COMMUN. MATH. SCI.
Vol. 21, No. 2, pp. 581–590


*FAST COMMUNICATION*

# BLIND SUPER-RESOLUTION OF POINT SOURCES VIA FAST ITERATIVE HARD THRESHOLDING*

ZENGYING ZHU†, JINCHI CHEN‡, AND WEIGUO GAO§

**Abstract.** In this work, we develop a provable fast algorithm for blind super-resolution based on the low rank structure of vectorized Hankel matrix associated with the target matrix. Theoretical results show that the proposed method converges to the ground truth with linear convergence rate. Numerical experiments are also conducted to illustrate the linear convergence and effectiveness of the proposed approach.

**Keywords.** Blind Super-resolution; Low Rank Matrix Recovery; Fast Iterative Hard Threshoding; Vectorized Hankel Lift.

**AMS subject classifications.** 65F55; 15A83; 90C26; 94A12.

## 1. Introduction

Blind super-resolution is the problem of estimating $\{\tau_k, d_k, \boldsymbol{g}_k\}_{k=1}^r$ from the observations

$$\boldsymbol{y}[j] = \sum_{k=1}^{r} d_k e^{-2\pi\imath\tau_k(j-1)} \boldsymbol{g}_k[j], \text{ for } j = 1,\dots,n, \tag{1.1}$$

where $\{d_k, \tau_k\}_{k=1}^r$ are unknown coefficients and frequencies associated with the complex exponentials, and $\{\boldsymbol{g}_k\}_{k=1}^r$ are unknown point spread functions. It has many applications, such as computational photography [12], multi-user communication system [15] and seismic data analysis [17]. It is worth noting that the measurement model (1.1) also includes blind sparse spike deconvolution problem [9, 14] in which the point spread function is shared among all point sources. In particular, when the knowledge of $\{\boldsymbol{g}_k\}_{k=1}^r$ is available, blind super-resolution reduces to the super-resolution problem [2, 10].

Since the number of unknowns in (1.1) is larger than the number of samples, it is an ill-posed problem without any additional constraints on $\boldsymbol{g}_k$. To alleviate this issue, it is typically assumed that $\{\boldsymbol{g}_k\}_{k=1}^r$ belong to a known low-dimensional subspace spanned by the columns of $\boldsymbol{B} \in \mathbb{C}^{n\times s}$ with $s < n$, i.e.,

$$\boldsymbol{g}_k = \boldsymbol{B}\boldsymbol{h}_k, \tag{1.2}$$

where $\boldsymbol{h}_k \in \mathbb{C}^s$ represents the unknown coefficient of $\boldsymbol{g}_k$ in the subspace [7, 9, 21]. Under the subspace assumption (1.2) and applying the lift technique [1, 14], blind super-resolution can be cast as the problem of recovering the matrix $\boldsymbol{X}^\natural = \sum_{k=1}^r d_k \boldsymbol{h}_k \boldsymbol{a}_{\tau_k}^{\mathsf{H}} \in \mathbb{C}^{s\times n}$ from a set of linear measurements

$$\boldsymbol{y}[j] = \left\langle \boldsymbol{b}_j \boldsymbol{e}_j^{\mathsf{H}}, \boldsymbol{X}^\natural \right\rangle, \quad j = 0,\dots,n-1, \tag{1.3}$$

*Received: January 25, 2022; Accepted (in revised form): November 04, 2022. Communicated by Lexing Ying.

†School of Mathematical Sciences, Fudan University, Shanghai, China (zengyingzhu@fudan.edu.cn).

‡School of Data Science, Fudan University, Shanghai, China (jcchen.phys@gmail.com).

§School of Mathematical Sciences and School of Data Science, Fudan University, Shanghai, China (wggao@fudan.edu.cn).

where $\boldsymbol{a}_\tau = \left[1\ e^{-2\pi\imath\tau\cdot 1}\ \cdots\ e^{-2\pi\imath\tau\cdot(n-1)}\right]^{\mathsf{H}}$, $\boldsymbol{b}_j \in \mathbb{C}^s$ is the $j$-th column of $\boldsymbol{B}^{\mathsf{H}}$, $\boldsymbol{e}_j$ is the $(j+1)$-th standard basis of $\mathbb{R}^n$, and the inner product between two matrices is given by $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathrm{trace}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{B})$. The measurement model (1.3) can be rewritten as a more compact form

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}^\natural), \tag{1.4}$$

where $\mathcal{A}: \mathbb{C}^{s\times n} \to \mathbb{C}^m$ is the linear operator. When the data matrix $\boldsymbol{X}^\natural$ is recovered, the frequencies $\{\tau_k\}_{k=1}^r$ can be extracted from it via spatial smoothing MUSIC [11] and coefficients $\{d_k, \boldsymbol{h}_k\}_{k=1}^r$ can be estimated by solving an over–determined linear system [21]. Therefore in this work we focus on the problem of estimating $\boldsymbol{X}^\natural$ from its linear measurements (1.4).

Let $\mathcal{H}$ be the vectorized Hankel lift operator [7,22] which maps a matrix $\boldsymbol{X} \in \mathbb{C}^{s\times n}$ into an $sn_1 \times n_2$ matrix,

$$\mathcal{H}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{n_2} \\ \boldsymbol{x}_2 & \boldsymbol{x}_3 & \cdots & \boldsymbol{x}_{n_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_{n_1} & \boldsymbol{x}_{n_1+1} & \cdots & \boldsymbol{x}_n, \end{bmatrix} \in \mathbb{C}^{sn_1\times n_2} \tag{1.5}$$

and $\boldsymbol{x}_i \in \mathbb{C}^s$ is the $i$-th column of $\boldsymbol{X}$ and $n_1 + n_2 = n+1$. It has been shown that the rank of $\mathcal{H}(\boldsymbol{X}^\natural)$ is at most $r$ and thus it is a low rank matrix when $r \ll \min(sn_1, n_2)$ [7]. To recover $\boldsymbol{X}^\natural$, we seek a rank-$r$ vectorized Hankel matrix consistent with the linear measurements (1.4) by solving the following low rank vectorized Hankel matrix sensing problem

$$\min_{\boldsymbol{X}\in\mathbb{C}^{s\times n}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 \text{ s.t. } \mathrm{rank}(\mathcal{H}(\boldsymbol{X})) = r, \tag{1.6}$$

Inspired by [5,22], we develop a non-convex algorithm called Fast Iterative Hard Thresholding via Vectorized Hankel Lift (FIHT-VHL) to solve (1.6). We apply the low rank structure of the vectorized Hankel matrix associated with the target matrix while Cai et al. [5] investigated the low rank Hankel matrix recovery problem in the context of spectrally sparse recovery. Moreover, the measurement model in this work is different from that in [22]. Therefore the theoretical guarantee in [22] is not applicable for the blind super-resolution setting. We show that FIHT-VHL is able to converge linearly to the unknown data matrix with high probability if the number of measurements is of the order $\mathcal{O}(s^2r^2\log^2(sn))$ and the algorithm is properly initialized.

**Related works:** The problem of recovering $\boldsymbol{X}^\natural$ from (1.6) is also studied in [7,16]. Chen et al. [7] developed a nuclear norm minimization method based on the vectorized Hankel lift. Recently, Mao and Chen [16] developed a Projected Gradient Descent via Vectorized Hankel Lift (PGD-VHL) method for the problem (1.6). Their theoretical results show that the matrix $\boldsymbol{X}^\natural$ can be exactly recovered from $\mathcal{O}(s^2r^2\log^2(sn))$ measurements.

Another line of related work addresses the problem of recovering $\boldsymbol{X}^\natural$ from (1.4) [13,18,21]. More specifically, Yang et al. [21] proposed an atomic norm minimization method to recover the data matrix. Their theoretical result shows that $\mathcal{O}(sr\log n)$ measurements are sufficient to guarantee exact recovery of $\boldsymbol{X}^\natural$ with high probability under certain incoherence condition. The stable analysis of blind super-resolution is also provided in [13,18]. The approaches developed in [13,18,21] are based on convex relaxation and the equivalent semi-definite programmings are computationally inefficient for large-scale problems.

**Organization:** The remainder of this paper is organized as follows. In Section 2, we will introduce FIHT-VHL algorithm. In Section 3, we will introduce two assumptions and establish our main result. The performance of FIHT-VHL is evaluated by numerical experiments in Section 4. In Section 5, we give the detailed proofs for main result. We close with a conclusion in Section 6.

**Notations and preliminaries:** Throughout this work, we use bold lowercase letters, bold uppercase letters and calligraphic letters for vectors, matrices and operators, respectively. The letter $\mathcal{I}$ denotes the identity operator. We use $\boldsymbol{x}[i]$ to denote the $i$-th entry of vector $\boldsymbol{x}$ and $\boldsymbol{X}[j,k]$ to denote the $(j,k)$-th entry of matrix $\boldsymbol{X}$. Additionally, we use $\boldsymbol{Z}[i:j,k]$ to denote a $j-i+1$ vector with entries $\boldsymbol{Z}[i,k],\dots,\boldsymbol{Z}[j,k]$. The adjoint of $\mathcal{H}$, denoted by $\mathcal{H}^*$, is a linear mapping from $sn_1\times n_2$ matrices to matrices of size $s\times n$. In particular, for any matrix $\boldsymbol{Z}\in\mathbb{C}^{sn_1\times n_2}$, the $i$-th column of $\mathcal{H}^*(\boldsymbol{Z})$ is given by

$$\mathcal{H}^*(\boldsymbol{Z})\boldsymbol{e}_i=\sum_{(j,k)\in\mathcal{W}_i}\boldsymbol{z}_{j,k},$$

where $\boldsymbol{z}_{j,k}=\boldsymbol{Z}[js:(j+1)s-1,k]$ and $\mathcal{W}_i$ is the set

$$\{(j,k)\,|\,j+k=i,0\le j\le n_1-1,0\le k\le n_2-1\}.$$

Let $\mathcal{D}:\mathbb{C}^{s\times n}\to\mathbb{C}^{s\times n}$ be an operator such that

$$\mathcal{D}(\boldsymbol{X})=\boldsymbol{X}\,\mathrm{diag}\left(\sqrt{w_0},\dots,\sqrt{w_{n-1}}\right)$$

for any $\boldsymbol{X}$, where the scalar $w_i$ is defined as the number of $\mathcal{W}_i$ for $i=0,\dots,n-1$. The Moore-Penrose pseudoinverse of $\mathcal{H}$ is given by $\mathcal{H}^\dagger=\mathcal{D}^{-2}\mathcal{H}^*$ which satisfies $\mathcal{H}^\dagger\mathcal{H}=\mathcal{I}$ [7]. The adjoint of the operator $\mathcal{A}(\cdot)$, denoted by $\mathcal{A}^*(\cdot)$, is defined as $\mathcal{A}^*(\boldsymbol{y})=\sum_{j=0}^{n-1}\boldsymbol{y}[j]\boldsymbol{b}_j\boldsymbol{e}_j^{\mathsf{H}}$. Notice that $\mathcal{A}$ and $\mathcal{A}^*$ satisfy $\|\mathcal{A}\mathcal{A}^*-\mathcal{I}\|\le s\mu_0$ and $\|\mathcal{A}\|\le\sqrt{s\mu_0}$, whose proof is provided in [7, Lemma 3.2]. Define $\mathcal{G}=\mathcal{H}\mathcal{D}^{-1}$. Then the adjoint of $\mathcal{G}$, denoted by $\mathcal{G}^*$, is given by $\mathcal{G}^*=\mathcal{D}^{-1}\mathcal{H}^*$. Additionally, $\mathcal{G}$ and $\mathcal{G}^*$ obey that $\mathcal{G}^*\mathcal{G}=\mathcal{I}$, $\|\mathcal{G}\|=1$ and $\|\mathcal{G}^*\|\le 1$.

Let $\boldsymbol{Z}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}\in\mathbb{C}^{sn_1\times n_2}$ be the compact singular value decomposition of a rank-$r$ matrix, where $\boldsymbol{U}\in\mathbb{C}^{sn_1\times r}$, $\boldsymbol{V}\in\mathbb{C}^{n_2\times r}$ and $\boldsymbol{\Sigma}\in\mathbb{R}^{r\times r}$. It is known that the tangent space of the fixed rank-$r$ matrix manifold at $\boldsymbol{Z}$ is given by [19]

$$\mathfrak{T}=\left\{\boldsymbol{U}\boldsymbol{N}^{\mathsf{H}}+\boldsymbol{M}\boldsymbol{V}^{\mathsf{H}}:\boldsymbol{M}\in\mathbb{C}^{sn_1\times r},\boldsymbol{N}\in\mathbb{C}^{n_2\times r}\right\}.$$

Given any matrix $\boldsymbol{W}\in\mathbb{C}^{sn_1\times n_2}$, the projection of $\boldsymbol{W}$ onto $\mathfrak{T}$ can be computed using the formula [19]

$$\mathcal{P}_{\mathfrak{T}}(\boldsymbol{W})=\boldsymbol{U}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{W}+\boldsymbol{W}\boldsymbol{V}\boldsymbol{V}^{\mathsf{H}}-\boldsymbol{U}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{V}\boldsymbol{V}^{\mathsf{H}}.$$

## 2. Fast Iterative Hard Thresholding via Vectorized Hankel Lift

We develop a fast iterative hard thresholding method for the problem (1.6), which is summarized in Algorithm **1**. The initial guess was obtained with the spectrum method. In the $t$-th iteration of FIHT-VHL, the current estimate $\boldsymbol{X}^t$ is first updated along the gradient descent direction of the objective function in (1.6). Then, the vectorized Hankel matrix corresponding to the update is formed via the application of the vectorized Hankel lift operator $\mathcal{H}$, followed by a projection operator $\mathcal{P}_{\mathfrak{T}_t}$ onto the $\mathfrak{T}_t$ space. After that, it imposes a hard thresholding operator $\mathcal{T}_r$ to $\boldsymbol{W}^t$ by truncated SVD process.

Finally, it applies $\mathcal{H}^\dagger$ on the low rank matrix $\boldsymbol{Z}^{t+1}$. Indeed, FIHT-VHL algorithm can be efficiently implemented. The authors in [7] show that $\mathcal{H}(\boldsymbol{X})$ and $\mathcal{H}^\dagger(\boldsymbol{Z})$ can be computed by using $\mathcal{O}(srn\log n)$ flops. Moreover, instead of computing a SVD directly, FIHT-VHL first projects a matrix onto a $2r$-dimensional subspace and then calculates the SVD of a rank-$2r$ matrix, which requires $\mathcal{O}(r^2sn+r^3)$ flops. Thus the main computational complexity in each step is $\mathcal{O}(r^2sn+r^3+srn\log n)$ and therefore our algorithm is very efficient compared to existing algorithms.

**Algorithm 1** FIHT-VHL

**Input:** Initialization $\boldsymbol{X}^0=\mathcal{H}^\dagger\mathcal{T}_r\mathcal{H}\mathcal{A}^*(\boldsymbol{y})$.
**Output:** $\boldsymbol{X}^T$

```
1: for t = 0,1,...,T−1 do
2:     X̃^t = X^t − A*(A(X^t) − y)
3:     W^t = P_{𝔗_t} H(X̃^t)
4:     Z^{t+1} = T_r(W^t)
5:     X^{t+1} = H†(Z^{t+1})
6: end for
```

1: **for** $t=0,1,\dots,T-1$ **do**
2: $\widetilde{\boldsymbol{X}}^t=\boldsymbol{X}^t-\mathcal{A}^*(\mathcal{A}(\boldsymbol{X}^t)-\boldsymbol{y})$
3: $\boldsymbol{W}^t=\mathcal{P}_{\mathfrak{T}_t}\mathcal{H}(\widetilde{\boldsymbol{X}}^t)$
4: $\boldsymbol{Z}^{t+1}=\mathcal{T}_r(\boldsymbol{W}^t)$
5: $\boldsymbol{X}^{t+1}=\mathcal{H}^\dagger(\boldsymbol{Z}^{t+1})$
6: **end for**

## 3. Main result

In this section, we establish our main result. To this end, we make two assumptions.

ASSUMPTION 3.1. *The column vectors $\{\boldsymbol{b}_j\}_{j=0}^{n-1}\subset\mathbb{C}^s$ of the subspace matrix $\boldsymbol{B}^{\mathsf{H}}$ are i.i.d random vectors which obey*

$$\mathbb{E}\left[\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\right]=\boldsymbol{I}_s \quad \text{and} \quad \max_{0\le\ell\le s-1}|\boldsymbol{b}_j[\ell]|^2\le\mu_0,$$

*for some constant $\mu_0$. Here, $\boldsymbol{b}_j[\ell]$ denotes the $\ell$-th entry of $\boldsymbol{b}_j$.*

REMARK 3.1. This assumption is standard in compressed sensing [6] and blind super-resolution [7,9,13,21], and holds with $\mu_0=1$ when $\boldsymbol{b}$ is uniformly sampled from the rows of a Discrete Fourier Transform (DFT) matrix.

ASSUMPTION 3.2. *There exists a constant $\mu_1>0$ such that*

$$\max_{0\le i\le n_1-1}\|\boldsymbol{U}_i\|_{\mathsf{F}}^2\le\frac{\mu_1 r}{n} \quad \text{and} \quad \max_{0\le j\le n_2-1}\left\|\boldsymbol{e}_j^{\mathsf{H}}\boldsymbol{V}\right\|_2^2\le\frac{\mu_1 r}{n},$$

*where the columns of $\boldsymbol{U}\in\mathbb{C}^{sn_1\times r}$ and $\boldsymbol{V}\in\mathbb{C}^{n_2\times r}$ are the left and right singular vectors of $\boldsymbol{Z}^\natural=\mathcal{H}(\boldsymbol{X}^\natural)$ separately, and $\boldsymbol{U}_i=\boldsymbol{U}[is:(i+1)s-1]$ is the $i$-th block of $\boldsymbol{U}$.*

REMARK 3.2. Assumption 3.2 is commonly used in spectrally sparse signal recovery [3,4,8] and blind super-resolution [7,16], and is satisfied when the minimum separation distance between $\{\tau_k\}_{k=1}^r$ is greater than about $1/n$.

Now, we are in the position to state our main result, whose proof is deferred to Section 5.

THEOREM 3.1. *Let $0<\varepsilon<\frac{1}{10}$ be a constant. Under Assumptions 3.1 and 3.2, with probability at least $1-(sn)^{-c_1}$, the iterations generated by FIHT-VHL with the initial guess $\boldsymbol{X}^0=\mathcal{H}^\dagger\mathcal{T}_r\mathcal{H}(\mathcal{A}^*(\boldsymbol{y}))$ satisfy*

$$\left\|\boldsymbol{X}^t-\boldsymbol{X}^\natural\right\|_{\mathsf{F}}\le\left(\frac{1}{2}\right)^t\left\|\boldsymbol{X}^0-\boldsymbol{X}^\natural\right\|_{\mathsf{F}} \tag{3.1}$$

*provided the sample complexity obeys that*

$$n \geq C\kappa^2\mu_0^2\mu_1 s^2 r^2 \log^2(sn)/\varepsilon^2$$

*where* $c_1$ *and* $C$ *are absolute constants and* $\kappa = \sigma_1(\mathcal{H}(\boldsymbol{X}^\natural))/\sigma_r(\mathcal{H}(\boldsymbol{X}^\natural))$.

REMARK 3.3. The sample complexity established in [7] for the Vectorized Hankel Lift is $n \geq c\mu_0\mu_1 \cdot sr\log^4(sn)$. While the sample complexity is sub-optimal dependence on $s$ and $r$, our recovery method requires low per iteration computational complexity.

REMARK 3.4. It is shown in [16] that PGD-VHL can exactly recover $\boldsymbol{X}^\natural$ when the measurements are of order $\mathcal{O}(s^2r^2\log^2(sn))$. Moreover the exact recovery of PGD-VHL relies on a more complicated regularization scheme.

## 4. Numerical simulations

Numerical experiments are conducted to evaluate the performance of FIHT-VHL. In the experiments, the target matrix $\boldsymbol{X}^\natural$ is generated by $\boldsymbol{X}^\natural = \sum_{k=1}^r d_k \boldsymbol{h}_k \boldsymbol{a}_{\tau_k}^{\mathsf{H}}$ and the measurements are obtained by (1.1), where the locations $\{\tau_k\}_{k=1}^r$ are generated from a standard uniform distribution $U(0,1)$, and the amplitudes $\{d_k\}_{k=1}^r$ are generated via $d_k = (1+10^{c_k})e^{-i\psi_k}$ where $\psi_k$ follows $U(0,2\pi)$ and $c_k$ follows $U(0,1)$. Each row of subspace matrix $\boldsymbol{B}$ is uniformly sampled from the rows of a Discrete Fourier Transform matrix. The coefficient vectors $\{\boldsymbol{h}_k\}_{k=1}^r$ are generated from a standardized multivariate Gaussian distribution $MVN_s(0, I_{s\times s})$, where $I_{s\times s}$ is the identity matrix.

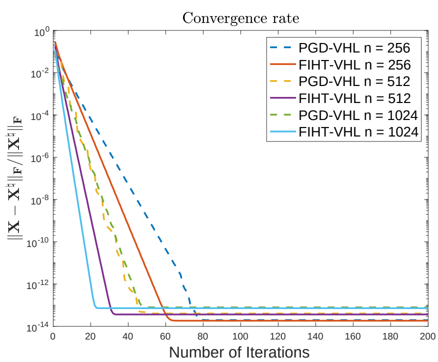


FIG. 4.1. *Convergences of FIHT-VHL and PGD-VHL for* $n = 256, 512, 1024$ *when* $s = 4$ *and* $r = 4$.

In the first experiment, we study the convergence rate of FIHT-VHL under different number of observations and compare it with PGD-VHL [16]. The step size for FIHT-VHL in the second step is given by

$$\alpha = \frac{\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}(\boldsymbol{y} - \mathcal{A}(\boldsymbol{X}^t))\|_2^2}{\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}(\boldsymbol{y} - \mathcal{A}(\boldsymbol{X}^t))\|_2^2} \tag{4.1}$$

in each iteration and for PGD-VHL, the step size is chosen using the line search method. We set dimensions of subspaces $s = 4$ and the number of point sources $r = 4$. Figure 4.1 presents the logarithmic recovery error $\log_{10}\left\|\boldsymbol{X}_t - \boldsymbol{X}^\natural\right\|_{\mathsf{F}} / \left\|\boldsymbol{X}^\natural\right\|_{\mathsf{F}}$ with respect to the number of iterations. Figure 4.1 shows that FIHT-VHL converges linearly which is in accordance with our main theorem. Compared to PGD-VHL [16], FIHT-VHL requires fewer number of iterations to achieve convergence when $s = r = 4$.

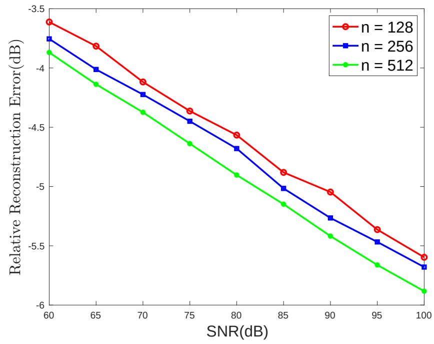


FIG. 4.2. *Performance of FIHT-VHL under different noise levels when $s=r=2$.*

In the second experiment, we conduct tests to illustrate the robustness of FIHT-VHL to additive noise. More specifically, we add noise vector $\boldsymbol{e}=\sigma_{\boldsymbol{e}}\cdot\|\boldsymbol{y}\|_2\cdot\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}$ to the measurements where $\boldsymbol{y}$ is the noiseless observations (1.3), $\sigma_{\boldsymbol{e}}$ denotes the noise level and $\boldsymbol{w}$ is the standard Gaussian vector with i.i.d entries. In the tests, the noise level $\sigma_{\boldsymbol{e}}$ is evenly spaced from $10^{-5}$ to $10^{-3}$, corresponding to the signal-to-noise ratio (SNR) from 100 to 60 dB. For each noise level, 10 random trails are conducted with $s=r=2$. We choose $n=128$ and $n=256$ for the number of measurements. In Figure 4.2, we demonstrate the linear relationship between the average relative reconstruction error and the noise level. It can be seen that the relative recovery error decreases with the increase of the number of measurements.

## 5. Proof of main result

We first introduce three auxiliary lemmas that will be used in our proof.

LEMMA 5.1 ([7, Corollary III.10]). *Suppose $0<\varepsilon<1$ and $n\geq C\varepsilon^{-2}\mu_0\mu_1 sr\log(sn)$. Then the event*

$$\|\mathcal{P}_{\mathfrak{T}}(\mathcal{G}\mathcal{G}^*-\mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*)\mathcal{P}_{\mathfrak{T}}\|\leq\varepsilon \tag{5.1}$$

*occurs with probability at least $1-(sn)^{-c_1}$ for absolute constants $c_1$ and $C$.*

LEMMA 5.2 ([16, Lemma 6.3]). *Suppose that $\boldsymbol{Z}^\natural$ is $\mu_1$-incoherent. Then with probability at least $1-(sn)^{-c_1}$, the initialization $\boldsymbol{Z}_0=\mathcal{T}_r\mathcal{H}(\boldsymbol{X}^0)$ obeys*

$$\left\|\boldsymbol{Z}^0-\boldsymbol{Z}^\natural\right\|\leq c_0\sigma_1(\boldsymbol{Z}^\natural)\sqrt{\frac{\mu_0\mu_1 sr\log^2(sn)}{n}},$$

*where $c_0$ and $c_1$ are absolute constants.*

LEMMA 5.3. *Suppose that*

$$\left\|\boldsymbol{Z}^t-\boldsymbol{Z}^\natural\right\|_{\mathsf{F}}\leq\frac{\sigma_r(\boldsymbol{Z}^\natural)\varepsilon}{16\sqrt{(1+\varepsilon)\cdot\mu_0 s}}. \tag{5.2}$$

*Conditioned on (5.1), one has*

$$\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\|\leq 3\sqrt{1+\varepsilon}, \tag{5.3}$$

$$\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\|\leq 2\varepsilon. \tag{5.4}$$

Now we are in the position to prove our main result. The iteration in FIHT-VHL can be rewritten as

$$\boldsymbol{Z}^{t+1} = \mathcal{T}_r \mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right). \tag{5.5}$$

For ease of exposition, we will prove our main results in this section in terms of $\boldsymbol{Z}^t$ and $\boldsymbol{Z}^\natural$ but note that results in terms of $\boldsymbol{X}^t$ and $\boldsymbol{X}^\natural$ follow immediately due to

$$\left\|\boldsymbol{X}^t - \boldsymbol{X}^\natural\right\|_{\mathsf{F}} = \left\|\mathcal{H}^\dagger(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural)\right\|_{\mathsf{F}} \le \left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}. \tag{5.6}$$

Motivated by [5], we prove our main result by induction. When $t=0$, the condition (5.2) can be verified as follows.

$$\begin{aligned}
\left\|\boldsymbol{Z}^0 - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} &\le \sqrt{2r}\left\|\boldsymbol{Z}^0 - \boldsymbol{Z}^\natural\right\| \\
&\le \sqrt{2r}\cdot c_0\sigma_1(\boldsymbol{Z}^\natural)\sqrt{\frac{\mu_0\mu_1 sr\log^2(sn)}{n}} \\
&\le \frac{\sigma_r(\boldsymbol{Z}^\natural)\varepsilon}{16\sqrt{(1+\varepsilon)\mu_0 s}}
\end{aligned}$$

where the second line is due to Lemma 5.2 and the last line follows from $n \ge C\varepsilon^{-2}\kappa^2\mu_0^2\mu_1 s^2 r^2\log^2(sn)$. Next we assume (5.2) holds for the iterations $0,1,\ldots,t$, and then prove it also holds for $t+1$. Denote

$$\boldsymbol{W}^t = \mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right).$$

Then it can be seen that $\boldsymbol{Z}^{t+1} = \mathcal{T}_r(\boldsymbol{W}^t)$. Since $\boldsymbol{Z}^{t+1}$ is the best rank-$r$ approximation of $\boldsymbol{W}^t$, it implies that $\left\|\boldsymbol{Z}^{t+1} - \boldsymbol{W}^t\right\|_{\mathsf{F}} \le \left\|\boldsymbol{Z}^\natural - \boldsymbol{W}^t\right\|_{\mathsf{F}}$. A direct computation yields that

$$\begin{aligned}
\left\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} &\le \left\|\boldsymbol{Z}^{t+1} - \boldsymbol{W}^t\right\|_{\mathsf{F}} + \left\|\boldsymbol{W}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\
&\le 2\left\|\boldsymbol{W}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\
&= 2\left\|\mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right) - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\
&\le 2\left\|\mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right)\right\|_{\mathsf{F}} + 2\left\|(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\le 2\left\|(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} + 2\left\|\mathcal{P}_{\mathfrak{T}_t}(\mathcal{G}\mathcal{G}^* - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*)\mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\quad + 2\left\|\mathcal{P}_{\mathfrak{T}_t}(\mathcal{G}\mathcal{G}^* - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*)(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\le 2\left\|(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} + 2\left\|\mathcal{P}_{\mathfrak{T}_t}(\mathcal{G}\mathcal{G}^* - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*)\mathcal{P}_{\mathfrak{T}_t}\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\quad + 2\left\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}\mathcal{G}^*(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\quad + 2\left\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*(\mathcal{I} - \mathcal{P}_{\mathfrak{T}_t})\left(\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right)\right\|_{\mathsf{F}} \\
&\triangleq I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

We bound $I_1, I_2, I_3$ and $I_4$, respectively. Applying [20, Lemma 4.1] yields that

$$I_1 \le 2\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}^2/\sigma_r(\boldsymbol{Z}^\natural), \quad I_3 \le 2\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}^2/\sigma_r(\boldsymbol{Z}^\natural)$$

For $I_2$, a simple computation yields that

$$\begin{aligned}
I_2 &\le 2\left\|\mathcal{P}_{\mathfrak{T}_t}(\mathcal{G}\mathcal{G}^* - \mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*)\mathcal{P}_{\mathfrak{T}_t}\right\| \cdot \left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\
&\le 4\varepsilon\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}
\end{aligned}$$

where the last line is due to (5.4). Finally, Lemma 5.3 implies that

$$I_4 \le 3\sqrt{1+\varepsilon}\cdot \left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}^2/\sigma_r(\boldsymbol{Z}^\natural).$$

Combining these terms together, we have

$$\begin{aligned}\left\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} &\le \left(4\varepsilon + \frac{4+3\sqrt{1+\varepsilon}}{\sigma_r(\boldsymbol{Z}^\natural)}\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}\right)\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\ &\le \left(4\varepsilon + \frac{4+3\sqrt{1+\varepsilon}}{16\sqrt{(1+\varepsilon)\mu_0 s}}\cdot\varepsilon\right)\cdot\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \qquad (5.7) \\ &\le 5\varepsilon\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}} \\ &\le \frac{1}{2}\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}, \qquad (5.8)\end{aligned}$$

where (5.7) follows from (5.2) and (5.8) is due to $\varepsilon < 1/10$. Since $\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}$ is a contractive sequence following from (5.8), the inequality (5.2) holds for all $t \ge 0$ by induction. Thus we complete the proof.

**5.1. Proof of Lemma 5.3.** For any $\boldsymbol{Z} \in \mathbb{C}^{sn_1\times n_2}$, we have

$$\begin{aligned}\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z})\|_{\mathsf{F}}^2 &= \langle \mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z}), \mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z})\rangle \\ &= \langle \mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z}), \mathcal{A}^*\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z})\rangle \\ &= \langle \boldsymbol{Z}, \mathcal{P}_{\mathfrak{T}}\mathcal{G}(\mathcal{A}^*\mathcal{A}-\mathcal{I})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z})\rangle + \langle \boldsymbol{Z}, \mathcal{P}_{\mathfrak{T}}\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}(\boldsymbol{Z})\rangle \\ &\le (1+\varepsilon)\|\boldsymbol{Z}\|_{\mathsf{F}}^2,\end{aligned}$$

where the last inequality is due to Lemma 5.1. So it follows that $\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}\| \le \sqrt{1+\varepsilon}$ and

$$\begin{aligned}\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| &\le \|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}\| + \|\mathcal{A}\mathcal{G}^*(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\| \\ &\le \sqrt{1+\varepsilon} + \|\mathcal{A}\|\cdot\|\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}}\| \\ &\le \sqrt{1+\varepsilon} + \frac{2\sqrt{\mu_0 s}\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}}{\sigma_{\min}(\boldsymbol{Z}^\natural)} \\ &\le 3\sqrt{1+\varepsilon},\end{aligned}$$

where the last inequality is due to (5.2). Finally, a straightforward computation yields that

$$\begin{aligned}\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| &\le \|(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| \\ &\le \|(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| \\ &\quad + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\| \\ &\quad + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}\| \\ &\le \|(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\mathcal{G}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}\mathcal{G}^*(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\| \\ &\quad + \|(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\| \\ &\quad + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}\mathcal{A}^*\mathcal{A}\mathcal{G}^*(\mathcal{P}_{\mathfrak{T}_t} - \mathcal{P}_{\mathfrak{T}})\| + \|\mathcal{P}_{\mathfrak{T}}\mathcal{G}(\mathcal{I}-\mathcal{A}^*\mathcal{A})\mathcal{G}^*\mathcal{P}_{\mathfrak{T}}\| \\ &\le \frac{4\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}}{\sigma_r(\boldsymbol{Z}^\natural)} + \frac{2\left\|\boldsymbol{Z}^t - \boldsymbol{Z}^\natural\right\|_{\mathsf{F}}}{\sigma_r(\boldsymbol{Z}^\natural)}\cdot\sqrt{\mu_0 s}\cdot\big(\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}\mathcal{A}^*\|\end{aligned}$$

$$+\|\mathcal{P}_{\mathfrak{T}}\mathcal{G}\mathcal{A}^*\|\big)+\varepsilon \tag{5.9}$$
$$\le \frac{4\varepsilon}{16\sqrt{(1+\varepsilon)\cdot\mu_0 s}}+\frac{8\varepsilon\sqrt{\mu_0 s(1+\varepsilon)}}{16\sqrt{(1+\varepsilon)\cdot\mu_0 s}}+\varepsilon \tag{5.10}$$
$$\le 2\varepsilon,$$

where (5.9) is due to [20, Lemma 4.1] and the fact that $\|\mathcal{A}^*\|=\|\mathcal{A}\|\le\sqrt{\mu_0 s}$ and $\|\mathcal{P}_{\mathfrak{T}_t}\mathcal{G}\mathcal{A}^*\|=\|\mathcal{A}\mathcal{G}^*\mathcal{P}_{\mathfrak{T}_t}\|$, (5.10) follows from (5.2).

## 6. Conclusion

We propose a FIHT-VHL method to solve the blind super-resolution problem in a non-convex scheme. The convergence analysis has been established for FIHT-VHL, showing that the algorithm linearly converges to the target matrix given suitable initialization and provided the number of samples is large enough. The numerical experiments illustrate our theoretical results.

**Acknowledgement.** The authors would like to thank professor Ke Wei and Yingzhou Li for fruitful discussions. The authors would also like to thank the anonymous reviewers and the Associate Editor for their useful comments which have helped to improve the quality of the present work. The research of this work was supported by National Key R&D Program of China under Grant No. 2021YFA1003300. Jinchi Chen was partially supported by National Natural Science Foundation of China under Grant No. 12001108. Weiguo Gao was partially supported by National Natural Science Foundation of China under Grant No. 71991471, U1811461.

## REFERENCES


[1] A. Ahmed, B. Recht, and J. Romberg, *Blind deconvolution using convex programming*, IEEE Trans. Inf. Theory, 60(3):1711–1732, 2013. 1

[2] B. Bernstein and C. Fernandez-Granda, *Deconvolution of point sources: a sampling theorem and robustness guarantees*, Commun. Pure Appl. Math., 72(6):1152–1230, 2019. 1

[3] J.-F. Cai, S. Liu, and W. Xu, *A fast algorithm for reconstruction of spectrally sparse signals in super-resolution*, Proc. SPIE 9597, Wavelets and Sparsity XVI, 9597:95970A, 2015. 3.2

[4] J.-F. Cai, T. Wang, and K. Wei, *Spectral compressed sensing via projected gradient descent*, SIAM J. Optim., 28(3):2625–2653, 2018. 3.2

[5] J.-F. Cai, T. Wang, and K. Wei, *Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion*, Appl. Comput. Harmon. Anal., 46(1):94–121, 2019. 1, 5

[6] E.J. Candes and Y. Plan, *A probabilistic and ripless theory of compressed sensing*, IEEE Trans. Inf. Theory, 57(11):7235–7254, 2011. 3.1

[7] J. Chen, W. Gao, S. Mao, and K. Wei, *Vectorized Hankel lift: A convex approach for blind super-resolution of point sources*, IEEE Trans. Inf. Theory, 68(12):8280–8309, 2022. 1, 1, 1, 1, 2, 3.1, 3.2, 3.3, 5.1

[8] Y. Chen and Y. Chi, *Robust spectral compressed sensing via structured matrix completion*, IEEE Trans. Inf. Theory, 60(10):6576–6601, 2014. 3.2

[9] Y. Chi, *Guaranteed blind sparse spikes deconvolution via lifting and convex optimization*, IEEE J. Sel. Top. Signal Process., 10(4):782–794, 2016. 1, 1, 3.1

[10] A. Eftekhari, J. Tanner, A. Thompson, B. Toader, and H. Tyagi, *Sparse non-negative super-resolution - simplified and stabilised*, Appl. Comput. Harmon. Anal., 50:216–280, 2021. 1

[11] J.E. Evans, J.R. Johnson, and D.F. Sun, *Application of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance systems*, Technical report, Massachusetts Institute of Technology, Lincoln Laboratory, 1982. 1

[12] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, *Removing camera shake from a single photograph*, ACM Trans. Graph., 25(3):787–794, 2006. 1

[13] S. Li, M.B. Wakin, and G. Tang, *Atomic norm denoising for complex exponentials with unknown waveform modulations*, IEEE Trans. Inf. Theory, 66(6):3893–3913, 2019. 1, 3.1

[14] X. Li, S. Ling, T. Strohmer, and K. Wei, *Rapid, robust, and reliable blind deconvolution via nonconvex optimization*, Appl. Comput. Harmon. Anal., 47(3):893–934, 2019. 1, 1
[15] X. Luo and G.B. Giannakis, *Low-complexity blind synchronization and demodulation for (ultra-) wideband multi-user ad hoc access*, IEEE Trans. Wirel. Commun., 5(7):1930–1941, 2006. 1
[16] S. Mao and J. Chen, *Blind super-resolution of point sources via projected gradient descent*, IEEE Trans. Signal Process., 70:4649–4664, 2022. 1, 3.2, 3.4, 4, 4, 5.2
[17] G.F. Margrave, M.P. Lamoureux, and D.C. Henley, *Gabor deconvolution: Estimating reflectivity by nonstationary deconvolution of seismic data*, Geophysics, 76(3):W15–W30, 2011. 1
[18] M.A. Suliman and W. Dai, *Mathematical theory of atomic norm denoising in blind two-dimensional super-resolution*, IEEE Trans. Signal Process., 69:1681–1696, 2021. 1
[19] B. Vandereycken, *Low-rank matrix completion by Riemannian optimization*, SIAM J. Optim., 23(2):1214–1236, 2013. 1
[20] K. Wei, J.-F. Cai, T.F. Chan, and S. Leung, *Guarantees of Riemannian optimization for low rank matrix recovery*, SIAM J. Matrix. Anal. A., 37(3):1198–1222, 2016. 5, 5.1
[21] D. Yang, G. Tang, and M.B. Wakin, *Super-resolution of complex exponentials from modulations with unknown waveforms*, IEEE Trans. Inf. Theory, 62(10):5809–5830, 2016. 1, 1, 1, 3.1
[22] S. Zhang, Y. Hao, M. Wang, and J.H. Chow, *Multichannel Hankel matrix completion through nonconvex optimization*, IEEE J. Sel. Top. Signal Process., 12(4):617–632, 2018. 1, 1