((r^2 n + m \log n) \log(1/\varepsilon))$ flops. However, ASAP studies only the fully observed

cases. HSGD [4] solves (4) via projected gradient descent in $\mathcal{O}(r^2n + m \log n)$ flops per iteration, with a small constant. Although HSGD has a guaranteed linear convergence, its iteration complexity is $\mathcal{O}(\kappa \log(1/\varepsilon))$, which is problematic for heavily ill-conditioned problems. The recovery guarantee of HSGD requires $m \gtrsim \mathcal{O}(\max\{c_s^2 \mu^2 r^2 \log n, c_s \mu \kappa^3 r^2 \log n\})$ and $\alpha \lesssim \mathcal{O}(1/(\max\{c_s \mu \kappa^{3/2} r^{3/2}, c_s \mu r \kappa^2\}))$.

In this paper, our main contributions are three-fold:

- (i) We propose a highly efficient algorithm for heavily ill-conditioned robust Hankel recovery problems, coined Hankel structured Newton-like descent (HSNLD). Its overall computational cost is $\mathcal{O}((r^2n + m \log n) \log(1/\varepsilon))$ flops which beats all state-of-the-art algorithms. Especially when compared with HSGD, our convergence rate is independent of the condition number κ ; when compared with SAP, our complexity is $\mathcal{O}(\min\{\log(n), r\})$ lower and the hidden constant is much smaller since no iterative truncated SVD involved.
- (ii) The recovery guarantee, with linear convergence, has been established for HSNLD under some mild conditions. The sample complexity and outlier tolerance are $m \gtrsim \mathcal{O}(c_s^2 \mu^2 r^2 \kappa^4 \log n)$ and $\alpha \lesssim \mathcal{O}(1/(c_s \mu r \kappa^2))$ respectively. Compared to HSGD, the balance regularization is omitted from the objective, a new distance measurement is employed and a series of new technical lemmas has been developed accordingly, which will benefit future studies in the field.
- (iii) We verify the empirical advantages of HSNLD with synthetic datasets, as well as DOA and NMR signals. HSNLD achieves the best recoverability and computational efficiency in all ill-conditioned test problems. It gains more advantages when the condition numbers become even worse.

2. Proposed method

Hankel structured gradient descent (HSGD) [4] is a state-of-the-art non-convex algorithm for Hankel matrix recovery problems. It has shown superior performance for well-conditioned Hankel matrices, both theoretically and empirically. However, when the underlying Hankel matrix is ill-conditioned, the convergence of HSGD is significantly slower in theory. In particular, HSGD requires $\mathcal{O}(\kappa \log \varepsilon^{-1})$ iterations to find a ε -solution. This phenomenon is empirically verified in figure 3, where we observe that HSGD suffers from the larger condition number of the underlying Hankel matrix. To overcome this major weakness of HSGD, we propose a novel non-convex algorithm inspired by recent developments on Newton-like gradient descent algorithms [35]. The proposed algorithm, dubbed HSNLD, is highly efficient even for ill-conditioned problems.

To recover the low-rank Hankel matrix from missing data and sparse outliers, we aim at a similar objective function as used in [4]:

$$\ell(\mathbf{L}, \mathbf{R}; \mathbf{s}) = \frac{1}{2p} \langle \Pi_{\Omega}(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f}), \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + \mathbf{s} - \mathbf{f} \rangle + \frac{1}{2} \|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*)\|_{\mathbb{F}}^2, \quad (5)$$

where $\mathbf{z} = \mathcal{G}^*\mathbf{X} \in \mathbb{C}^n$ and we parameterize $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ as a product of $\mathbf{L} \in \mathbb{C}^{n_1 \times r}$ and $\mathbf{R} \in \mathbb{C}^{n_2 \times r}$, thus the low-rank constraint is implicitly encoded. Moreover, a regularization term

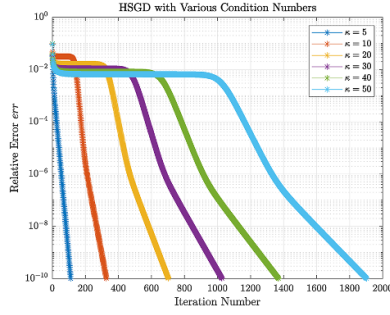


Figure 3. The convergence performance of HSGD [4] with different condition numbers κ .

Algorithm 1. Hankel Structured Newton-Like Descent (HSNLD).

- 1: **Input:** $\Pi_{\Omega}f$: partial observation on the corrupted Hankel matrix in reweighted vector form; r : the rank of underlying Hankel matrix; p : observation rate; α : outlier density; $\{\gamma_k\}$: parameters for sparsification operator; η : step size; K : maximum number of iterations.
 - 2: **Initialization:** Let L_0, R_0 be produced by (7).
 - 3: **for** $k = 0, 1, \dots, K-1$ **do**
 - 4: $z_{k+1} = \mathcal{G}^*(L_k R_k^*)$
 - 5: $s_{k+1} = \Gamma_{\gamma_k \alpha p}(\Pi_{\Omega}(f - z_{k+1}))$
 - 6: $\begin{bmatrix} L_{k+1} \\ R_{k+1} \end{bmatrix} = \mathcal{P}_C \begin{bmatrix} L_k - \eta \nabla_L \ell(L_k, R_k; s_{k+1}) (R_k^* R_k)^{-1} \\ R_k - \eta \nabla_R \ell(L_k, R_k; s_{k+1}) (L_k^* L_k)^{-1} \end{bmatrix}$
 - 7: **end for**
 - 8: **Output:** $z_{\text{output}} = \mathcal{G}^*(L_K R_K^*)$: recovered Hankel matrix in vector form.
-

$\|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(L R^*)\|_{\mathbb{F}}^2$ is added to enforce the Hankel structure of $L R^*$ ⁶. Targeting (5), the proposed HSNLD (summarized in algorithm 1) will be discussed step by step in the following paragraphs. For the sake of presentation, the iterative steps will be discussed first, and then initialization.

Remark 1. Compared to the objective posted in [4], our objective (5) omits the balance regularization term $\|L^* L - R^* R\|_{\mathbb{F}}^2$. The balance regularization is not empirically needed for the original HSGD algorithm but is required for its convergence analysis. In this paper, we drop the balance regularization for the proposed HSNLD with an improved convergence guarantee.

Iterative updates on s . For the sake of computational efficiency, one should avoid forming the full-size Hankel matrix and process the calculation in the corresponding vector form. Hence, we first update z with the current low-rank components, i.e. $z_{k+1} = \mathcal{G}^*(L_k R_k^*)$. By the definition of \mathcal{G}^* , the t th entry of z can be written as

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

⁶ The choice of the regularization parameter $\frac{1}{2}$ for the term $\|(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(L R^*)\|_{\mathbb{F}}^2$ is dictated by technical considerations. In particular, the application of lemma 12 in the proof of theorem 3 relies critically on this specific choice of $\frac{1}{2}$.

Thus, z_{k+1} can be obtained via r fast convolutions, with a computational cost in $\mathcal{O}(rn \log n)$ flops.

To detect the outliers, we compute the residue $\Pi_\Omega(\mathbf{f} - \mathbf{z}_{k+1})$ over the observation and project it onto the space of sparse vectors via a sparsification operator. That is, $s_{k+1} = \Gamma_{\gamma_k \alpha p} \Pi_\Omega(\mathbf{f} - \mathbf{z}_{k+1})$, where $\Gamma_{\gamma_k \alpha p} : \mathbb{C}^n \rightarrow \mathbb{C}^n$ is the vector sparsification operator that keeps the $\gamma_k \alpha p n$ largest-in-magnitude entries unchanged and vanishes the rest to 0. Therein, $\gamma_k > 1$ is a parameter that tolerates slight overestimation on the outlier sparsity as the true α is usually unavailable to the user, and it also provides redundant robustness in outlier detection. We observe that taking $\gamma_k \rightarrow 1$ as $k \rightarrow \infty$ provides balanced performance in robustness and efficiency; however, γ_k should always be larger than 1. With off-the-shelf fast partial sorting algorithm, the computational cost of s_{k+1} is $\mathcal{O}(n \log n)$ flops or better.

Iterative updates on \mathbf{L} and \mathbf{R} . The state-of-the-art HSGD [4] updates the low-rank Hankel components via iterative gradient descent. As discussed, its convergence is hindered by ill-conditioning. Inspired by recent work [19, 35], we utilize an additional Newton-like preconditioner in the iterative updates of \mathbf{L} and \mathbf{R} to properly address the ill-conditioned Hankel matrices. We have

$$\begin{bmatrix} \mathbf{L}_{k+1} \\ \mathbf{R}_{k+1} \end{bmatrix} = \mathcal{P}_C \begin{bmatrix} \mathbf{L}_k - \eta \nabla_{\mathbf{L}} \ell(\mathbf{L}_k, \mathbf{R}_k; s_{k+1}) (\mathbf{R}_k^* \mathbf{R}_k)^{-1} \\ \mathbf{R}_k - \eta \nabla_{\mathbf{R}} \ell(\mathbf{L}_k, \mathbf{R}_k; s_{k+1}) (\mathbf{L}_k^* \mathbf{L}_k)^{-1} \end{bmatrix}, \quad (6)$$

where $\eta > 0$ is the step size and \mathcal{P}_C is a projection operator that enforces the weak incoherence condition, i.e. $\max\{\|\mathbf{L}\mathbf{R}^*\|_{2,\infty}, \|\mathbf{R}\mathbf{L}^*\|_{2,\infty}\} \leq C$, on the low-rank Hankel matrix. In particular, modified from [35] we define

$$\mathcal{P}_C \begin{bmatrix} \tilde{\mathbf{L}} \\ \tilde{\mathbf{R}} \end{bmatrix} = \begin{bmatrix} \mathbf{L} \\ \mathbf{R} \end{bmatrix}, \text{ where } \mathbf{L}_{i,:} = \min \left(1, C / \|\tilde{\mathbf{L}}_{i,:} (\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}}\|_2 \right) \tilde{\mathbf{L}}_{i,:}, \quad \forall i \in [n_1],$$

$$\mathbf{R}_{j,:} = \min \left(1, C / \|\tilde{\mathbf{R}}_{j,:} (\tilde{\mathbf{L}}^* \tilde{\mathbf{L}})^{\frac{1}{2}}\|_2 \right) \tilde{\mathbf{R}}_{j,:}, \quad \forall j \in [n_2],$$

for some numerical constant $C > 0$. The choice of C will be specified in section 3 and the local non-expansiveness of \mathcal{P}_C is verified in lemma 7.

To efficiently update \mathbf{L} and \mathbf{R} , we notice that

$$\begin{aligned} \nabla_{\mathbf{L}} \ell(\mathbf{L}, \mathbf{R}; s) (\mathbf{R}^* \mathbf{R})^{-1} &= \left[\mathcal{G} \left(\frac{1}{p} \Pi_\Omega(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + s - \mathbf{f}) \right) \mathbf{R} + (\mathcal{I} - \mathcal{G}\mathcal{G}^*)(\mathbf{L}\mathbf{R}^*) \mathbf{R} \right] (\mathbf{R}^* \mathbf{R})^{-1} \\ &= \mathcal{G} \left(\frac{1}{p} \Pi_\Omega(\mathcal{G}^*(\mathbf{L}\mathbf{R}^*) + s - \mathbf{f}) - \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) \right) \mathbf{R} (\mathbf{R}^* \mathbf{R})^{-1} + \mathbf{L}. \end{aligned}$$

One can compute $\mathbf{z} := \mathcal{G}^*(\mathbf{L}\mathbf{R}^*)$ via r fast convolutions which cost $\mathcal{O}(rn \log n)$ flops. Computing $\frac{1}{p} \Pi_\Omega(\mathbf{z} + s - \mathbf{f}) - \mathbf{z}$ costs $\mathcal{O}(n)$ flops or better. Since $\mathcal{G}(\cdot)\mathbf{R}$ can be viewed as r convolutions, it costs $\mathcal{O}(rn \log n)$ flops. Computing $\mathbf{R}(\mathbf{R}^* \mathbf{R})^{-1}$ costs $\mathcal{O}(r^2 n)$ flops, so does the projection \mathcal{P}_C . Thus, the total computational cost of updating \mathbf{L} (and similarly for \mathbf{R}) is $\mathcal{O}(rn \log n + r^2 n)$.

Initialization. A tight initialization is important to the success of HSNLD. We adopt the modified Hankel spectral initialization from [4]:

$$s_0 = \mathcal{W}^{-1} \Gamma_{\alpha p}(\mathcal{W} \Pi_{\Omega} f), U_0 \Sigma_0 V_0^* = \text{SVD}_r \left(\frac{1}{p} \mathcal{G}(\Pi_{\Omega} f - s_0) \right), \begin{bmatrix} L_0 \\ R_0 \end{bmatrix} = \mathcal{P}_C \begin{bmatrix} U_0 \Sigma_0^{1/2} \\ V_0 \Sigma_0^{1/2} \end{bmatrix}. \quad (7)$$

This initialization has shown promising performance and a low computational complexity of $\mathcal{O}(m \log n)$. Note that the truncated SVD of a Hankel matrix costs as low as $\mathcal{O}(m \log n)$ flops since all matrix-vector multiplications therein can be viewed as convolutions.

Therefore, the overall computational complexity of the proposed algorithm is $\mathcal{O}(m \log n + r^2 n)$ which is tied to the current state-of-the-art HSGD [4]. In addition, HSNLD does not require forming the entire Hankel matrix, and only tracking the n distinct elements in the corresponding vector form is sufficient. Thus, the proposed HSNLD is computationally and memory efficient.

3. Theoretical results

We present the theoretical results of the proposed HSNLD in this section. We will first define some convenient notations and metrics for the presentation. Denote $L_{\star} := U_{\star} \Sigma_{\star}^{1/2}$ and $R_{\star} := V_{\star} \Sigma_{\star}^{1/2}$ where $U_{\star} \Sigma_{\star} V_{\star}^* = \mathcal{G}_{z_{\star}}$ is the compact SVD of the underlying Hankel matrix. Let (L_k, R_k) denote the update at the k th iteration in algorithm 1. Define the distance between (L_k, R_k) and (L_{\star}, R_{\star}) to be

$$d_k^2 := \text{dist}^2(L_k, R_k; L_{\star}, R_{\star}) := \inf_{Q \in \text{GL}(r, \mathbb{C})} \left\| (L_k Q - L_{\star}) \Sigma_{\star}^{1/2} \right\|_F^2 + \left\| (R_k Q^* - R_{\star}) \Sigma_{\star}^{1/2} \right\|_F^2. \quad (8)$$

Here, $\text{GL}(r, \mathbb{C})$ denotes the general linear group of degree r over \mathbb{C} and the invertible matrix Q is the best alignment between (L_k, R_k) and (L_{\star}, R_{\star}) . [35, lemma 22] has shown that Q exists if d_k is sufficiently small. In that case, the infimum in (8) can be replaced by the minimum, and the recovery error $\|L_k R_k^* - \mathcal{G}_{z_{\star}}\|_F$ can be bounded by d_k as stated in the following proposition.

Proposition 1. *For any $\varepsilon_0 \in (0, 1)$, provided $d_k < \varepsilon_0 \sigma_r^*$, then it holds that*

$$\|L_k R_k^* - \mathcal{G}_{z_{\star}}\|_F \leq \sqrt{2} (1 + \varepsilon_0/2) d_k.$$

Proof. The proof of this proposition is deferred to section 5.2. \square

The following theorem 1 guarantees that our initialization (7) will start HSNLD sufficiently close to the ground truth, under some mild conditions. Thus, it is sufficient to bound d_k for the recovery guarantee. In addition, our initial guess also satisfies the weak incoherence condition, which will be used and inherited iteratively in the convergence analysis. For ease of presentation, we denote the basin of attraction as:

$$\mathcal{E}(\zeta) := \left\{ (L, R) \mid \text{dist}(L, R; L_{\star}, R_{\star}) \leq \delta \text{ and } \max \left\{ \|L R^*\|_{2, \infty}, \|R L^*\|_{2, \infty} \right\} \leq C \right\}.$$

Theorem 1. For any $\varepsilon_0 \in (0, 1)$, let $C = c\sqrt{c_s\mu r/n}\sigma_1^*$ with $c \geq 1 + \varepsilon_0$, and suppose assumptions 1–3 hold with $m \geq c_0\varepsilon_0^{-2}c_s\mu\kappa^2r^2\log n$ and $\alpha \leq \varepsilon_0/(50c_s\mu r^{1.5}\kappa)$, where c_0 is some universal constant. Then, the initialization step (7) satisfies

$$(\mathbf{L}_0, \mathbf{R}_0) \in \mathcal{E}(\varepsilon_0\sigma_r^*, C)$$

with probability at least $1 - 2n^{-2}$. In the case of full observation, probability rises to 1.

Proof. The proof of this theorem is deferred to section 5.3. \square

3.1. Full observation

We start the analysis with the simpler full observation case, i.e. $\Omega = [n]$. Theorem 2 presents the local linear convergence of HSNLD.

Theorem 2. Suppose assumptions 2 and 3 hold with $\alpha \lesssim \mathcal{O}(1/(c_s\mu r\kappa^2))$. Choose parameters $C = c\sqrt{c_s\mu r/n}\sigma_1^*$ with $c \geq 1.01$, $\gamma_k \in [1 + 1/b_0, 2]$ with some fixed $b_0 \geq 1$, and $\eta \in (0, 1]$. Under full observation, for any $(\mathbf{L}_k, \mathbf{R}_k) \in \mathcal{E}(0.01\sigma_r^*, C)$, it holds that

$$d_{k+1} \leq (1 - 0.6\eta)d_k \quad \text{and} \quad (\mathbf{L}_{k+1}, \mathbf{R}_{k+1}) \in \mathcal{E}(0.01\sigma_r^*, C).$$

Proof. The proof of this theorem is deferred to section 5.4.1. \square

Combining proposition 1 and theorems 1 and 2, we have the recovery guarantee, with a linear convergence, for the proposed HSNLD under full observation.

Corollary 1 (recovery guarantee with full observation). Suppose assumptions 2 and 3 hold with $\alpha \lesssim \mathcal{O}(1/(c_s\mu r\kappa^2))$. Choose parameters $C = c\sqrt{c_s\mu r/n}\sigma_1^*$ with $c \geq 1.01$, $\gamma_k \in [1 + 1/b_0, 2]$ with some fixed $b_0 \geq 1$, and $\eta \in (0, 1]$. Then, under full observation, in $K = \mathcal{O}(\log(1/\varepsilon))$ iterations, the outputs of algorithm 1 satisfy

$$\|\mathbf{L}_K \mathbf{R}_K^* - \mathcal{G}z_*\|_F \leq 0.02(1 - 0.6\eta)^K \sigma_r^* \leq \varepsilon.$$

3.2. Partial observation

Now we will process the partial observation cases. One major challenge is the dependency among the entries on the antidiagonals of the Hankel matrix, which causes difficulty in uniformly bounding the subsampling error (see lemma 12). Now, we present the local convergence under partial observation.

Theorem 3. Suppose assumptions 1 and 2, 3 hold with $m \gtrsim \mathcal{O}(c_s^2\mu^2r^2\kappa^4\log n)$ and $\alpha \lesssim \mathcal{O}(1/(c_s\mu r\kappa^2))$. Choose the parameters $C = c\sqrt{c_s\mu r/n}\sigma_1^*$ with $c \geq 1.01$, $\gamma_k \in [1 + 1/b_0, 2]$ with some fixed $b_0 \geq 1$, and $\eta \in (0, 0.6]$. For any fixed $(\mathbf{L}_k, \mathbf{R}_k) \in \mathcal{E}(0.01\sigma_r^*, C)$, with probability at least $1 - 8n^{-2}$ the following results hold

$$d_{k+1} \leq (1 - 0.6\eta)d_k \quad \text{and} \quad (\mathbf{L}_{k+1}, \mathbf{R}_{k+1}) \in \mathcal{E}(0.01\sigma_r^*, C).$$

Proof. The proof of this theorem is deferred to section 5.4.2. \square

Combining proposition 1 and theorems 1 and 3, we have the recovery guarantee, with a linear convergence, for the proposed HSNLD under partial observation.

Corollary 2 (recovery guarantee with partial observation). Suppose assumptions 1–3 hold with $m \gtrsim \mathcal{O}(c_s^2 \mu^2 r^2 \kappa^4 \log n)$ and $\alpha \lesssim \mathcal{O}(1/(c_s \mu r \kappa^2))$. Let the observation set Ω be resampled independently every iteration. Choose parameters $C = c \sqrt{c_s \mu r / n \sigma_1^*}$ with $c \geq 1.01$, $\gamma_k \in [1 + 1/b_0, 2]$ with some fixed $b_0 \geq 1$, and $\eta \in (0, 0.6]$. Then, in $K = \mathcal{O}(\log(1/\varepsilon))$ iterations, the outputs of algorithm 1 satisfy

$$\|\mathbf{L}_K \mathbf{R}_K^* - \mathcal{G} \mathbf{z}_\star\|_F \leq 0.02 (1 - 0.6\eta)^K \sigma_r^* \leq \varepsilon$$

with probability at least $1 - \mathcal{O}(n^{-1})$, provided $\log(1/\varepsilon) \ll n$.

Remark 2. Compared to the state-of-the-art HSGD which runs $\mathcal{O}(\kappa \log(1/\varepsilon))$ iterations for a ε -optimizer, the convergence rate of HSNLD is independent of the condition number κ . This makes HSNLD a powerful tool for heavily ill-conditioned Hankel applications, e.g. spectrally sparse signals with inseparable frequencies [28].

4. Numerical experiments

In this section, we will empirically demonstrate the performance of the proposed HSNLD with synthetic and real datasets. We compare our algorithm with HSGD [4], SAP [42], and RobustEMaC [14]. Parameters are set according to the original paper and further hand-tuned for best performance when applicable. Particularly, both HSNLD and HSGD use the same outlier sparsification parameters: $\gamma_k = 1.05 + 0.45 \times 0.95^k$ as suggested in [4], and the step size $\eta = 0.5$. Moreover, we estimate the incoherence parameter μ via one-step Cadzow method [2] for all algorithms, and set the projection parameter C according to theorem 2 for HSNLD. The resampling is not implemented for HSNLD as it is just a proof technique and is not needed in practice. All results were conducted on Matlab with Intel i9-12950HX and 64GB RAM. The code is available at <https://github.com/caesarcai/HSNLD>.

4.1. Synthetic dataset

We generate an ill-conditioned Hankal matrix in its vector form $\mathbf{z} \in \mathbb{C}^n$ with

$$z_t = \sqrt{\varsigma_t} \sum_{j=1}^r a_j e^{2\pi i f_j t},$$

where i denotes the imaginary unit, $\{f_j\} \in \mathbb{R}$ is a randomly selected set of normalized off-grid frequencies, and $\{a_j\} \in \mathbb{R}$ is a set of r evenly spaced numbers over $[1/\kappa, 1]$. Thus, $\mathcal{G}\mathbf{z} \in \mathbb{C}^{n_1 \times n_2}$ is exactly rank- r with condition number κ . Then, m observations are sampled uniformly without replacement from \mathbf{z} , and therein αm entries are randomly chosen for corruption. We add complex outliers to the corrupted entries whose real and imaginary parts are drawn uniformly over the intervals $[-10\mathbb{E}(|\text{Re}(z_t)|), 10\mathbb{E}(|\text{Re}(z_t)|)]$ and $[-10\mathbb{E}(|\text{Im}(z_t)|), 10\mathbb{E}(|\text{Im}(z_t)|)]$.

Phase transition. We demonstrate the recoverability and robustness of HSNLD via empirical phase transition under various settings. When an algorithm stops at $\|\mathcal{G}\Pi_\Omega \mathbf{z}_k + \mathcal{G}\Pi_\Omega \mathbf{s}_k - \mathcal{G}\Pi_\Omega \mathbf{f}\|_F / \|\mathcal{G}\Pi_\Omega \mathbf{f}\|_F \leq 10^{-5}$, we considered it a successful recovery if $\|\mathcal{G}\mathbf{z}_k - \mathcal{G}\mathbf{z}_\star\|_F / \|\mathcal{G}\mathbf{z}_\star\|_F \leq 10^{-3}$. For every problem setting, each algorithm runs for 20 trials and reports

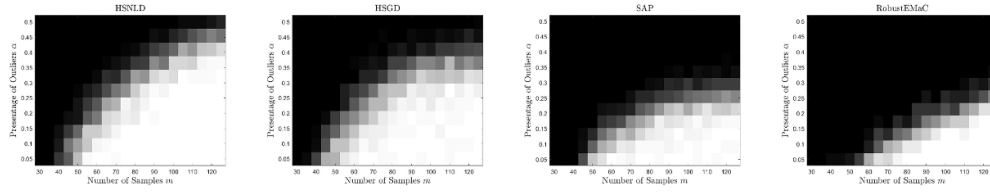


Figure 4. Phase transition: number of samples vs. rate of outliers. Problem dimension $n = 125$, condition number $\kappa = 10$, and rank $r = 10$ for every problem.

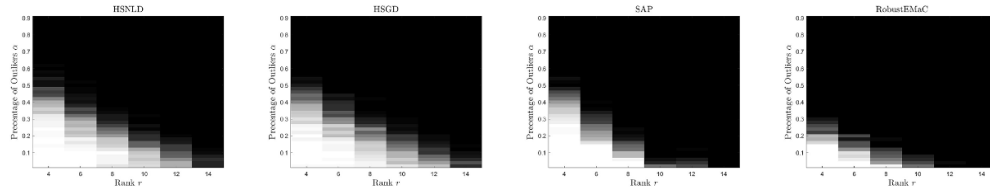


Figure 5. Phase transition: rank vs. rate of outliers. Problem dimension $n = 125$, condition number $\kappa = 10$, and $m = 50$ entries are sampled in every problem.

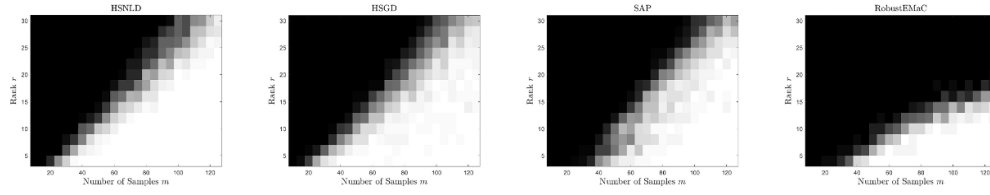


Figure 6. Phase transition: number of samples vs. rank. Problem dimension $n = 125$, condition number $\kappa = 10$, and outlier rate $\alpha = 10\%$ for every problem.

the result as a pixel on the phase transition, where a white pixel means all 20 trials were recovered and a black pixel means all 20 trials were failed. The phase transition results are reported in figures 4–6 where one can see HSNLD delivers the best recoverability and robustness for ill-conditioned problems. Note that we only run the phase transition with a reasonably large $\kappa = 10$ since the very ill-conditioned cases, such as $\kappa = 2000$, will take HSGD a forbidding long runtime.

Convergence performance. In figure 7, we plot the relative error v.s. runtime for the tested algorithms with different condition numbers. All problems have dimension $n = 2^{16} - 1$ and the results are averaged over five trials. When the condition number is perfect, i.e. $\kappa = 1$, the HSNLD has tied convergence performance with HSGD. When the condition number grows, HSGD’s performance drops dramatically while HSNLD takes no impact—this matches our theoretical results. Generally speaking, SAP’s convergence performance is not impacted by the condition number either; however, the truncated SVD it applies every iteration may fail

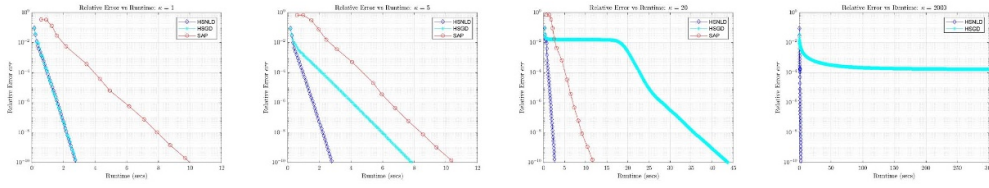


Figure 7. Convergence performance: relative error v.s. runtime with condition numbers $\kappa = 1, 5, 20, 2000$. Note that the singular values of complex-valued SVD failed to stay real with vast condition number $\kappa = 2000$, thus no convergence result can be reported for SAP.

to function with vast condition number. RobustEMaC is not tested here because its sublinear convergence is too slow for the comparison.

4.2. DOA estimation

Consider r narrowband far-field source signals $\{\tilde{g}_i\}_{i=1}^r \subseteq \mathbb{C}$ impinging on an array of omnidirectional sensors from distinct yet close directions as shown in figure 2, and the corresponding DOAs are $\{\theta_i\}_{i=1}^r$. Recalling that we have the DOA signal

$$x_j = \sum_{i=1}^r \tilde{g}_i e^{-\pi i(j-1) \sin \theta_i}, \quad j = 1, 2, \dots, n.$$

Consider the scenario where only a subset $\Omega \subseteq [n]$ of the sensors is available, and the received signals are partially corrupted. The resulting observations are of the form $y_j = x_j + o_j$, $j \in \Omega$, where o_j represents the outlier at the j th sensor. The problem of recovering $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$ from observed data on Ω (i.e. $\{y_j\}_{j \in \Omega}$) is equivalent to recover $\mathbf{z} = \mathcal{W}^{-1} \mathbf{x}$ as defined in equation (3) and can be formulated as a robust rank- r Hankel matrix completion problem. For the interested reader, please see [40, section 6.3] for further details.

We test HSNLD against the state-of-the-art with DOA data generated from $n = 2^{12}$ sensors and $r = 3$ far-field sources, with directions $\theta = 87^\circ, 87.1^\circ, 87.3^\circ$ as shown in figure 2. The underlying square Hankel matrix corresponding to the DOA data has a condition number $\kappa \approx 1083$. In this experiment, only 1.5% sensors are activated (or deployed), and 10% of the collected data are corrupted by complex outliers whose components are uniformly drawn from the intervals $[-\sum_{j=1}^n \text{Re}(x_j)/n, \sum_{j=1}^n \text{Re}(x_j)/n]$ and $[-\sum_{j=1}^n \text{Im}(x_j)/n, \sum_{j=1}^n \text{Im}(x_j)/n]$. All test algorithms are terminated at $\|z_k - z\|/\|z\| \leq 10^{-5}$. Figure 8 plots the signal recovered by HSNLD, compared to the original signal and corrupted partial observations. HSNLD successfully recovers the ground truth. The state-of-the-art algorithms, HSGD and SAP, also successfully recover the DOA signal. Due to space constraints, their plots are omitted. However, the runtime reported in table 1 shows the significant speed advantage of HSNLD; it achieves nearly $200 \times$ speedup to HSGD and $10 \times$ speedup to SAP.

4.3. NMR

We test HSNLD with the NMR signal recovery problem, a classic application of low-rank Hankel recovery. The 1D NMR signal being tested has a length of 32768 with an approximate rank of 40 and condition number 30. The observation rate is set to be 30% and 10% outliers are added. All algorithms stop at $|\text{err}_{k+1} - \text{err}_k|/\text{err}_k < 10^{-4}$, where $\text{err}_k =$

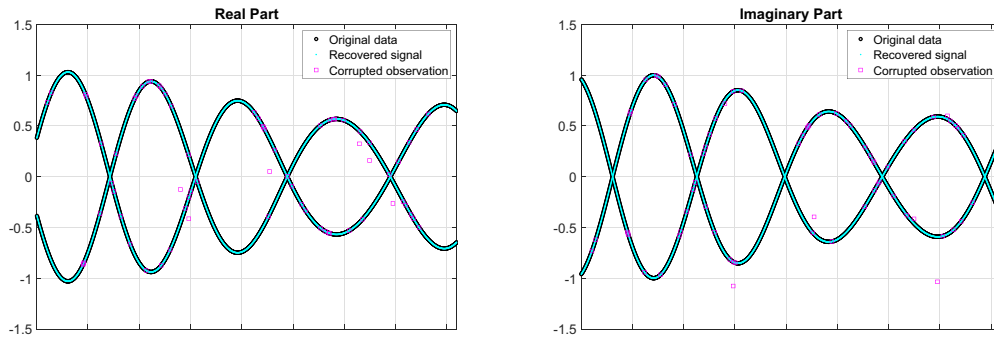


Figure 8. DOA data recover by HSNLD. Only 1.5% of the sensors are activated/deployed and 10% of the observations are corrupted by outliers. Left and right: Real and imaginary parts of the data.

Table 1. The number of iterations and runtime for DOA data recovery.

METHOD	HSNLD	HSGD	SAP
ITERATION	72	29 186	32
RUNTIME (SEC)	0.236	38.611	2.040

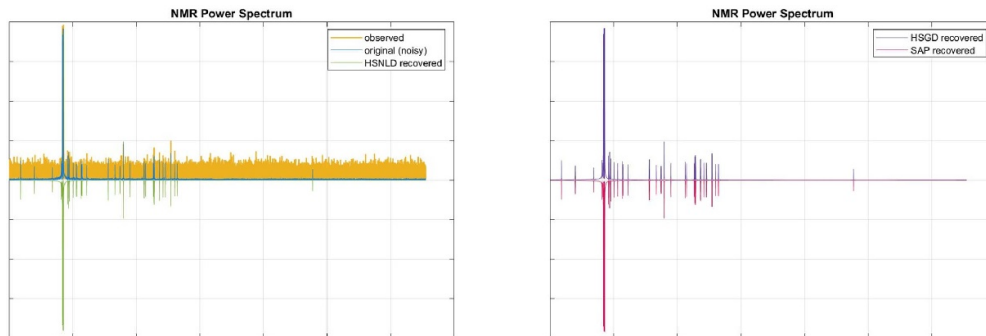


Figure 9. NMR recovery results in the form of power spectrum.

Table 2. The number of iterations and runtime for NMR recovery.

METHOD	HSNLD	HSGD	SAP
ITERATION	43	314	40
RUNTIME (SEC)	5.22	28.82	79.68

$\langle \Pi_{\Omega}(\mathbf{z}_k + \mathbf{s}_k - \mathbf{f}), \mathbf{z}_k + \mathbf{s}_k - \mathbf{f} \rangle / \langle \Pi_{\Omega} \mathbf{f}, \mathbf{f} \rangle$ due to no available groundtruth. The recovery results are reported in figure 9 and table 2. One can see that all three tested algorithms recovered the signal successfully, while HSNLD is significantly faster.

5. Proofs

In this section, we provide the analysis of the claimed theoretical results. All the proofs are under assumptions 1–3. We start with introducing some additional notations used in the analysis.

For ease of presentation, in addition to section 1.3, we denote the following slightly abused notation for the proofs. We let Ω be any subset of $[n]$ that satisfies assumption 1 with $m := |\Omega|$ and $p := m/n$. We denote the tangent space of rank- r matrix manifold at $\mathcal{G}\mathbf{z}_\star$ by

$$T := \{X \mid X = U_\star C^* + D V_\star^* \text{ for any } C \in \mathbb{C}^{n_1 \times r}, D \in \mathbb{C}^{n_2 \times r}\}.$$

Let Ω_a denote the index set of the a th antidiagonal of $n_1 \times n_2$ matrix, i.e.

$$\Omega_a := \{(i, j) : i + j = a + 1, 1 \leq i \leq n_1, 1 \leq j \leq n_2\}, \quad a = 1, 2, \dots, n. \quad (9)$$

We also introduce the following notations for the ease of presentation: let $\mathbf{Q}_k \in \text{GL}(r, \mathbb{C})$ be the invertible matrix that best aligns (L_k, R_k) and (L_\star, R_\star) , and according to [35, lemma 22] we know \mathbf{Q}_k always exists if $d_k < \sigma_r^*$. Denote

$$L_{\natural} := L_k \mathbf{Q}_k, \quad R_{\natural} := R_k \mathbf{Q}_k^{-*}, \quad \Delta_{L_k} := L_{\natural} - L_\star, \quad \Delta_{R_k} := R_{\natural} - R_\star. \quad (10)$$

5.1. Technical lemmas

In this subsection, we mainly introduce some supporting lemmas, including bounds estimated from probability inequalities.

Lemma 1 ([7, lemma 2], [4, lemma 9]). *It holds with probability at least $1 - 2n^{-2}$ that*

$$\left\| \left(\frac{1}{p} \mathcal{G} \Pi_\Omega - \mathcal{G} \right) \mathbf{z}_\star \right\|_2 \leq \tilde{c}_0 \sqrt{(\mu c_s r \log n) / m} \|\mathcal{G} \mathbf{z}_\star\|_2 \quad (11)$$

provided $m \geq \mu c_s r \log n$. \tilde{c}_0 is a universal constant.

Proof. The bound of $\left\| \left(\frac{1}{p} \mathcal{G} \Pi_\Omega - \mathcal{G} \right) \mathbf{z}_\star \right\|_2$ is provided in [7, lemma 2] (see [7, equation (25)]), and also proved with a slightly different sampling model in [4, lemma 9]. \square

Lemma 2 ([4, lemma 10]). *There exists a universal constant \tilde{c}_1 such that*

$$\left\| \frac{1}{p} \mathcal{P}_T \mathcal{G} \Pi_\Omega \mathcal{G}^* \mathcal{P}_T - \mathcal{P}_T \mathcal{G} \mathcal{G}^* \mathcal{P}_T \right\| \leq \varepsilon_0 \quad (12)$$

holds with probability at least $1 - 2n^{-2}$, provided $m \geq \varepsilon_0^{-2} \tilde{c}_1 \mu r \log n$.

Proof. The result is proved in [4, lemma 9] under a slightly different sampling model and can be extended to our sampling model, which is omitted here for simplicity. \square

Lemma 3 ([6, lemma 5]). *For all $\mathbf{u} \in \mathbb{R}^{n_1}$, $\mathbf{v} \in \mathbb{R}^{n_2}$, it holds that*

$$\frac{1}{p} \sum_{a_k \in \Omega} \sum_{(i,j) \in \Omega_{a_k}} u_i v_j \leq \|\mathbf{u}\|_1 \|\mathbf{v}\|_1 + \sqrt{\frac{24n \log n}{p}} \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \quad (13)$$

with probability at least $1 - 2n^{-2}$, provided $m \geq 3 \log n$.

Lemma 4. If $m \geq \log n$, then under event (13) it holds that

$$\begin{aligned} & \frac{1}{p} |\langle \Pi_{\Omega} \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*) \rangle| \\ & \leq \|\mathbf{L}_A\|_F^2 \|\mathbf{R}_A\|_F^2 + \sqrt{\frac{24n \log n}{p}} \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{L}_A\|_F \|\mathbf{R}_A\|_F, \end{aligned}$$

for all $\mathbf{L}_A \in \mathbb{C}^{n_1 \times r}$ and $\mathbf{R}_A \in \mathbb{C}^{n_2 \times r}$.

Proof. For $\Omega = \{a_k\}_{k=1}^m$, we have

$$\begin{aligned} & \frac{1}{p} |\langle \Pi_{\Omega} \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*) \rangle| \\ & = \frac{1}{p} \sum_{k=1}^m \left| \sum_{(i,j) \in \Omega_{a_k}} \frac{1}{\sqrt{a_k}} \langle \mathbf{L}_A \mathbf{R}_A^*, \mathbf{e}_i \mathbf{e}_j^\top \rangle \right|^2 \\ & \leq \frac{1}{p} \sum_{k=1}^m \sum_{(i,j) \in \Omega_{a_k}} \left| \langle \mathbf{L}_A \mathbf{R}_A^*, \mathbf{e}_i \mathbf{e}_j^\top \rangle \right|^2 \\ & \leq \frac{1}{p} \sum_{k=1}^m \sum_{(i,j) \in \Omega_{a_k}} \|\mathbf{L}_A(i, :)\|_2^2 \|\mathbf{R}_A(j, :)\|_2^2 \\ & \leq \|\mathbf{L}_A\|_F^2 \|\mathbf{R}_A\|_F^2 + \sqrt{\frac{24n \log n}{p}} \sqrt{\sum_i \|\mathbf{L}_A(i, :)\|_2^4} \sqrt{\sum_j \|\mathbf{R}_A(j, :)\|_2^4} \\ & \leq \|\mathbf{L}_A\|_F^2 \|\mathbf{R}_A\|_F^2 + \sqrt{\frac{24n \log n}{p}} \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{L}_A\|_F \|\mathbf{R}_A\|_F. \end{aligned}$$

where the first and second inequalities follow from the Cauchy–Schwarz inequality, and the third inequality follows from lemma 3. \square

Lemma 5. For all matrices $\mathbf{A}, \mathbf{B} \in T$, under event (12), it holds that

$$\left| \left\langle \mathcal{G} \left(\frac{1}{p} \Pi_{\Omega} - \mathcal{I} \right) \mathcal{G}^* \mathbf{A}, \mathbf{B} \right\rangle \right| \leq \varepsilon_0 \|\mathbf{A}\|_F \|\mathbf{B}\|_F,$$

and

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

Proof. For any matrices $\mathbf{A}, \mathbf{B} \in T$, we have $\mathcal{P}_T \mathbf{A} = \mathbf{A}$ and $\mathcal{P}_T \mathbf{B} = \mathbf{B}$. Then, under event (12), it holds that

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

Thus, together with the inequality $\|\mathcal{G}^* \mathbf{A}\|_2^2 = \langle \mathcal{G} \mathcal{G}^* \mathbf{A}, \mathbf{A} \rangle \leq \|\mathbf{A}\|_F^2$ we have

$$|\langle \mathcal{G} \Pi_{\Omega} \mathcal{G}^* \mathbf{A}, \mathbf{A} \rangle| \leq p \varepsilon_0 \|\mathbf{A}\|_F^2 + p \|\mathcal{G}^* \mathbf{A}\|_2^2 \leq p(1 + \varepsilon_0) \|\mathbf{A}\|_F^2.$$

\square

Lemma 6. For any $\mathbf{z} \in \mathbb{C}^n$ and $\mathbf{s} = \mathcal{W}^{-1}\Gamma_{\gamma\alpha p}(\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z}))$ with $\gamma \geq 1$, we have

$$\|\mathcal{W}\Pi_{\Omega}\mathbf{s}_{\star} - \mathcal{W}\mathbf{s}\|_{\infty} \leq 2\|\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})\|_{\infty}. \quad (14)$$

Proof. Recall $\mathbf{f} = \mathbf{z}_{\star} + \mathbf{s}_{\star}$. By the definition of \mathbf{s} , we have $\mathcal{W}\mathbf{s} = \Gamma_{\gamma\alpha p}(\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} + \mathbf{s}_{\star} - \mathbf{z}))$, thus

$$[\mathcal{W}\Pi_{\Omega}\mathbf{s}_{\star} - \mathcal{W}\mathbf{s}]_i = -[\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})]_i, \quad \text{for } i \in \text{supp}(\mathbf{s}).$$

Given that $\|\Pi_{\Omega}\mathbf{s}_{\star}\|_0 \leq \alpha p n = \alpha m$, there can be at most αm elements in $\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z})$ satisfying $|[\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z})]_i| > \|\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})\|_{\infty}$. Then, for $\gamma \geq 1$ we have

$$|[\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z})]_i| \leq \|\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})\|_{\infty}, \quad \text{for } i \in \text{supp}(\mathbf{s}_{\star}) \setminus \text{supp}(\mathbf{s}).$$

Thus, for $i \in \text{supp}(\mathbf{s}_{\star}) \setminus \text{supp}(\mathbf{s})$ we have

$$\begin{aligned} [\mathcal{W}\Pi_{\Omega}\mathbf{s}_{\star} - \mathcal{W}\mathbf{s}]_i &= [\mathcal{W}\Pi_{\Omega}\mathbf{s}_{\star}]_i = [\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z}) - \mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})]_i \\ &\leq |[\mathcal{W}\Pi_{\Omega}(\mathbf{f} - \mathbf{z})]_i| + |[\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})]_i| \\ &\leq 2\|\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})\|_{\infty}. \end{aligned}$$

Combing all the pieces we have $\|\mathcal{W}\Pi_{\Omega}\mathbf{s}_{\star} - \mathcal{W}\mathbf{s}\|_{\infty} \leq 2\|\mathcal{W}\Pi_{\Omega}(\mathbf{z}_{\star} - \mathbf{z})\|_{\infty}$. \square

Lemma 7. (Non-expansiveness of $\mathcal{P}_{\mathcal{C}}$) For all $(\tilde{\mathbf{L}}, \tilde{\mathbf{R}}) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, C)$ with $C \geq (1 + \varepsilon_0)\sqrt{\mu c_s r / n \sigma_1^*}$, $\varepsilon_0 \in (0, 1)$, and $\mathcal{P}_{\mathcal{C}} \begin{bmatrix} \tilde{\mathbf{L}} \\ \tilde{\mathbf{R}} \end{bmatrix} = \begin{bmatrix} \mathbf{L} \\ \mathbf{R} \end{bmatrix}$, we have

$$\text{dist}(\mathbf{L}, \mathbf{R}; \mathbf{L}_{\star}, \mathbf{R}_{\star}) \leq \text{dist}(\tilde{\mathbf{L}}, \tilde{\mathbf{R}}; \mathbf{L}_{\star}, \mathbf{R}_{\star})$$

and

$$\max\{\|\mathbf{L}\mathbf{R}^*\|_{2,\infty}, \|\mathbf{R}\mathbf{L}^*\|_{2,\infty}\} \leq C.$$

Proof. We first show the non-expansiveness of $\mathcal{P}_{\mathcal{C}}$. Let $\tilde{\mathbf{Q}} \in \text{GL}(r, \mathbb{C})$ be the best matrix that aligns $(\tilde{\mathbf{L}}, \tilde{\mathbf{R}})$ and $(\mathbf{L}_{\star}, \mathbf{R}_{\star})$. Notice that for $i \in [n_1]$ we have

$$\begin{aligned} \|\tilde{\mathbf{L}}_{i,:} (\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}}\|_2 &\leq \|\tilde{\mathbf{L}}_{i,:} \tilde{\mathbf{Q}} \Sigma_{\star}^{1/2}\|_2 \|(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}} \tilde{\mathbf{Q}}^{-*} \Sigma_{\star}^{-1/2}\|_2 \\ &\leq \|\tilde{\mathbf{L}}_{i,:} \tilde{\mathbf{Q}} \Sigma_{\star}^{1/2}\|_2 \|(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{-1/2} \tilde{\mathbf{R}}^*\|_2 \|\tilde{\mathbf{R}} \tilde{\mathbf{Q}}^{-*} \Sigma_{\star}^{-1/2}\|_2 \\ &\leq \|\tilde{\mathbf{L}}_{i,:} \tilde{\mathbf{Q}} \Sigma_{\star}^{1/2}\|_2 \left(\|(\tilde{\mathbf{R}} \tilde{\mathbf{Q}}^{-*} - \mathbf{R}_{\star}) \Sigma_{\star}^{-1/2}\|_2 + \|\mathbf{R}_{\star} \Sigma_{\star}^{-1/2}\|_2 \right) \\ &\leq (1 + \varepsilon_0) \|\tilde{\mathbf{L}}_{i,:} \tilde{\mathbf{Q}} \Sigma_{\star}^{1/2}\|_2, \end{aligned}$$

where the third inequality follows from $\|(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{-1/2} \tilde{\mathbf{R}}^*\|_2 = 1$, and the last inequality follows from

$$\begin{aligned} \|(\tilde{\mathbf{R}} \tilde{\mathbf{Q}}^{-*} - \mathbf{R}_{\star}) \Sigma_{\star}^{-1/2}\|_2 &\leq \|(\tilde{\mathbf{R}} \tilde{\mathbf{Q}}^{-*} - \mathbf{R}_{\star}) \Sigma_{\star}^{1/2} \Sigma_{\star}^{-1}\|_{\text{F}} \\ &\leq \|\Sigma_{\star}^{-1}\|_2 \text{dist}(\tilde{\mathbf{L}}, \tilde{\mathbf{R}}; \mathbf{L}_{\star}, \mathbf{R}_{\star}) \leq \varepsilon_0. \end{aligned} \quad (15)$$

Thus, as long as $C \geq (1 + \varepsilon_0) \sqrt{\mu c_s r / n \sigma_1^*}$, $\forall i \in [n_1]$ we have

$$\frac{C}{\|\tilde{L}_{i,:}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}}\|_2} \geq \frac{C}{(1 + \varepsilon_0) \|\tilde{L}_{i,:} \tilde{\mathbf{Q}} \Sigma_\star^{1/2}\|_2} \geq \frac{\|[L_\star]_{i,:} \Sigma_\star^{1/2}\|_2}{\|\tilde{L}_{i,:} \tilde{\mathbf{Q}} \Sigma_\star^{1/2}\|_2}, \quad (16)$$

where the last inequality follows from $L_\star \Sigma_\star^{1/2} = U_\star \Sigma_\star$ and

$$\|[L_\star]_{i,:} \Sigma_\star^{1/2}\|_2 \leq \|U_\star\|_{2,\infty} \|\Sigma_\star\|_2 \leq \sqrt{\mu c_s r / n \sigma_1^*}, \quad \forall i \in [n_1].$$

For any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, it holds that $\|\min(1, \lambda) \mathbf{u} - \mathbf{v}\|_2 \leq \|\mathbf{u} - \mathbf{v}\|_2$ provided $\lambda \geq \frac{\|\mathbf{v}\|_2}{\|\mathbf{u}\|_2}$ (see [35, claim 5]). Recalling that $L_{i,:} = \min\left(1, \frac{C}{\|\tilde{L}_{i,:}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}}\|_2}\right) \tilde{L}_{i,:}$, thus by equation (16) we have

$$\|(L_{i,:} \tilde{\mathbf{Q}} - [L_\star]_{i,:}) \Sigma_\star^{1/2}\|_2^2 \leq \|(\tilde{L}_{i,:} \tilde{\mathbf{Q}} - [L_\star]_{i,:}) \Sigma_\star^{1/2}\|_2^2$$

and similarly $\|(\mathbf{R}_{i,:} \tilde{\mathbf{Q}}^{-*} - [\mathbf{R}_\star]_{i,:}) \Sigma_\star^{1/2}\|_2^2 \leq \|(\tilde{\mathbf{R}}_{i,:} \tilde{\mathbf{Q}}^{-*} - [\mathbf{R}_\star]_{i,:}) \Sigma_\star^{1/2}\|_2^2$. Then, by definition of the metric we have the local non-expansiveness of \mathcal{P}_C , i.e.

$$\begin{aligned} \text{dist}^2(\mathbf{L}, \mathbf{R}; \mathbf{L}_\star, \mathbf{R}_\star) &\leq \|(\mathbf{L} \tilde{\mathbf{Q}} - \mathbf{L}_\star) \Sigma_\star^{1/2}\|_F^2 + \|(\mathbf{R} \tilde{\mathbf{Q}}^{-*} - \mathbf{R}_\star) \Sigma_\star^{1/2}\|_F^2 \\ &\leq \text{dist}^2(\tilde{\mathbf{L}}, \tilde{\mathbf{R}}; \mathbf{L}_\star, \mathbf{R}_\star). \end{aligned}$$

Noticing that $\|\tilde{L}_{i,:}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{1/2}\|_2 = \|\min(1, \frac{C}{\|\tilde{L}_{i,:}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{\frac{1}{2}}\|_2}) \tilde{L}_{i,:}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{1/2}\|_2 \leq C$, $\forall i \in [n_1]$, thus $\|\mathbf{L} \mathbf{R}^*\|_{2,\infty} \leq \|\mathbf{L}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{1/2}\|_{2,\infty} \|\mathbf{R}(\tilde{\mathbf{R}}^* \tilde{\mathbf{R}})^{-1/2}\|_2 \leq C$. Similarly, we have $\|\mathbf{R} \mathbf{L}^*\|_{2,\infty} \leq C$. \square

5.2. Proof of proposition 1

Proof. For $d_k \leq \varepsilon_0 \sigma_r^*$, following equation (15), we have $\|\Delta_{L_k} \Sigma_\star^{-1/2}\|_2 \leq \varepsilon_0$, thus

$$\|\Delta_{L_k} \Delta_{R_k}^*\|_F = \|\Delta_{L_k} \Sigma_\star^{-1/2} (\Delta_{R_k} \Sigma_\star^{1/2})^*\|_F \leq \varepsilon_0 \|\Delta_{R_k} \Sigma_\star^{1/2}\|_F.$$

Similarly, we have $\|\Delta_{R_k} \Sigma_\star^{-1/2}\|_2 \leq \varepsilon_0$ and $\|\Delta_{L_k} \Delta_{R_k}^*\|_F \leq \varepsilon_0 \|\Delta_{L_k} \Sigma_\star^{1/2}\|_F$. Therefore, we have

$$\begin{aligned} \|\mathbf{L}_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_\star\|_F &= \|\Delta_{L_k} \mathbf{R}_\star^* + \mathbf{L}_\star \Delta_{R_k}^* + \Delta_{L_k} \Delta_{R_k}^*\|_F \\ &\leq \|\Delta_{L_k} \Sigma_\star^{1/2} \mathbf{V}_\star^*\|_F + \|\mathbf{U}_\star \Sigma_\star^{1/2} \Delta_{R_k}^*\|_F + \|\Delta_{L_k} \Delta_{R_k}^*\|_F \\ &\leq \left(1 + \frac{\varepsilon_0}{2}\right) \left(\|\Delta_{L_k} \Sigma_\star^{1/2}\|_F + \|\Delta_{R_k} \Sigma_\star^{1/2}\|_F\right). \end{aligned} \quad (17)$$

Thus, by noticing $\left(\|\Delta_{L_k} \Sigma_\star^{1/2}\|_F + \|\Delta_{R_k} \Sigma_\star^{1/2}\|_F\right)^2 \leq 2d_k^2$, we complete the proof. \square

5.3. Proof of theorem 1 (guaranteed initialization)

Proof. This proof is under event (11) and we bound d_0 in the following. Following the notation of algorithm 1, we denote

$$\mathbf{M}_0 := \frac{1}{p} \mathcal{G}(\Pi_\Omega \mathbf{f} - \mathbf{s}_0), \quad \text{and } \mathbf{M}_0^{(r)} \text{ be the top-} r \text{ SVD of } \mathbf{M}_0. \quad (18)$$

For some universal constant \tilde{c}_0 , provided $m \geq 64\tilde{c}_0^2 \varepsilon_0^{-2} \kappa^2 c_s \mu r^2 \log n$, we have

$$\begin{aligned} \|\mathbf{M}_0 - \mathcal{G}\mathbf{z}_\star\|_2 &\leq \left\| \mathbf{M}_0 - \frac{1}{p} \mathcal{G} \Pi_\Omega \mathbf{z}_\star \right\|_2 + \left\| \frac{1}{p} \mathcal{G} \Pi_\Omega \mathbf{z}_\star - \mathcal{G}\mathbf{z}_\star \right\|_2 \\ &= \left\| \frac{1}{p} \mathcal{G}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0) \right\|_2 + \left\| \left(\frac{1}{p} \mathcal{G} \Pi_\Omega - \mathcal{G} \right) \mathbf{z}_\star \right\|_2 \\ &\leq 2\alpha n \|\mathcal{G}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0)\|_\infty + \tilde{c}_0 \sqrt{(\mu c_s r \log n) / m \sigma_1^\star} \\ &\leq 4\alpha n \|\mathcal{G}\mathbf{z}_\star\|_\infty + 0.125 \varepsilon_0 \sigma_r^\star r^{-\frac{1}{2}} \\ &\leq 4\alpha c_s \mu r \sigma_1^\star + 0.125 \varepsilon_0 \sigma_r^\star r^{-\frac{1}{2}}, \end{aligned} \quad (19)$$

where the second step follows from $\|\mathcal{G}(\Pi_\Omega \mathbf{f} - \mathbf{s}_0) - \mathcal{G} \Pi_\Omega \mathbf{z}_\star\|_2 = \|\mathcal{G}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0)\|_2$. The third line follows from [3, lemma 6] (as $\mathcal{G} = \mathcal{H}\mathcal{W}$ and $\mathcal{W}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0)$ is at most $2\alpha m$ -sparse) and lemma 1. The fourth line follows from (14) with $\mathbf{z} = \mathbf{0}$, i.e. $\|\mathcal{G}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0)\|_\infty = \|\mathcal{W}(\Pi_\Omega \mathbf{s}_\star - \mathbf{s}_0)\|_\infty \leq 2\|\mathcal{W} \Pi_\Omega \mathbf{z}_\star\|_\infty = 2\|\mathcal{G} \Pi_\Omega \mathbf{z}_\star\|_\infty$, because $\|\mathcal{G}\mathbf{z}\|_\infty$ and $\|\mathcal{W}\mathbf{z}\|_\infty$ with $\mathcal{W}\mathbf{z} = \mathbf{x}$ both equal the largest entry-wise magnitude in \mathbf{x} , see (3). And the last inequality follows from

$$\|\mathcal{G}\mathbf{z}_\star\|_\infty \leq \|\mathbf{U}_\star\|_{2,\infty} \|\mathcal{G}\mathbf{z}_\star\|_2 \|\mathbf{V}_\star\|_{2,\infty} \leq \mu \frac{c_s r}{n} \|\mathcal{G}\mathbf{z}_\star\|_2.$$

Therefore, we have

$$\begin{aligned} \left\| \mathbf{M}_0^{(r)} - \mathcal{G}\mathbf{z}_\star \right\|_2 &\leq \left\| \mathbf{M}_0^{(r)} - \mathbf{M}_0 \right\|_2 + \|\mathbf{M}_0 - \mathcal{G}\mathbf{z}_\star\|_2 \\ &\leq 2 \|\mathbf{M}_0 - \mathcal{G}\mathbf{z}_\star\|_2 \leq 8\alpha c_s \mu r \kappa \sigma_r^\star + 0.25 \varepsilon_0 \sigma_r^\star r^{-\frac{1}{2}}, \end{aligned} \quad (20)$$

where the second inequality follows from the definition of $\mathbf{M}_0^{(r)}$ and Eckart–Young–Mirsky theorem. Then, by [35, lemma 24] we have

$$\begin{aligned} \text{dist}\left(\mathbf{U}_0 \Sigma_0^{1/2}, \mathbf{V}_0 \Sigma_0^{1/2}; \mathbf{L}_\star, \mathbf{R}_\star\right) &\leq \left(\sqrt{2} + 1\right)^{1/2} \left\| \mathbf{M}_0^{(r)} - \mathcal{G}\mathbf{z}_\star \right\|_F \\ &\leq \sqrt{2} \left(\sqrt{2} + 1\right) r \left\| \mathbf{M}_0^{(r)} - \mathcal{G}\mathbf{z}_\star \right\|_2 \leq 18\alpha c_s \kappa \mu r \sqrt{r} \sigma_r^\star + 0.55 \varepsilon_0 \sigma_r^\star. \end{aligned}$$

Then, for any given $\varepsilon_0 \in (0, 1)$, we have $\text{dist}\left(\mathbf{U}_0 \Sigma_0^{1/2}, \mathbf{V}_0 \Sigma_0^{1/2}; \mathbf{L}_\star, \mathbf{R}_\star\right) \leq \varepsilon_0 \sigma_r^\star < \sigma_r^\star$ provided $0 \leq \alpha \leq \frac{\varepsilon_0}{50c_s \kappa \mu r \sqrt{r}}$. Thus, for $c \geq 1 + \varepsilon_0$, by lemma 7 we have

$$d_0 \leq \text{dist}\left(\mathbf{U}_0 \Sigma_0^{1/2}, \mathbf{V}_0 \Sigma_0^{1/2}; \mathbf{L}_\star, \mathbf{R}_\star\right) \leq \varepsilon_0 \sigma_r^\star,$$

and $\max\{\|\mathbf{L}_0 \mathbf{R}_0^\star\|_{2,\infty}, \|\mathbf{R}_0 \mathbf{L}_0^\star\|_{2,\infty}\} \leq c \sqrt{\frac{c_s \mu r}{n}} \sigma_1^\star$. This completes the proof. \square

5.4. Proof of theorems 2 and 3 (local convergence)

In this subsection, based on the probabilistic events in section 5.1, we prove several inequalities that are crucial for the proof of our main theorems. The proof follows a similar route established in [35], but the techniques and details are quite involved due to the extra Hankel structure, as well as the simultaneous presence of outliers and missing data.

In the following, we let $\varepsilon_0 \in (0, 1)$ be some fixed constant, and show that for any given $(\mathbf{L}_k, \mathbf{R}_k) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, c \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*)$ we have the contraction $d_{k+1} \leq \tau d_k$ for some $\tau \in (0, 1)$ provided ε_0 sufficiently small and $c \geq 1 + \varepsilon_0$. We begin with some auxiliary lemmas.

Lemma 8. For any $(\mathbf{L}_k, \mathbf{R}_k) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, c \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*)$, we have

$$\max \left\{ \left\| \Delta_{L_k} \Sigma_\star^{-1/2} \right\|_2, \left\| \Delta_{R_k} \Sigma_\star^{-1/2} \right\|_2 \right\} \leq \varepsilon_0, \quad (21a)$$

$$\max \left\{ \left\| L_\natural \Sigma_\star^{1/2} \right\|_{2,\infty}, \left\| R_\natural \Sigma_\star^{1/2} \right\|_{2,\infty} \right\} \leq \frac{c}{1-\varepsilon_0} \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*, \quad (21b)$$

$$\max \left\{ \left\| L_\natural \Sigma_\star^{-1/2} \right\|_{2,\infty}, \left\| R_\natural \Sigma_\star^{-1/2} \right\|_{2,\infty} \right\} \leq \frac{c\kappa}{1-\varepsilon_0} \sqrt{\frac{c_s \mu r}{n}}, \quad (21c)$$

$$\left\| R_\natural (R_\natural^* R_\natural)^{-1} \Sigma_\star^{1/2} \right\|_2 \leq \frac{1}{1-\varepsilon_0}. \quad (21d)$$

Proof. The inequality (21a) follows from (15). To show (21b), noticing that

$$\sigma_r \left(L_\natural \Sigma_\star^{-1/2} \right) \geq \sigma_r \left(L_\star \Sigma_\star^{-1/2} \right) - \sigma_1 \left(\Delta_{L_k} \Sigma_\star^{-1/2} \right) \geq 1 - \varepsilon_0,$$

then by $\|L_k R_k^*\|_{2,\infty} \geq \sigma_r(R_\natural \Sigma_\star^{-1/2}) \left\| L_\natural \Sigma_\star^{1/2} \right\|_{2,\infty}$ we have $\left\| L_\natural \Sigma_\star^{1/2} \right\|_{2,\infty} \leq \frac{\|L_k R_k^*\|_{2,\infty}}{1-\varepsilon_0}$.

Similarly, we have $\left\| R_\natural \Sigma_\star^{1/2} \right\|_{2,\infty} \leq \frac{\|L_k R_k^*\|_{2,\infty}}{1-\varepsilon_0}$ and conclude (21b). By using $\|AB\|_{2,\infty} \leq \|A\|_{2,\infty} \|B\|_2$, (21c) follows directly from (21b). (21d) follows from [35, lemma 25] and (21a). \square

Lemma 9. For all matrices $(\mathbf{L}_k, \mathbf{R}_k) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, c \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*)$, it holds that

$$\begin{aligned} & \left| \langle \Pi_\Omega \mathcal{G}^* (\mathbf{L}_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_\star), \mathcal{G}^* (\mathbf{L}_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_\star) \rangle \right|^{\frac{1}{2}} \\ & \leq \sqrt{p} (1 + 3\sqrt{\varepsilon_0}) \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right) \end{aligned}$$

provided $m \geq 6C_{\varepsilon_0, \kappa}^2 \varepsilon_0^{-2} c_s^2 \mu^2 r^2 \log n$, where $C_{\varepsilon_0, \kappa} = \left(1 + \frac{c}{1-\varepsilon_0}\right) \left(1 + \frac{c\kappa}{1-\varepsilon_0}\right) \kappa$.

Proof. The decomposition $L_k R_k^* - \mathcal{G} \mathbf{z}_\star = \Delta_{L_k} R_\star^* + L_\star \Delta_{R_k}^* + \Delta_{L_k} \Delta_{R_k}^*$ gives

$$\begin{aligned} & \left| \langle \Pi_\Omega \mathcal{G}^* (\mathbf{L}_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_\star), \mathcal{G}^* (\mathbf{L}_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_\star) \rangle \right|^{\frac{1}{2}} \\ & \leq \left| \langle \Pi_\Omega \mathcal{G}^* (\Delta_{L_k} R_\star^* + L_\star \Delta_{R_k}^*), \mathcal{G}^* (\Delta_{L_k} R_\star^* + L_\star \Delta_{R_k}^*) \rangle \right|^{\frac{1}{2}} \\ & \quad + \left| \langle \Pi_\Omega \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*) \rangle \right|^{\frac{1}{2}} \\ & \leq \sqrt{p(1+\varepsilon_0)} \left\| \left(\Delta_{L_k} \Sigma_\star^{1/2} V_\star^* + U_\star \Sigma_\star^{1/2} \Delta_{R_k}^* \right) \right\|_F + \left| \langle \Pi_\Omega \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*) \rangle \right|^{\frac{1}{2}} \\ & \leq \sqrt{p(1+\varepsilon_0)} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right) + \left| \langle \Pi_\Omega \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^* (\Delta_{L_k} \Delta_{R_k}^*) \rangle \right|^{\frac{1}{2}}, \end{aligned}$$

where the first step follows from $|\langle \Pi_\Omega(\mathbf{u} + \mathbf{v}), \mathbf{u} + \mathbf{v} \rangle|^{\frac{1}{2}} \leq |\langle \Pi_\Omega \mathbf{u}, \mathbf{u} \rangle|^{\frac{1}{2}} + |\langle \Pi_\Omega \mathbf{v}, \mathbf{v} \rangle|^{\frac{1}{2}}$ for $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$, the second inequality follows from lemma 5, and in the last inequality we have used $\|\Delta_{L_k} \Sigma_\star^{1/2} \mathbf{V}_\star^*\|_F = \|\Delta_{L_k} \Sigma_\star^{1/2}\|_F$. For the last term, provided $m \geq 6\varepsilon_0^{-2} C_{\varepsilon_0, \kappa}^2 c_s^2 \mu^2 r^2 \log n$, we have

$$\begin{aligned} & \frac{1}{p} |\langle \Pi_\Omega \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*) \rangle| \\ & \leq \left\| \Delta_{L_k} \Sigma_\star^{-1/2} \right\|_2^2 \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \\ & \quad + \sqrt{\frac{24n \log n}{p}} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_{2,\infty} \left\| \Delta_{R_k} \Sigma_\star^{-1/2} \right\|_{2,\infty} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F \left\| \Delta_{R_k} \Sigma_\star^{-1/2} \right\|_F \\ & \leq \left\| \Delta_{L_k} \Sigma_\star^{-1/2} \right\|_2^2 \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 + \varepsilon_0 \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \right) \end{aligned}$$

where the first inequality follows from lemma 4, the second inequality follows from the inequalities $\left\| \Delta_{R_k} \Sigma_\star^{-1/2} \right\|_F = \left\| \Delta_{R_k} \Sigma_\star^{1/2} \Sigma_\star^{-1} \right\|_F \leq \frac{1}{\sigma_r^*} \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F$, and (by (21c))

$$\begin{aligned} \left\| \Delta_{L_k} \Sigma_\star^{-\frac{1}{2}} \right\|_{2,\infty} & \leq \left\| L_\natural \Sigma_\star^{-\frac{1}{2}} \right\|_{2,\infty} + \left\| L_\star \Sigma_\star^{-\frac{1}{2}} \right\|_{2,\infty} \leq \left(1 + \frac{c\kappa}{1 - \varepsilon_0} \right) \sqrt{\frac{c_s \mu r}{n}}, \\ \left\| \Delta_{L_k} \Sigma_\star^{\frac{1}{2}} \right\|_{2,\infty} & \leq \left\| L_\natural \Sigma_\star^{\frac{1}{2}} \right\|_{2,\infty} + \left\| L_\star \Sigma_\star^{\frac{1}{2}} \right\|_{2,\infty} \leq \left(1 + \frac{c}{1 - \varepsilon_0} \right) \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*. \end{aligned} \quad (22)$$

Therefore,

$$\begin{aligned} & |\langle \Pi_\Omega \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*) \rangle|^{\frac{1}{2}} \\ & \leq \sqrt{p} \left\| \Delta_{L_k} \Sigma_\star^{-1/2} \right\|_2 \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F + \sqrt{\varepsilon_0 p} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right). \end{aligned}$$

Similarly, we have

$$\begin{aligned} & |\langle \Pi_\Omega \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*) \rangle|^{\frac{1}{2}} \\ & \leq \sqrt{p} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F \left\| \Delta_{R_k} \Sigma_\star^{-1/2} \right\|_2 + \sqrt{\varepsilon_0 p} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right). \end{aligned}$$

By (21a) and averaging the above two bounds, we conclude

$$|\langle \Pi_\Omega \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*), \mathcal{G}^*(\Delta_{L_k} \Delta_{R_k}^*) \rangle|^{\frac{1}{2}} \leq 2\sqrt{p\varepsilon_0} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right).$$

Putting all the bound of terms together, we conclude the lemma. \square

Lemma 10. For any indices set $\widehat{\Omega} \subseteq [n]$ we have

$$\begin{aligned} & |\langle \Pi_{\widehat{\Omega}} \mathcal{G}^*(L_k R_k^* - \mathcal{G}z_\star), \mathcal{G}^*(L_k R_k^* - \mathcal{G}z_\star) \rangle|^{\frac{1}{2}} \\ & \leq \left(1 + \frac{c\kappa}{1 - \varepsilon_0} \right) \sqrt{c_s \mu r |\widehat{\Omega}| / (2n)} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right), \end{aligned} \quad (23)$$

and for all $\mathbf{L} \in \mathbb{C}^{n_1 \times r}$, $\mathbf{R} \in \mathbb{C}^{n_2 \times r}$ it holds that

$$|\langle \Pi_{\widehat{\Omega}} \mathcal{G}^*(\mathbf{L} \mathbf{R}^*), \mathcal{G}^*(\mathbf{L} \mathbf{R}^*) \rangle|^{\frac{1}{2}} \leq \sqrt{|\widehat{\Omega}|} \min \left\{ \|\mathbf{L}\|_{2,\infty} \|\mathbf{R}\|_F, \|\mathbf{R}\|_{2,\infty} \|\mathbf{L}\|_F \right\}.$$

Proof. By using the decomposition $L_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_* = (\mathbf{L}_{\natural} - \mathbf{L}_*) \mathbf{R}_{\natural}^* + \mathbf{L}_* (\mathbf{R}_{\natural} - \mathbf{R}_*) = (\mathbf{L}_{\natural} - \mathbf{L}_*) \mathbf{R}_{\natural}^* + \mathbf{L}_{\natural} (\mathbf{R}_{\natural} - \mathbf{R}_*)$, we have

$$\begin{aligned} & 2 \left| [\mathbf{L}_{\natural} \mathbf{R}_{\natural}^* - \mathcal{G} \mathbf{z}_*]_{i,j} \right| \\ &= \left| [(\mathbf{L}_{\natural} - \mathbf{L}_*) \mathbf{R}_{\natural}^* + \mathbf{L}_* (\mathbf{R}_{\natural} - \mathbf{R}_*)]_{i,j} + [(\mathbf{L}_{\natural} - \mathbf{L}_*) \mathbf{R}_{\natural}^* + \mathbf{L}_{\natural} (\mathbf{R}_{\natural} - \mathbf{R}_*)]_{i,j} \right| \\ &\leq \left\| \mathbf{R}_{\natural} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2 + \left\| \mathbf{L}_* \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2 \\ &\quad + \left\| \mathbf{R}_{\star} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2 + \left\| \mathbf{L}_{\natural} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2 \\ &\leq \theta \left(\left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2 + \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2 \right). \end{aligned}$$

Where

$$\begin{aligned} \theta &= \max \left\{ \left\| \mathbf{R}_{\natural} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} + \left\| \mathbf{R}_{\star} \Sigma_{\star}^{-1/2} \right\|_{2,\infty}, \left\| \mathbf{L}_* \Sigma_{\star}^{-1/2} \right\|_{2,\infty} + \left\| \mathbf{L}_{\natural} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \right\} \\ &\leq \left(1 + \frac{c\kappa}{1 - \varepsilon_0} \right) \sqrt{c_s \mu r / n} \end{aligned}$$

due to lemma 8. Let $\hat{\Omega} = \{\hat{a}_t\}_{1 \leq t \leq |\hat{\Omega}|}$, denote $\hat{\Phi}_t := \{(i, j) : (i + j - 1) = \hat{a}_t\}$ and $\hat{\Phi} = \cup_{1 \leq t \leq |\hat{\Omega}|} \hat{\Phi}_t$. Then, by definition in (10), we have

$$\begin{aligned} & \left| \langle \Pi_{\hat{\Omega}} \mathcal{G}^* (L_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_*), \mathcal{G}^* (L_k \mathbf{R}_k^* - \mathcal{G} \mathbf{z}_*) \rangle \right| = \sum_{\hat{a}_t \in \hat{\Omega}} \frac{1}{\zeta_{\hat{a}_t}} \left| \sum_{(i,j) \in \hat{\Phi}_t} [\mathbf{L}_{\natural} \mathbf{R}_{\natural}^* - \mathcal{G} \mathbf{z}_*]_{i,j} \right|^2 \\ &\leq \sum_{\hat{a}_t \in \hat{\Omega}} \sum_{(i,j) \in \hat{\Phi}_t} |[\mathbf{L}_{\natural} \mathbf{R}_{\natural}^* - \mathcal{G} \mathbf{z}_*]_{i,j}|^2 = \sum_{(i,j) \in \hat{\Phi}} |[\mathbf{L}_{\natural} \mathbf{R}_{\natural}^* - \mathcal{G} \mathbf{z}_*]_{i,j}|^2 \\ &\leq \sum_{(i,j) \in \hat{\Phi}} \frac{\theta^2}{4} \left(\left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2 + \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2 \right)^2 \\ &\leq \sum_{(i,j) \in \hat{\Phi}} \frac{\theta^2}{2} \left(\left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2^2 + \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2^2 \right) \\ &= \frac{\theta^2}{2} \left(\sum_i \sum_{1 \leq t \leq |\hat{\Omega}|} \chi_{[1,\infty)}(\hat{a}_t + 1 - i) \left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2^2 \right. \\ &\quad \left. + \sum_j \sum_{1 \leq t \leq |\hat{\Omega}|} \chi_{[1,\infty)}(\hat{a}_t + 1 - j) \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2^2 \right) \\ &\leq \frac{\theta^2}{2} \left(\sum_i |\hat{\Omega}| \left\| [\Delta_{L_k} \Sigma_{\star}^{1/2}] (i, :) \right\|_2^2 + \sum_j |\hat{\Omega}| \left\| [\Delta_{R_k} \Sigma_{\star}^{1/2}] (j, :) \right\|_2^2 \right) \\ &\leq \frac{\theta^2}{2} \left(|\hat{\Omega}| \left\| \Delta_{L_k} \Sigma_{\star}^{1/2} \right\|_{\text{F}}^2 + |\hat{\Omega}| \left\| \Delta_{R_k} \Sigma_{\star}^{1/2} \right\|_{\text{F}}^2 \right), \end{aligned}$$

where the first inequality follows from $|\widehat{\Phi}_t| = \varsigma_{\hat{a}_t}$ and the Cauchy–Schwarz inequality, the third inequality follows from $(a+b)^2 \leq 2a^2 + 2b^2$, and $\chi_{[1,\infty)}$ is the indicator function of $[1, \infty)$. Therefore, by $\sqrt{a^2 + b^2} \leq |a| + |b|$ we conclude the inequality (23). Similarly, by

$$\begin{aligned} |\langle \Pi_{\widehat{\Omega}} \mathcal{G}^*(\mathbf{L}\mathbf{R}^*), \mathcal{G}^*(\mathbf{L}\mathbf{R}^*) \rangle| &\leq \sum_{(i,j) \in \widehat{\Phi}} |[\mathbf{L}(i, :)\mathbf{R}^*(j, :)]^2| \leq \sum_{(i,j) \in \widehat{\Phi}} \|\mathbf{L}(i, :)\|_2^2 \|\mathbf{R}^*(j, :)\|_2^2 \\ &\leq |\widehat{\Omega}| \min \left\{ \|\mathbf{L}\|_{2,\infty}^2 \|\mathbf{R}\|_{\mathbb{F}}^2, \|\mathbf{R}\|_{2,\infty}^2 \|\mathbf{L}\|_{\mathbb{F}}^2 \right\}, \end{aligned}$$

we conclude the second statement. \square

Lemma 11. Let $\mathbf{s} = \Gamma_{\gamma\alpha}(\Pi_{\Omega}(\mathbf{f} - \mathcal{G}^*(\mathbf{L}_k\mathbf{R}_k^*)))$ with $\gamma > 1$. Then, for all $\mathbf{L}_A \in \mathbb{C}^{n_1 \times r}$, $\mathbf{R}_A \in \mathbb{C}^{n_2 \times r}$, under event (12) and (13) and provided $m \geq 6C_{\varepsilon_0, \kappa}^2 \varepsilon_0^{-2} c_s^2 \mu^2 r^2 \log n$, we have

$$\begin{aligned} &\frac{1}{p} |\langle \mathcal{G}\Pi_{\Omega}(\mathbf{s} - \mathbf{s}_*), \mathbf{L}_A\mathbf{R}_A^* \rangle| \\ &\leq \sqrt{n}\xi \left(\left\| \left[\Delta_{\mathbf{L}} \Sigma_{\star}^{1/2} \right] \right\|_{\mathbb{F}} + \left\| \left[\Delta_{\mathbf{R}} \Sigma_{\star}^{1/2} \right] \right\|_{\mathbb{F}} \right) \min \left\{ \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{R}_A\|_{\mathbb{F}}, \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{L}_A\|_{\mathbb{F}} \right\}, \end{aligned}$$

where $\xi = \left((\gamma+1)\alpha \left(1 + \frac{c\kappa}{1-\varepsilon_0} \right) \sqrt{\frac{c_s \mu r}{2}} + (1+3\sqrt{\varepsilon_0}) \frac{\sqrt{(\gamma+1)\alpha}}{\sqrt{(\gamma-1)}} \right)$.

Proof. Denote $\Delta_k := \mathbf{L}_k\mathbf{R}_k^* - \mathcal{G}\mathbf{z}_{\star}$, and $\Omega^s := \text{supp}(\Pi_{\Omega}\mathbf{s})$, $\Omega^{s*} := \text{supp}(\Pi_{\Omega}\mathbf{s}_{\star})$. By the definition of \mathbf{s} , we have

$$|[\Pi_{\Omega}\mathbf{s}_{\star} - \mathbf{s}]_i| = |[\Pi_{\Omega}(\mathbf{z}_{\star} - \mathcal{G}^*(\mathbf{L}_k\mathbf{R}_k^*))]_i| = |[\Pi_{\Omega}(\mathcal{G}^*\Delta_k)]_i|, \text{ for } i \in \Omega^s, \quad (24)$$

where the last equation follows from $\mathcal{G}^*\mathcal{G} = \mathcal{I}$. For $i \in \Omega^{s*} \setminus \Omega^s$, by

$$\mathbf{s} = \Gamma_{\gamma\alpha}(\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*(\mathbf{L}_k\mathbf{R}_k^* - \mathcal{G}\mathbf{z}_{\star}))) = \Gamma_{\gamma\alpha}(\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*\Delta_k)),$$

we know $|[\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*\Delta_k)]_i|$ is smaller than the $\gamma\alpha m$ th largest-in-magnitude entry of $|\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*\Delta_k)|$. Noticing $\Pi_{\Omega}\mathbf{s}_{\star}$ is at most αm sparse, thus $|[\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*\Delta_k)]_i|$ is smaller than the $(\gamma-1)\alpha m$ th largest entry of $|\Pi_{\Omega}(\mathcal{G}^*\Delta_k)|$, and $|[\Pi_{\Omega}(\mathbf{s}_{\star} - \mathcal{G}^*\Delta_k)]_i|^2 \leq \frac{|\langle \Pi_{\Omega}(\mathcal{G}^*\Delta_k), \mathcal{G}^*\Delta_k \rangle|}{(\gamma-1)\alpha m}$. Then, we have

$$|[\Pi_{\Omega}\mathbf{s}_{\star}]_i| \leq \frac{|\langle \Pi_{\Omega}(\mathcal{G}^*\Delta_k), \mathcal{G}^*\Delta_k \rangle|^{1/2}}{\sqrt{(\gamma-1)\alpha m}} + |[\Pi_{\Omega}(\mathcal{G}^*\Delta_k)]_i|, \text{ for } i \in \Omega^{s*} \setminus \Omega^s. \quad (25)$$

Hence,

$$\begin{aligned} &|\langle \mathcal{G}\Pi_{\Omega}(\mathbf{s} - \mathbf{s}_{\star}), \mathbf{L}_A\mathbf{R}_A^* \rangle| = |\langle \Pi_{\Omega}(\mathbf{s} - \mathbf{s}_{\star}), \mathcal{G}^*(\mathbf{L}_A\mathbf{R}_A^*) \rangle| \\ &\leq \sum_{i \in \Omega^s} |[\Pi_{\Omega}(\mathbf{s} - \mathbf{s}_{\star})]_i| |[\mathcal{G}^*(\mathbf{L}_A\mathbf{R}_A^*)]_i| + \sum_{i \in \Omega^{s*} \setminus \Omega^s} |[\Pi_{\Omega}(\mathbf{s}_{\star})]_i| |[\mathcal{G}^*(\mathbf{L}_A\mathbf{R}_A^*)]_i| \\ &\leq \sum_{i \in \Omega^s \cup \Omega^{s*}} |[\Pi_{\Omega}(\mathcal{G}^*\Delta_k)]_i| |[\mathcal{G}^*(\mathbf{L}_A\mathbf{R}_A^*)]_i| \\ &\quad + \sum_{i \in \Omega^{s*} \setminus \Omega^s} \frac{|\langle \Pi_{\Omega}(\mathcal{G}^*\Delta_k), \mathcal{G}^*\Delta_k \rangle|^{1/2}}{\sqrt{(\gamma-1)\alpha m}} |[\mathcal{G}^*(\mathbf{L}_A\mathbf{R}_A^*)]_i| \end{aligned}$$

$$\begin{aligned}
&\leq |\langle \Pi_{\Omega^s \cup \Omega^{s*}}(\mathcal{G}^* \Delta_k), \mathcal{G}^* \Delta_k \rangle|^{\frac{1}{2}} |\langle \Pi_{\Omega^s \cup \Omega^{s*}} \mathcal{G}^*(L_A R_A^*), \mathcal{G}^*(L_A R_A^*) \rangle|^{\frac{1}{2}} \\
&\quad + \frac{|\langle \Pi_{\Omega}(\mathcal{G}^* \Delta_k), \mathcal{G}^* \Delta_k \rangle|}{2\beta(\gamma-1)} + \frac{\beta}{2} |\langle \Pi_{\Omega^s \cup \Omega^{s*}} \mathcal{G}^*(L_A R_A^*), \mathcal{G}^*(L_A R_A^*) \rangle| \\
&\leq \left(1 + \frac{c\kappa}{1-\varepsilon_0}\right) \sqrt{\frac{c_s \mu r}{2n}} (\gamma+1) \alpha m \left(\|\Delta_{L_k} \Sigma_{\star}^{1/2}\|_{\text{F}} + \|\Delta_{R_k} \Sigma_{\star}^{1/2}\|_{\text{F}} \right) \\
&\quad \times \min \left\{ \|L_A\|_{2,\infty} \|R_A\|_{\text{F}}, \|R_A\|_{2,\infty} \|L_A\|_{\text{F}} \right\} + \frac{|\langle \Pi_{\Omega} \mathcal{G}^*(\Delta_k), \mathcal{G}^*(\Delta_k) \rangle|}{2\beta(\gamma-1)} \\
&\quad + \frac{\beta(\gamma+1)\alpha m}{2} \min \left\{ \|L_A\|_{2,\infty}^2 \|R_A\|_{\text{F}}^2, \|R_A\|_{2,\infty}^2 \|L_A\|_{\text{F}}^2 \right\},
\end{aligned}$$

where the third inequality follows from $ab \leq \frac{\beta}{2}a^2 + \frac{b^2}{2\beta}$ for any $\beta > 0$, and the Cauchy–Schwarz inequality, the fourth inequality follows from lemma 10. Letting

$$\beta = \frac{|\langle \Pi_{\Omega} \mathcal{G}^*(\Delta_k), \mathcal{G}^*(\Delta_k) \rangle|^{\frac{1}{2}}}{\left(\sqrt{(\gamma-1)(\gamma+1)\alpha m} \min \left\{ \|L_A\|_{2,\infty} \|R_A\|_{\text{F}}, \|R_A\|_{2,\infty} \|L_A\|_{\text{F}} \right\} \right)},$$

we then have

$$\begin{aligned}
&|\langle \mathcal{G} \Pi_{\Omega}(s - s_{\star}), L_A R_A^* \rangle| \\
&\leq \left[\left(1 + \frac{c\kappa}{1-\varepsilon_0}\right) \sqrt{\frac{c_s \mu r}{2n}} (\gamma+1) \alpha m \left(\|\Delta_{L_k} \Sigma_{\star}^{1/2}\|_{\text{F}} + \|\Delta_{R_k} \Sigma_{\star}^{1/2}\|_{\text{F}} \right) \right. \\
&\quad \left. + \frac{\sqrt{(\gamma+1)\alpha m}}{\sqrt{(\gamma-1)}} |\langle \Pi_{\Omega} \mathcal{G}^*(\Delta_k), \mathcal{G}^*(\Delta_k) \rangle|^{\frac{1}{2}} \right] \min \left\{ \|L_A\|_{2,\infty} \|R_A\|_{\text{F}}, \|R_A\|_{2,\infty} \|L_A\|_{\text{F}} \right\} \\
&\leq p\sqrt{n}\xi \left(\|\Delta_{L_k} \Sigma_{\star}^{1/2}\|_{\text{F}} + \|\Delta_{R_k} \Sigma_{\star}^{1/2}\|_{\text{F}} \right) \min \left\{ \|L_A\|_{2,\infty} \|R_A\|_{\text{F}}, \|R_A\|_{2,\infty} \|L_A\|_{\text{F}} \right\},
\end{aligned}$$

where $\xi = \left((\gamma+1)\alpha \left(1 + \frac{c\kappa}{1-\varepsilon_0}\right) \sqrt{\frac{c_s \mu r}{2}} + (1 + 3\sqrt{\varepsilon_0}) \frac{\sqrt{(\gamma+1)\alpha}}{\sqrt{(\gamma-1)}} \right)$ and the last inequality follows from lemma 9. \square

5.4.1. Contraction property in full observation.

Proof of theorem 2. For the case of full observation $m = n$ ($p = 1$), we show the contraction property for one step update of L_k, R_k in terms of the metric d_k . Denote

$$\begin{aligned}
\tilde{L}_{k+1} &:= L_k - \eta \nabla_{L} \ell(L_k, R_k; s_{k+1}) (R_k^* R_k)^{-1} \\
&= L_k - \eta \mathcal{G}(s_{k+1} - f) R_k (R_k^* R_k)^{-1} - \eta L_k.
\end{aligned}$$

Then, we have

$$\begin{aligned}
 (\tilde{L}_{k+1}Q_k - L_\star) \Sigma_\star^{1/2} &= [\Delta_{L_k} - \eta \mathcal{G}[s_{k+1} - z_\star - s_\star] R_\natural (R_\natural^\star R_\natural)^{-1} - \eta L_\natural] \Sigma_\star^{1/2} \\
 &= \Delta_{L_k} \Sigma_\star^{1/2} - \eta [\mathcal{G}(s_{k+1} - s_\star) + (L_\natural R_\natural^\star - L_\star R_\star^\star)] R_\natural (R_\natural^\star R_\natural)^{-1} \Sigma_\star^{1/2} \\
 &= \underbrace{(1 - \eta) \Delta_{L_k} \Sigma_\star^{1/2} - \eta L_\star \Delta_{R_k}^\star R_\natural (R_\natural^\star R_\natural)^{-1} \Sigma_\star^{1/2}}_{T_1} \\
 &\quad - \underbrace{\eta [\mathcal{G}(s_{k+1} - s_\star)] R_\natural (R_\natural^\star R_\natural)^{-1} \Sigma_\star^{1/2}}_{T_2},
 \end{aligned}$$

where in the third equation we have used the decomposition $L_\natural R_\natural^\star - L_\star R_\star^\star = \Delta_{L_k} R_\natural^\star + L_\star \Delta_{R_k}^\star = \Delta_{L_k} R_\star^\star + L_\star \Delta_{R_k}^\star + \Delta_{L_k} \Delta_{R_k}^\star$. Noticing that

$$\left\| (\tilde{L}_{k+1}Q_k - L_\star) \Sigma_\star^{1/2} \right\|_F^2 = \|T_1\|_F^2 + \|T_2\|_F^2 - 2\operatorname{Re} \langle T_1, T_2 \rangle, \quad (26)$$

and we have the following claims, whose proof is deferred later in this subsection:

Claim 1. For $(L_k, R_k) \in \mathcal{E}(\varepsilon_0 \sigma_r^\star, c\sqrt{\frac{c_s \mu r}{n}} \sigma_1^\star)$, we have

$$\begin{aligned}
 \|T_1\|_F^2 &\leq \left((1 - \eta)^2 + \frac{2\varepsilon_0}{1 - \varepsilon_0} \eta(1 - \eta) + \frac{2\varepsilon_0 + \varepsilon_0^2}{(1 - \varepsilon_0)^2} \eta^2 \right) d_k^2, \\
 |\langle T_1, T_2 \rangle| &\leq \frac{(\eta - \eta^2) c\kappa \sqrt{c_s \mu r} \xi}{2(1 - \varepsilon_0)^3} \left(3 \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \right) \\
 &\quad + \frac{\eta^2 \sqrt{c_s \mu r} \xi}{2(1 - \varepsilon_0)^2} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + 3 \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \right), \\
 \|T_2\|_F &\leq \frac{2\eta c\kappa \sqrt{c_s \mu r} \xi}{(1 - \varepsilon_0)^3} \left(\left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \right).
 \end{aligned}$$

Recall that $\xi = \left((\gamma + 1)\alpha \left(1 + \frac{c\kappa}{1 - \varepsilon_0} \right) \sqrt{\frac{c_s \mu r}{2}} + (1 + 3\sqrt{\varepsilon_0}) \frac{\sqrt{(\gamma + 1)\alpha}}{\sqrt{(\gamma - 1)}} \right)$. Therefore, provided $\alpha \leq \frac{\varepsilon_0^2(\gamma - 1)}{64(\gamma + 1)(1 + c\kappa)^2 c_s \mu r}$, by substituting the above three inequalities in claim 1 to (26), we have

$$\begin{aligned}
 &\left\| (\tilde{L}_{k+1}Q_k - L_\star) \Sigma_\star^{1/2} \right\|_F^2 \\
 &\leq C_1(\varepsilon_0, \eta) d_k^2 + C_2(\varepsilon_0, \eta) \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + C_3(\varepsilon_0, \eta) \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \\
 &\leq (1 - 0.6\eta)^2 d_k^2, \quad \text{for } \varepsilon_0 = 0.01, 0 < \eta \leq 1,
 \end{aligned}$$

where we have used the Cauchy–Schwarz inequality and the constants satisfy $C_1(\varepsilon_0, \eta) \leq (1 - \eta)^2 + \frac{2\varepsilon_0}{1 - \varepsilon_0} \eta(1 - \eta) + \frac{8\eta^2 \varepsilon_0^2}{(1 - \varepsilon_0)^6}$, $C_2(\varepsilon_0, \eta) \leq \frac{3\eta \varepsilon_0 - 2\eta^2 \varepsilon_0 - \eta^2 \varepsilon_0^2}{2(1 - \varepsilon_0)^3}$, $C_3(\varepsilon_0, \eta) \leq \frac{\eta \varepsilon_0 + 2\eta^2 \varepsilon_0 - 3\eta^2 \varepsilon_0^2}{2(1 - \varepsilon_0)^3}$. Similarly, let $\tilde{R}_{k+1} := R_k - \eta \nabla_{R\ell}(L_k, R_k; s_{k+1})(R_k^\star R_k)^{-1}$, we can show $\left\| (\tilde{R}_{k+1}Q_k^\star - R_\star) \Sigma_\star^{1/2} \right\|_F^2 \leq (1 - 0.6\eta)^2 d_k^2$. Thus we have

$$\text{dist}(\tilde{\mathbf{L}}_{k+1}, \tilde{\mathbf{R}}_{k+1}; \mathbf{L}_*, \mathbf{R}_*) \leq (1 - 0.6\eta) d_k.$$

Then, lemma 7 gives $d_{k+1} \leq (1 - 0.6\eta)d_k$ and $(\mathbf{L}_{k+1}, \mathbf{R}_{k+1}) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, c \sqrt{\frac{c_s \mu F}{n}} \sigma_1^*)$. Finally, we will complete the proof by proving claim 1. \square

Proof of claim 1. The upper bound of $\|\mathbf{T}_1\|_F$ follows the argument as in [35, equation (46)]. For $|\langle \mathbf{T}_1, \mathbf{T}_2 \rangle|$, we denote

$$\begin{aligned} \mathbf{L}_{A_1} &:= \Delta_{L_k} \Sigma_* (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2}, \quad \mathbf{R}_{A_1} := \mathbf{R}_\natural \Sigma_*^{-1/2}, \\ \mathbf{L}_{A_2} &:= \mathbf{L}_* \Sigma_*^{-1/2} = \mathbf{U}_*, \quad \mathbf{R}_{A_2} := \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_* (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \mathbf{R}_\natural^* \Delta_{R_k} \Sigma_*^{1/2}. \end{aligned} \quad (27)$$

Then, by applying lemma 11 with $m = n$ we have

$$\begin{aligned} |\langle \mathbf{T}_1, \mathbf{T}_2 \rangle| &= \left| \eta(1 - \eta) \text{tr} \left([\mathcal{G}(s_{k+1} - s_*)] \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_* \Delta_{L_k}^* \right) \right. \\ &\quad \left. - \eta^2 \text{tr} \left([\mathcal{G}(s_{k+1} - s_*)] \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_* (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \mathbf{R}_\natural^* \Delta_{R_k} \mathbf{L}_*^* \right) \right| \\ &\leq \eta(1 - \eta) |\langle \mathcal{G}(s_{k+1} - s_*), \mathbf{L}_{A_1} \mathbf{R}_{A_1}^* \rangle| + \eta^2 |\langle \mathcal{G}(s_{k+1} - s_*), \mathbf{L}_{A_2} \mathbf{R}_{A_2}^* \rangle| \\ &\leq \sqrt{n} \xi \left(\left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F \right) \\ &\quad \times \left((\eta - \eta^2) \|\mathbf{R}_{A_1}\|_{2,\infty} \|\mathbf{L}_{A_1}\|_F + \eta^2 \|\mathbf{L}_{A_2}\|_{2,\infty} \|\mathbf{R}_{A_2}\|_F \right), \end{aligned}$$

together with $ab \leq \frac{a^2+b^2}{2}$, (21c) and (21d) and the fact that

$$\begin{aligned} \left\| \Sigma_*^{1/2} (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2} \right\|_2 &= \left\| \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2} \right\|_2^2, \\ \left\| \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_* (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \mathbf{R}_\natural^* \right\|_2 &= \left\| \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2} \right\|_2^2, \\ \stackrel{(21d)}{\implies} \|\mathbf{L}_{A_1}\|_F &\leq \frac{1}{(1 - \varepsilon_0)^2} \left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F, \quad \|\mathbf{R}_{A_2}\|_F \leq \frac{1}{(1 - \varepsilon_0)^2} \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F, \end{aligned} \quad (28)$$

we conclude the second inequality in claim 1. For $\|\mathbf{T}_2\|_F$, using the variational representation of the Frobenius norm, for some $\mathbf{L}_A \in \mathbb{C}^{m \times r}$, $\|\mathbf{L}_A\|_F = 1$ we have

$$\begin{aligned} \|\mathbf{T}_2\|_F &= \eta \left| \left\langle [\mathcal{G}(s_{k+1} - s_*)] \mathbf{R}_\natural (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2}, \mathbf{L}_A \right\rangle \right| \\ &= \eta \left| \left\langle [\mathcal{G}(s_{k+1} - s_*)], \mathbf{L}_A \Sigma_*^{1/2} (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2} (\mathbf{R}_\natural \Sigma_*^{-1/2})^* \right\rangle \right| \\ &\leq \eta \sqrt{n} \xi \left(\left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F \right) \left\| \mathbf{R}_\natural \Sigma_*^{-1/2} \right\|_{2,\infty} \left\| \mathbf{L}_A \Sigma_*^{1/2} (\mathbf{R}_\natural^* \mathbf{R}_\natural)^{-1} \Sigma_*^{1/2} \right\|_F \\ &\leq \frac{2\eta c \kappa \sqrt{c_s \mu F} \xi}{(1 - \varepsilon_0)^3} \left(\left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F \right), \end{aligned}$$

where the first inequality follows from lemma 11, the last inequality follows from (21c) and similar argument to (28). \square

5.4.2. Contraction property in partial observation. The following lemma is crucial for the proof of the theorem.

Lemma 12. For any $\mathbf{L}_A, \mathbf{L}_B \in \mathbb{C}^{n_1 \times r}$, $\mathbf{R}_A, \mathbf{R}_B \in \mathbb{C}^{n_2 \times r}$ independent of Ω , with probability at least $1 - 2n^{-2}$ it holds that

$$\begin{aligned} & \left| \left\langle \mathcal{G} \left(\frac{1}{p} \Pi_\Omega - \mathcal{I} \right) \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathbf{L}_B \mathbf{R}_B^* \right\rangle \right| \\ & \leq \sqrt{\frac{24n \log n}{p}} \min \left\{ \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_F, \|\mathbf{L}_A\|_F \|\mathbf{L}_B\|_{2,\infty} \right\} \\ & \quad \times \min \left\{ \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{R}_B\|_F, \|\mathbf{R}_A\|_F \|\mathbf{R}_B\|_{2,\infty} \right\} \end{aligned}$$

provided $m \geq 24 \log n$.

Proof. First, we have

$$\begin{aligned} & \left| \left\langle \mathcal{G} \left(\frac{1}{p} \Pi_\Omega - \mathcal{I} \right) \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathbf{L}_B \mathbf{R}_B^* \right\rangle \right| \\ & = \frac{1}{p} \left| \langle (\Pi_\Omega - p\mathcal{I}) \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathcal{G}^* (\mathbf{L}_B \mathbf{R}_B^*) \rangle \right| = \frac{1}{p} \left| \sum_{k=1}^m r_k \right|, \end{aligned}$$

where

$$r_k := [\mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*)]_{a_k} [\mathcal{G}^* (\mathbf{L}_B \mathbf{R}_B^*)]_{a_k} - n^{-1} \langle \mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*), \mathcal{G}^* (\mathbf{L}_B \mathbf{R}_B^*) \rangle, \quad k \in [m],$$

then we have $\mathbb{E}[r_k] = 0$. Further, noticing that for any $a \in [n]$, we have

$$\begin{aligned} & |[\mathcal{G}^* (\mathbf{L}_A \mathbf{R}_A^*)]_a [\mathcal{G}^* (\mathbf{L}_B \mathbf{R}_B^*)]_a| \\ & = \left| \left(\sum_{(i,j) \in \Omega_a} \frac{1}{\varsigma_a} \mathbf{L}_A(i,:) \mathbf{R}_A(j,:)^* \right) \left(\sum_{(i,j) \in \Omega_a} \mathbf{L}_B(i,:) \mathbf{R}_B(j,:)^* \right) \right| \\ & \leq \left(\frac{1}{\varsigma_a} \sum_{(i,j) \in \Omega_a} \|\mathbf{L}_A(i,:)\|_2 \|\mathbf{R}_A(j,:)\|_2 \right) \left(\sum_{(i,j) \in \Omega_a} \|\mathbf{L}_B(i,:)\|_2 \|\mathbf{R}_B(j,:)\|_2 \right) \\ & \leq \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{L}_B\|_F \|\mathbf{R}_B\|_F, \end{aligned}$$

where the second and the inequalities follows from the Cauchy–Schwarz inequality and the fact that $|\Omega_a| = \varsigma_a$. Thus

$$|r_k| \leq 2 \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{L}_B\|_F \|\mathbf{R}_B\|_F.$$

Similarly, we have $|r_k| \leq 2 \|\mathbf{L}_A\|_F \|\mathbf{R}_A\|_F \|\mathbf{L}_B\|_{2,\infty} \|\mathbf{R}_B\|_{2,\infty}$. Also, for any $a \in [n]$, by Cauchy–Schwarz inequality and $|\Omega_a| = \varsigma_a$ we have

$$\begin{aligned}
& |[\mathcal{G}^*(\mathbf{L}_A \mathbf{R}_A^*)]_a [\mathcal{G}^*(\mathbf{L}_B \mathbf{R}_B^*)]_a| \\
& \leq \left(\sum_{(i,j) \in \Omega_a} \frac{1}{\sqrt{\zeta_a}} \|\mathbf{L}_A(i, :)\|_2 \|\mathbf{R}_A(j, :)\|_2 \right) \left(\frac{1}{\sqrt{\zeta_a}} \sum_{(i,j) \in \Omega_a} \|\mathbf{L}_B(i, :)\|_2 \|\mathbf{R}_B(j, :)\|_2 \right) \\
& \leq \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_{2,\infty} \left(\sum_{(i,j) \in \Omega_a} \frac{1}{\sqrt{\zeta_a}} \|\mathbf{R}_A(j, :)\|_2 \right) \left(\frac{1}{\sqrt{\zeta_a}} \sum_{(i,j) \in \Omega_a} \|\mathbf{R}_B(j, :)\|_2 \right) \\
& \leq \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_{2,\infty} \sqrt{\sum_{i=1}^a \|\mathbf{R}_A(i, :)\|_2^2} \sqrt{\sum_{j=1}^a \|\mathbf{R}_B(j, :)\|_2^2} \\
& \leq \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_{2,\infty} \|\mathbf{R}_A\|_F \|\mathbf{R}_B\|_F,
\end{aligned}$$

Then we have $|r_k| \leq 2 \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_{2,\infty} \|\mathbf{R}_A\|_F \|\mathbf{R}_B\|_F$. Similarly,

$$|r_k| \leq 2 \|\mathbf{L}_A\|_F \|\mathbf{L}_B\|_F \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{R}_B\|_{2,\infty}.$$

Denote

$$\zeta := \min \left\{ \|\mathbf{L}_A\|_{2,\infty} \|\mathbf{L}_B\|_F, \|\mathbf{L}_A\|_F \|\mathbf{L}_B\|_{2,\infty} \right\} \min \left\{ \|\mathbf{R}_A\|_{2,\infty} \|\mathbf{R}_B\|_F, \|\mathbf{R}_A\|_F \|\mathbf{R}_B\|_{2,\infty} \right\}$$

Then, combining all the pieces above we obtain

$$|r_k| \leq 2\zeta.$$

By using the above bounds for $|[\mathcal{G}^*(\mathbf{L}_A \mathbf{R}_A^*)]_a [\mathcal{G}^*(\mathbf{L}_B \mathbf{R}_B^*)]_a|$ and the inequality $(x+y)^2 \leq 2x^2 + 2y^2$, we further have

$$\sum_{k=1}^m \mathbb{E}[r_k^2] \leq \sum_{k=1}^m 4 \max_{a \in [n]} |[\mathcal{G}^*(\mathbf{L}_A \mathbf{R}_A^*)]_a [\mathcal{G}^*(\mathbf{L}_B \mathbf{R}_B^*)]_a|^2 \leq 4m\zeta^2.$$

By Bernstein's inequality, it then implies

$$\mathbb{P} \left(\left| \sum_{k=1}^m r_k \right| \geq t \right) \leq 2 \exp \left(\frac{-t^2/2}{4m\zeta^2 + \frac{4t}{3}\zeta} \right).$$

Letting $t = \sqrt{24np \log n} \zeta$, by the definition of r_k we then have

$$\left| \left\langle \mathcal{G} \left(\frac{1}{p} \Pi_\Omega - \mathcal{I} \right) \mathcal{G}^*(\mathbf{L}_A \mathbf{R}_A^*), \mathbf{L}_B \mathbf{R}_B^* \right\rangle \right| \leq \frac{t}{p} = \sqrt{\frac{24n \log n}{p}} \zeta$$

with probability at least $1 - 2n^{-2}$ provided $p \geq (24 \log n)/n$. \square

Proof of theorem 3. Similar to the proof of theorem 2, let

$$\tilde{L}_{k+1} := L_k - \eta \nabla_L \ell(L_k, R_k; s_{k+1}) (R_k^* R_k)^{-1}$$

we have

$$\begin{aligned} & (\tilde{L}_{k+1} Q_k - L_*) \Sigma_*^{1/2} \\ &= \left[\Delta_{L_k} - \eta \mathcal{G} \left[\frac{1}{p} \Pi_\Omega (\mathcal{G}^* (L_k R_k^*) + s_{k+1} - f) - \mathcal{G}^* (L_k R_k^*) \right] R_k (R_k^* R_k)^{-1} - \eta L_k \right] \Sigma_*^{1/2} \\ &= \Delta_{L_k} \Sigma_*^{1/2} - \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (L_k R_k^* - L_* R_*^*) R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} \\ &\quad - \frac{\eta}{p} [\mathcal{G} \Pi_\Omega (s_{k+1} - s_*)] R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} - \eta (L_k R_k^* - L_* R_*^*) R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} \\ &= (1 - \eta) \Delta_{L_k} \Sigma_*^{1/2} - \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_*^* + L_k \Delta_{R_k}^*) R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} \\ &\quad - \frac{\eta}{p} [\mathcal{G} \Pi_\Omega (s_{k+1} - s_*)] R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} - \eta L_* \Delta_{R_k}^* R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2} \\ &= \underbrace{(1 - \eta) \Delta_{L_k} \Sigma_*^{1/2} - \eta L_* \Delta_{R_k}^* R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2}}_{T_1} \\ &\quad - \underbrace{\frac{\eta}{p} [\mathcal{G} \Pi_\Omega (s_{k+1} - s_*)] R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2}}_{T_3} \\ &\quad - \underbrace{\eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_*^* + L_k \Delta_{R_k}^*) R_k (R_k^* R_k)^{-1} \Sigma_*^{1/2}}_{T_4}, \end{aligned}$$

where the second equation follows from $f = z_* + s_*$, $\mathcal{G}^*(L_* R_*^*) = z_*$ and $\mathcal{G} \mathcal{G}^*(L_* R_*^*) = \mathcal{G} z_* = L_* R_*^*$, the third equation follows from $L_k R_k^* - L_* R_*^* = \Delta_{L_k} R_*^* + L_k \Delta_{R_k}^* = \Delta_{L_k} R_*^* + L_* \Delta_{R_k}^* + \Delta_{L_k} \Delta_{R_k}^*$. Noticing that

$$\begin{aligned} & \left\| (\tilde{L}_{k+1} Q_k - L_*) \Sigma_*^{1/2} \right\|_F^2 \\ &= \|T_1\|_F^2 + \|T_3\|_F^2 + \|T_4\|_F^2 - 2\text{Re} \langle T_1, T_3 \rangle - 2\text{Re} \langle T_1, T_4 \rangle + 2\text{Re} \langle T_3, T_4 \rangle, \end{aligned} \quad (29)$$

thus we bound $\left\| (\tilde{L}_{k+1} Q_k - L_*) \Sigma_*^{1/2} \right\|_F$ term by term by the following claims:

Claim 2. For $(L_k, R_k) \in \mathcal{E}(\varepsilon_0 \sigma_r^*, c \sqrt{\frac{c_s \mu r}{n}} \sigma_1^*)$, $c \geq 1 + \varepsilon_0$, $\varepsilon_0, \eta \in (0, 1)$, provided $m \geq 24c^2(1 + c\kappa)^2 \varepsilon_0^{-2} c_s^2 \mu^2 r^2 \log n$, we have

$$\begin{aligned} |\langle T_1, T_3 \rangle| &\leq \frac{(\eta - \eta^2) c \kappa \sqrt{c_s \mu r} \xi}{2(1 - \varepsilon_0)^3} \left(3 \left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F^2 + \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F^2 \right) \\ &\quad + \frac{\eta^2 \sqrt{c_s \mu r} \xi}{2(1 - \varepsilon_0)^2} \left(\left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F^2 + 3 \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F^2 \right), \\ \|T_3\|_F &\leq \frac{2\eta c \kappa \sqrt{c_s \mu r} \xi}{(1 - \varepsilon_0)^3} \left(\left\| \Delta_{L_k} \Sigma_*^{1/2} \right\|_F + \left\| \Delta_{R_k} \Sigma_*^{1/2} \right\|_F \right), \end{aligned}$$

$$\begin{aligned}
|\langle T_1, T_4 \rangle| &\leq \left(\frac{(1-\eta)\eta\varepsilon_0}{1-\varepsilon_0} + \frac{\varepsilon_0\eta(1+\varepsilon_0)}{(1-\varepsilon_0)^4} \right) \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 \\
&\quad + \left(\frac{\varepsilon_0\eta^2}{(1-\varepsilon_0)^2} + \frac{\varepsilon_0\eta(1+\varepsilon_0)}{(1-\varepsilon_0)^4} \right) \left\| \Sigma_\star^{1/2} \Delta_{R_k}^* \right\|_F^2, \\
\|T_4\|_F &\leq \frac{\varepsilon_0\eta}{(1-\varepsilon_0)^2} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F + \frac{2\eta(1+c\kappa)}{(1-\varepsilon_0)^3} \sqrt{\frac{24c_s^2\mu^2r^2\log n}{m}} \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F.
\end{aligned}$$

Therefore, for $c \geq 1 + \varepsilon_0$, $\varepsilon_0, \eta \in (0, 1)$, provided $m \geq 24c^2(1+c\kappa)^2\varepsilon_0^{-2}c_s^2\mu^2r^2\log n$ and $\alpha \leq \frac{\varepsilon_0^2(\gamma-1)}{64(\gamma+1)(1+c\kappa)^2c_s\mu r}$, by combining all the pieces above together with $|\langle T_3, T_4 \rangle| \leq \|T_3\|_F \|T_4\|_F$ and (29) we have

$$\begin{aligned}
\left\| \left(\tilde{L}_{k+1} Q_k - L_\star \right) \Sigma_\star^{1/2} \right\|_F^2 &\leq C_1(\varepsilon_0, \eta) d_k^2 + C_2(\varepsilon_0, \eta) \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + C_3(\varepsilon_0, \eta) \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F^2 \\
&\leq (1-0.6\eta)^2 d_k^2, \quad \text{for } \varepsilon_0 = 0.01, 0 < \eta \leq 0.6,
\end{aligned}$$

where in the first inequality we have used the Cauchy-Schwarz inequality, and the constants satisfy $C_1(\varepsilon_0, \eta) \leq (1-\eta)^2 + \frac{2\varepsilon_0}{1-\varepsilon_0}\eta(1-\eta) + \frac{2\varepsilon_0+\varepsilon_0^2}{(1-\varepsilon_0)^2}\eta^2 + \frac{8\eta^2\varepsilon_0^2}{(1-\varepsilon_0)^8}$, $C_2(\varepsilon_0, \eta) \leq \frac{\varepsilon_0^2\eta^2+8\varepsilon_0\eta+2\varepsilon_0^2\eta-5\varepsilon_0\eta^2}{(1-\varepsilon_0)^4} + \frac{2(1-\eta)\eta\varepsilon_0}{1-\varepsilon_0} + \frac{8\eta^2\varepsilon_0^2}{(1-\varepsilon_0)^7}$, $C_3(\varepsilon_0, \eta) \leq \frac{4\eta\varepsilon_0+\eta^2\varepsilon_0+2\eta\varepsilon_0^2-3\eta^2\varepsilon_0^2}{(1-\varepsilon_0)^4} + \frac{2\varepsilon_0\eta^2}{(1-\varepsilon_0)^2} + \frac{12\eta^2\varepsilon_0^2}{(1-\varepsilon_0)^7}$. Similarly, let $\tilde{R}_{k+1} := R_k - \eta \nabla_{R\ell}(L_k, R_k; s_{k+1}) (R_k^* R_k)^{-1}$, we can use the same strategy to show $\|(\tilde{R}_{k+1} Q_k^* - R_\star) \Sigma_\star^{1/2}\|_F^2 \leq (1-0.6\eta)^2 d_k^2$. Thus we have

$$\text{dist}(\tilde{L}_{k+1}, \tilde{R}_{k+1}; L_\star, R_\star) \leq (1-0.6\eta) d_k.$$

By lemma 7, we conclude the statement in theorem 3. Finally, we finish the proof by proving claim 2. \square

Proof of claim 2. The proof of the upper bound for $|\langle T_1, T_3 \rangle|$ and $\|T_3\|_F$ are similar to $|\langle T_1, T_2 \rangle|$ and $\|T_2\|_F$ as in claim 1, respectively. For $|\langle T_1, T_4 \rangle|$, we have:

$$\begin{aligned}
\langle T_1, T_4 \rangle &= \left\langle (1-\eta) \Delta_{L_k} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_\star^*) R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2} \right\rangle \\
&\quad + \left\langle (1-\eta) \Delta_{L_k} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (L_{\natural} \Delta_{R_k}^*) R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2} \right\rangle \\
&\quad - \left\langle \eta L_\star \Delta_{R_k}^* R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_\star^*) R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2} \right\rangle \\
&\quad - \left\langle \eta L_\star \Delta_{R_k}^* R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (L_{\natural} \Delta_{R_k}^*) R_{\natural} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2} \right\rangle
\end{aligned}$$

Noticing $R_{\natural} = R_\star + \Delta_{R_k}$, let

$$\begin{aligned}
L_{A_3} &:= \Delta_{L_k} \Sigma_\star^{1/2}, \quad R_{A_3} := R_\star \Sigma_\star^{-1/2}, \\
L_{A_4} &:= L_{A_3}, \quad R_{A_4} := \Delta_{R_k} (R_{\natural}^* R_{\natural})^{-1} \Sigma_\star^{1/2},
\end{aligned} \tag{30}$$

by lemmas 4 and 12 we then have

$$\begin{aligned}
& \left| \left\langle (1-\eta) \Delta_{L_k} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_\star^*) (R_\star + \Delta_{R_k}) (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2} \right\rangle \right| \\
& \leq (1-\eta) \eta \varepsilon_0 \left\| \Delta_{L_k} \Sigma_\star (R_\star^* R_\star)^{-1} R_\star^* \right\|_F \left\| \Delta_{L_k} R_\star^* \right\|_F \\
& \quad + (1-\eta) \eta \left| \left\langle \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} \Sigma_\star^{1/2} (R_\star \Sigma_\star^{-1/2})^*), \Delta_{L_k} \Sigma_\star (R_\star^* R_\star)^{-1} \Delta_{R_k}^* \right\rangle \right| \\
& \leq \frac{(1-\eta) \eta \varepsilon_0}{1-\varepsilon_0} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F^2 + (1-\eta) \eta \sqrt{\frac{24n \log n}{p}} \|L_{A_3}\|_F \|R_{A_3}\|_{2,\infty} \|L_{A_4}\|_{2,\infty} \|R_{A_4}\|_F \\
& \leq \frac{(1-\eta) \eta \varepsilon_0}{1-\varepsilon_0} \left\| \Delta_{L_k} \Sigma_\star^{\frac{1}{2}} \right\|_F^2 + \frac{(1+c)(\eta-\eta^2)}{(1-\varepsilon_0)^3} \sqrt{\frac{24c_s^2 \mu^2 \kappa^2 r^2 \log n}{m}} \left\| \Delta_{L_k} \Sigma_\star^{\frac{1}{2}} \right\|_F \left\| \Delta_{R_k} \Sigma_\star^{\frac{1}{2}} \right\|_F,
\end{aligned}$$

where the last step follows from (21d) and (22), $\|R_{A_3}\|_{2,\infty} = \|U_\star\|_{2,\infty}$, and

$$\|R_{A_4}\|_F \leq \|\Delta_{R_k} \Sigma_\star^{1/2}\|_F \|\Sigma_\star^{-1}\|_2 \|\Sigma_\star^{1/2} (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2}\|_2 \leq \frac{1}{\sigma_r^* (1-\varepsilon_0)^2} \|\Delta_{R_k} \Sigma_\star^{1/2}\|_F.$$

Similarly, by lemma 12 and the definition of L_{A_2} , L_{A_3} , R_{A_2} , R_{A_4} in (27) and (30), we have

$$\begin{aligned}
& \left| \left\langle \eta L_\star \Delta_{R_k}^* R_\star (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_\star^*) R_\star (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2} \right\rangle \right| \\
& = \left| \left\langle \eta (L_\star \Sigma_\star^{-\frac{1}{2}}) \Sigma_\star^{\frac{1}{2}} \Delta_{R_k}^* R_\star (R_\star^* R_\star)^{-1} \Sigma_\star (R_\star^* R_\star)^{-1} R_\star^*, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} R_\star^*) \right\rangle \right| \\
& \leq \eta^2 \sqrt{\frac{24n \log n}{p}} \|L_{A_2}\|_{2,\infty} \|L_{A_3}\|_F \left\| R_\star \Sigma_\star^{-1/2} \right\|_{2,\infty} \|R_{A_2}\|_F \\
& \leq \frac{\eta^2}{(1-\varepsilon_0)^2} \sqrt{\frac{24c_s^2 \mu^2 r^2 \log n}{m}} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F,
\end{aligned}$$

and

$$\begin{aligned}
& \left| \left\langle (1-\eta) \Delta_{L_k} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (L_\star \Delta_{R_k}^*) R_\star (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2} \right\rangle \right| \\
& = \left| \left\langle (1-\eta) \Delta_{L_k} \Sigma_\star (R_\star^* R_\star)^{-1} \Sigma_\star^{1/2} (R_\star \Sigma_\star^{-1/2})^*, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (L_\star \Delta_{R_k}^*) \right\rangle \right| \\
& \leq (1-\eta) \eta \sqrt{\frac{24n \log n}{p}} \|L_{A_1}\|_F \left\| L_\star \Sigma_\star^{-1/2} \right\|_{2,\infty} \left\| R_\star \Sigma_\star^{-1/2} \right\|_{2,\infty} \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F \\
& \leq \frac{(1-\eta) \eta}{(1-\varepsilon_0)^4} \sqrt{\frac{24c^4 c_s^2 \kappa^4 \mu^2 r^2 \log n}{m}} \left\| \Delta_{L_k} \Sigma_\star^{1/2} \right\|_F \left\| \Delta_{R_k} \Sigma_\star^{1/2} \right\|_F.
\end{aligned}$$

By triangle inequality and lemmas 4 and 12 we have

$$\begin{aligned}
& \left| \left\langle \eta \mathbf{L}_\star \Delta_{\mathbf{R}_k}^* \mathbf{R}_{\mathfrak{h}} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \Sigma_\star^{1/2}, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\mathbf{L}_{\mathfrak{h}} \Delta_{\mathbf{R}_k}^*) \mathbf{R}_{\mathfrak{h}} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \Sigma_\star^{1/2} \right\rangle \right| \\
& \leq \left| \left\langle \eta \mathbf{L}_\star \Sigma_\star^{-1/2} \mathbf{R}_{A_2}^*, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\mathbf{L}_\star \Delta_{\mathbf{R}_k}^*) \right\rangle \right| \\
& \quad + \left| \left\langle \eta (\mathbf{L}_\star \Sigma_\star^{-1/2}) \mathbf{R}_{A_2}^*, \eta \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} \Delta_{\mathbf{R}_k}^*) \right\rangle \right| \\
& \leq \varepsilon_0 \eta^2 \|\mathbf{U}_\star \mathbf{R}_{A_2}^*\|_{\text{F}} \|\mathbf{L}_\star \Delta_{\mathbf{R}_k}^*\|_{\text{F}} \\
& \quad + \eta^2 \sqrt{\frac{24n \log n}{p}} \|\mathbf{L}_\star \Sigma_\star^{-1/2}\|_{2,\infty} \|\Delta_{L_k} \Sigma_\star^{1/2}\|_{\text{F}} \|\Delta_{\mathbf{R}_k} \Sigma_\star^{-1/2}\|_{2,\infty} \|\mathbf{R}_{A_2}\|_{\text{F}} \\
& \leq \frac{\varepsilon_0 \eta^2}{(1-\varepsilon_0)^2} \|\Sigma_\star^{1/2} \Delta_{\mathbf{R}_k}^*\|_{\text{F}}^2 + \frac{\eta^2 (1+c\kappa)}{(1-\varepsilon_0)^3} \sqrt{\frac{24c_s^2 \mu^2 r^2 \log n}{m}} \|\Delta_{L_k} \Sigma_\star^{1/2}\|_{\text{F}} \|\Delta_{\mathbf{R}_k} \Sigma_\star^{1/2}\|_{\text{F}},
\end{aligned}$$

where in the third inequality follows from $\|\mathbf{U}_\star \mathbf{R}_{A_2}^*\|_{\text{F}} \leq \|\mathbf{R}_{A_2}\|_{\text{F}}$, $\|\mathbf{L}_\star \Delta_{\mathbf{R}_k}^*\|_{\text{F}} = \|\mathbf{U}_\star \Sigma_\star^{1/2} \Delta_{\mathbf{R}_k}^*\|_{\text{F}} \leq \|\Sigma_\star^{1/2} \Delta_{\mathbf{R}_k}^*\|_{\text{F}}$ and (28), (22). Then, combining all the pieces, for $c \geq 1 + \varepsilon_0$, $\varepsilon_0, \eta \in (0, 1)$, provided $m \geq 24c^2(1+c\kappa)^2\varepsilon_0^{-2}c_s^2\mu^2r^2\log n$, we then have

$$\begin{aligned}
|\langle \mathbf{T}_1, \mathbf{T}_4 \rangle| & \leq \frac{(1-\eta)\eta\varepsilon_0}{1-\varepsilon_0} \|\Delta_{L_k} \Sigma_\star^{1/2}\|_{\text{F}}^2 + \frac{\varepsilon_0 \eta^2}{(1-\varepsilon_0)^2} \|\Sigma_\star^{1/2} \Delta_{\mathbf{R}_k}^*\|_{\text{F}}^2 \\
& \quad + \frac{2\varepsilon_0 \eta (1+\varepsilon_0)}{(1-\varepsilon_0)^4} \|\Delta_{L_k} \Sigma_\star^{1/2}\|_{\text{F}} \|\Delta_{\mathbf{R}_k} \Sigma_\star^{1/2}\|_{\text{F}} \\
& \leq \left(\frac{(1-\eta)\eta\varepsilon_0}{1-\varepsilon_0} + \frac{\varepsilon_0 \eta (1+\varepsilon_0)}{(1-\varepsilon_0)^4} \right) \|\Delta_{L_k} \Sigma_\star^{1/2}\|_{\text{F}}^2 \\
& \quad + \left(\frac{\varepsilon_0 \eta^2}{(1-\varepsilon_0)^2} + \frac{\varepsilon_0 \eta (1+\varepsilon_0)}{(1-\varepsilon_0)^4} \right) \|\Sigma_\star^{1/2} \Delta_{\mathbf{R}_k}^*\|_{\text{F}}^2.
\end{aligned}$$

For $\|\mathbf{T}_4\|_{\text{F}}$: Using the variational representation of the Frobenius norm, for some $\mathbf{L}_A \in \mathbb{C}^{n_1 \times r}$, $\|\mathbf{L}_A\|_{\text{F}} = 1$ we have

$$\begin{aligned}
& \|\mathbf{T}_4\|_{\text{F}} \\
& = \eta \left| \left\langle \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} \mathbf{R}_\star^* + \mathbf{L}_{\mathfrak{h}} \Delta_{\mathbf{R}_k}^*) \mathbf{R}_{\mathfrak{h}} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \Sigma_\star^{1/2}, \mathbf{L}_A \right\rangle \right| \\
& \leq \eta \left| \left\langle \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} \mathbf{R}_\star^*), \mathbf{L}_A \Sigma_\star^{1/2} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \mathbf{R}_\star^* \right\rangle \right| \\
& \quad + \eta \left| \left\langle \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\Delta_{L_k} \mathbf{R}_\star^*), \mathbf{L}_A \Sigma_\star^{1/2} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \Delta_{\mathbf{R}_k}^* \right\rangle \right| \\
& \quad + \eta \left| \left\langle \left(\frac{1}{p} \mathcal{G} \Pi_\Omega \mathcal{G}^* - \mathcal{G} \mathcal{G}^* \right) (\mathbf{L}_{\mathfrak{h}} \Delta_{\mathbf{R}_k}^*), \mathbf{L}_A \Sigma_\star^{1/2} (\mathbf{R}_{\mathfrak{h}}^* \mathbf{R}_{\mathfrak{h}})^{-1} \mathbf{R}_\star^* \right\rangle \right|
\end{aligned}$$

$$\begin{aligned}
&\leq \varepsilon_0 \eta \left\| \Delta_{L_k} \Sigma_{\star}^{1/2} V_{\star} \right\|_F \left\| L_A \Sigma_{\star}^{1/2} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} V_{\star} \right\|_F \\
&\quad + \eta \sqrt{\frac{24n \log n}{p}} \left(\left\| \Delta_{L_k} \Sigma_{\star}^{1/2} \right\|_{2,\infty} \|L_A\|_F \left\| R_{\star} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| \Delta_{R_k} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} \right\|_F \right. \\
&\quad \left. + \left\| L_{\natural} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \left\| L_A \Sigma_{\star}^{1/2} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} \right\|_F \left\| \Delta_{R_k} \Sigma_{\star}^{1/2} \right\|_F \left\| R_{\star} \Sigma_{\star}^{-1/2} \right\|_{2,\infty} \right) \\
&\leq \frac{\varepsilon_0 \eta}{(1-\varepsilon_0)^2} \left\| \Delta_{L_k} \Sigma_{\star}^{1/2} \right\|_F + \frac{2\eta(1+c\kappa)}{(1-\varepsilon_0)^3} \sqrt{\frac{24c_s^2 \mu^2 r^2 \log n}{m}} \left\| \Delta_{R_k} \Sigma_{\star}^{1/2} \right\|_F
\end{aligned}$$

where the second inequality follows from lemmas 4 and 12, and the last inequality follows from (21c), (21d), (22) and (28), as well as the following inequalities: $\left\| L_A \Sigma_{\star}^{1/2} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} V_{\star} \right\|_F \leq \left\| L_A \Sigma_{\star}^{1/2} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} \right\|_F \leq \frac{1}{(1-\varepsilon_0)^2}$ and $\left\| \Delta_{R_k} (R_{\natural}^* R_{\natural})^{-1} \Sigma_{\star}^{1/2} \right\|_F \leq \frac{1}{(1-\varepsilon_0)^2} \left\| \Delta_{R_k} \Sigma_{\star}^{-1/2} \right\|_F \leq \frac{1}{\sigma_r^* (1-\varepsilon_0)^2} \left\| \Delta_{R_k} \Sigma_{\star}^{1/2} \right\|_F$. \square

6. Conclusion

In this paper, we introduce a novel non-convex algorithm, dubbed Hankel Structured Newton-Like Descent, for accelerating ill-conditioned robust Hankel recovery problems. The linear convergence guarantee has been established for the proposed algorithm; especially, its convergence rate is independent of the condition number of the underlying Hankel matrix. The empirical advantages of the proposed algorithm have been verified via extensive experiments on synthetic and real datasets.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

HanQin Cai is partially supported by the National Science Foundation (DMS 2304489). Xiliang Lu is partially supported by the National Science Foundation of China (Nos. 12371424 and U24A2002), and the Natural Science Foundation of Hubei Province, China (No. 2024AFE006). Juntao You is partially supported by the National Science Foundation of China (No. 12401121) and the Shenzhen Municipal Stability Support Program Project (No. 8960117/0123).

References

- [1] Aldroubi A, Huang L, Krishtal I, Ledeczi A, Lederman R R and Volgyesi P 2018 Dynamical sampling with additive random noise *Samp. Theory Signal Image Process.* **17** 153–82
- [2] Cadzow J A 1988 Signal enhancement—a composite property mapping algorithm *IEEE Trans. Acoust., Speech, Signal. Process.* **36** 49–62
- [3] Cai H, Cai J-F, Wang T and Yin G 2021 Accelerated structured alternating projections for robust spectrally sparse signal recovery *IEEE Trans. Signal Process.* **69** 809–21

- [4] Cai H, Cai J-F and You J 2023 Structured gradient descent for fast robust low-rank Hankel matrix completion *SIAM J. Sci. Comput.* **45** A1172–98
- [5] Cai H, Hamm K, Huang L and Needell D 2021 Robust CUR decomposition: theory and imaging applications *SIAM J. Imaging Sci.* **14** 1472–503
- [6] Cai J-F, Wang T and Wei K 2018 Spectral compressed sensing via projected gradient descent *SIAM J. Optim.* **28** 2625–53
- [7] Cai J-F, Wang T and Wei K 2019 Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion *Appl. Comput. Harmon. Anal.* **46** 94–121
- [8] Candès E J, Li X, Ma Y and Wright J 2011 Robust principal component analysis? *J. ACM* **58** 1–37
- [9] Candès E J and Recht B 2009 Exact matrix completion via convex optimization *Found. Comput. Math.* **9** 717–72
- [10] Chandrasekaran V, Sanghavi S, Parrilo P A and Willsky A S 2011 Rank-sparsity incoherence for matrix decomposition *SIAM J. Optim.* **21** 572–96
- [11] Chen K and Sacchi M D 2015 Robust reduced-rank filtering for erratic seismic noise attenuation *Geophysics* **80** V1–V11
- [12] Chen X, Cheng Z, Cai H, Saunier N and Sun L 2024 Laplacian convolutional representation for traffic time series imputation *IEEE Trans. Knowl. Data Eng.* **36** 6490–502
- [13] Chen Y, Jalali A, Sanghavi S and Caramanis C 2013 Low-rank matrix recovery from errors and erasures *IEEE Trans. Inf. Theory* **59** 4324–37
- [14] Chen Y and Chi Y 2014 Robust spectral compressed sensing via structured matrix completion *IEEE Trans. Inf. Theory* **60** 6576–601
- [15] Cheng J, Sacchi M and Gao J 2019 Computational efficient multidimensional singular spectrum analysis for prestack seismic data reconstruction *Geophysics* **84** V111–9
- [16] Cherapanamjeri Y, Gupta K and Jain P 2017 Nearly optimal robust matrix completion *Int. Conf. on Machine Learning* pp 797–805
- [17] Chiron L, van Agthoven M A, Kieffer B, Rolando C and Delsuc M-A 2014 Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry *Proc. Natl Acad. Sci.* **111** 1385–90
- [18] Fazel M, Pong T K, Sun D and Tseng P 2013 Hankel matrix rank minimization with applications to system identification and realization *SIAM J. Matrix Anal. Appl.* **34** 946–77
- [19] Giampouras P, Cai H and Vidal R 2025 Guarantees of a preconditioned subgradient algorithm for overparameterized asymmetric low-rank matrix recovery *Int. Conf. on Machine Learning*
- [20] Gillard J and Usevich K 2018 Structured low-rank matrix completion for forecasting in time series analysis *Int. J. Forecast.* **34** 582–97
- [21] Hamm K, Meskini M and Cai H 2022 Riemannian CUR decompositions for robust principal component analysis *Topological, Algebraic and Geometric Learning Workshops 2022* (PMLR) pp 152–60
- [22] Holland D J, Bostock M J, Gladden L F and Nietlispach D 2011 Fast multidimensional NMR spectroscopy using compressed sensing *Angew. Chem. Int. Ed.* **50** 6548–51
- [23] Hu Q, Noll R J, Li H, Makarov A, Hardman M and Cooks R G 2005 The orbitrap: a new mass spectrometer *J. Mass Spectrom.* **40** 430–43
- [24] Jacob M, Mani M P and Ye J C 2020 Structured low-rank algorithms: theory, magnetic resonance applications and links to machine learning *IEEE Signal Process. Mag.* **37** 54–68
- [25] Jin K H, Lee D and Ye Y 2016 A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix *IEEE Trans. Comput. Imaging* **2** 480–95
- [26] Li J, Cai J-F and Zhao H 2020 Robust inexact alternating optimization for matrix completion with outliers *J. Comput. Math.* **38** 337–54
- [27] Li W and Liao W 2021 Stable super-resolution limit and smallest singular value of restricted Fourier matrices *Appl. Comput. Harmon. Anal.* **51** 118–56
- [28] Liao W and Fannjiang A 2016 MUSIC for single-snapshot spectral estimation: Stability and super-resolution *Appl. Comput. Harmon. Anal.* **40** 33–67
- [29] Netrapalli P, Niranjan U N, Sanghavi S, Anandkumar A and Jain P 2014 Non-convex robust PCA *Advances in Neural Information Processing Systems* pp 1107–15
- [30] Qu X, Mayzel M, Cai J-F, Chen Z and Orekhov V 2015 Accelerated NMR spectroscopy with low-rank reconstruction *Angew. Chem. Int. Ed.* **54** 852–4
- [31] Recht B 2011 A simpler approach to matrix completion *J. Mach. Learn. Res.* **12** 3413–30
- [32] Schmidt R 1986 Multiple emitter location and signal parameter estimation *IEEE Trans. Antennas Propag.* **34** 276–80

- [33] Smith R S 2014 Frequency domain subspace identification using nuclear norm minimization and Hankel matrix realizations *IEEE Trans. Autom. Control* **59** 2886–96
- [34] Sun W, Zhou Y, Xiang J, Chen B and Feng W 2021 Hankel matrix-based condition monitoring of rolling element bearings: an enhanced framework for time-series analysis *IEEE Trans. Instrum. Meas.* **70** 1–10
- [35] Tong T, Ma C and Chi Y 2021 Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent *J. Mach. Learn. Res.* **22** 150–1
- [36] Tropp J A, Laska J N, Duarte M F, Romberg J K and Baraniuk R G 2009 Beyond Nyquist: Efficient sampling of sparse bandlimited signals *IEEE Trans. Inf. Theory* **56** 520–44
- [37] Wang C, Zhu Z, Gu H, Wu X and Liu S 2018 Hankel low-rank approximation for seismic noise attenuation *IEEE Trans. Geosci. Remote Sens.* **57** 561–73
- [38] Wei K, Cai J-F, Chan T F and Leung S 2020 Guarantees of Riemannian optimization for low rank matrix completion *Inverse Problems Imaging* **14** 233
- [39] Xi Y and Rocke D M 2008 Baseline correction for NMR spectroscopic metabolomics data analysis *BMC Bioinform.* **9** 1–10
- [40] Yang Z, Li J, Stoica P and Xie L 2018 Sparse methods for direction-of-arrival estimation *Academic Library in Signal Processing* vol 7 (Elsevier) pp 509–81
- [41] Yi X, Park D, Chen Y and Caramanis C 2016 Fast algorithms for robust PCA via gradient descent *Advances in Neural Information Processing Systems* pp 4152–60
- [42] Zhang S and Wang M 2019 Correction of corrupted columns through fast robust Hankel matrix completion *IEEE Trans. Signal Process.* **67** 2580–94
- [43] Zhang T and Yang Y 2018 Robust PCA by manifold optimization *J. Mach. Learn. Res.* **19** 3101–39