Abstract—We consider the problem of resolving $r$ point sources from $n$ samples at the low end of the spectrum when point spread functions (PSFs) are not known. Assuming that the spectrum samples of the PSFs lie in low dimensional subspace (let $s$ denote the dimension), we can formulate it as a matrix recovery problem, followed by location estimation. By exploiting the low rank structure of the vectorized Hankel matrix associated with the target matrix, a convex approach called Vectorized Hankel Lift is proposed for the matrix recovery. It is shown that $n \gtrsim rs \log^4 n$ samples are sufficient for Vectorized Hankel Lift to achieve the exact recovery. For the location retrieval from the matrix, applying the single snapshot MUSIC method within the vectorized Hankel lift framework corresponds to the spatial smoothing technique proposed to improve the performance of the MMV MUSIC for the direction-of-arrival (DOA) estimation.

Index Terms—Blind super-resolution, vectorized Hankel lift, low rank, MUSIC.

I. INTRODUCTION

A. Problem Formulation

In this paper, we study the super-resolution of point sources when point spread functions (PSFs) are not known. More specifically, consider a point source signal $x(t)$ of the form

$$x(t) = \sum_{k=1}^{r} d_k \delta(t - \tau_k),$$

(1.1)

where $\delta(\cdot)$ is the Dirac function, $\{\tau_k\}$ and $\{d_k\}$ are the locations and amplitudes of the point source signal, respectively. Let $y(t)$ be its convolution with unknown point spread functions,

$$y(t) = \sum_{k=1}^{r} d_k \delta(t - \tau_k) * g_k(t) = \sum_{k=1}^{r} d_k \cdot g_k(t - \tau_k),$$

(1.2)

where $\{g_k\}_{k=1}^{r}$ are the point spread functions depending on the locations of the point sources.

Taking the Fourier transform on both sides of (1.2) yields

$$\hat{y}(f) = \int_{-\infty}^{+\infty} y(t) e^{-2\pi if t} dt = \sum_{k=1}^{r} d_k e^{-2\pi if \tau_k} \hat{g}_k(f).$$

(1.3)

The goal in blind super-resolution is to recover $\{d_k, \tau_k\}_{k=1}^{r}$ from the low end of the spectrum

$$y[j] = \sum_{k=1}^{r} d_k e^{-2\pi i j \tau_k} \hat{g}_k[j]$$

for $j = 0, \cdots, n - 1$ (1.4)

when $g_k = [\hat{g}_k(0), \cdots, \hat{g}_k(n-1)]^T$, $k = 1, \cdots, r$, are not known. Here we assume the index $j \in \{0, 1, \cdots, n-1\}$ rather than $j \in \{-n/2, \cdots, n/2\}$ only for convenience of notation. In addition to blind super-resolution, the observation model (1.4) also arises from many other important applications, such as 3D single-molecule microscopy [48], multi-user communication system [43] and nuclear magnetic resonance spectroscopy [47].

It is evident that the blind super-resolution problem is ill-posed without any further assumptions. To address this issue, we assume that the set of vectors $\{g_k\}_{k=1}^{r}$ corresponding to the unknown point spread functions belong to a common and known low-dimensional subspace represented by $B \in \mathbb{C}^{r \times s}$, i.e.,

$$g_k = Bh_k,$$

(1.5)

where $h_k \in \mathbb{C}^{s}$ is the unknown orientation of $g_k$ in this subspace. As is pointed out in [62], the subspace assumption is reasonable in several application scenarios. Moreover, it has been extensively used in the literature, see for example [1], [18], [33], [37], [62].

For any $\tau \in [0, 1]$, define the vector $a_\tau \in \mathbb{C}^{n}$ as

$$a_\tau = \begin{bmatrix} 1 & e^{-2\pi i \tau} & \cdots & e^{-2\pi i (n-1) \tau} \end{bmatrix}^T.$$

(1.6)

Let $b_j \in \mathbb{C}^{s}$ be the $j$th column vector of $B^*$. If we define the matrix $X^2 \in \mathbb{C}^{s \times n}$ as

$$X^2 = \sum_{k=1}^{r} d_k h_k a_{\tau_k}^T,$$

(1.7)

then under the subspace assumption (1.5) and using the lifting trick [1], [12], [18], [19], [37], [38], [42], [62], [66], the observation model (1.4) can be reformulated as a linear measurement of $X^2$:

$$y[j] = \langle b_j e_j^T, X^2 \rangle$$

for $j = 0, \cdots, n - 1$, (1.8)
where the inner product of two matrices is given by \( \langle A, B \rangle = \text{trace}(A^T B) \), \( e_j \) is \((j + 1)\)th column of the \( n \times n \) identity matrix \( I_n \), and throughout this paper vectors and matrices are indexed starting with zero. Moreover, we can further rewrite (I.8) in the following compact form,

\[
y = A(X^2),
\]

where \( A : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^n \) is a linear operator defined by \([A(X)]_j = \langle b_j, e_j^T X \rangle\). The adjoint of the operator \( A(\cdot) \), denoted \( A^*(\cdot) \), is defined as \( A^*(y) = \sum_{j=0}^{n-1} y[j] b_j e_j^T \).

Based on the above reformulation of blind super-resolution under the subspace assumption, it can be seen that the key is to recover \( X^2 \) from the linear measurement vector \( y \). Once \( X^2 \) is reconstructed, the frequency components can be extracted from \( X^2 \) by the subspace methods which will be detailed in Section II-B. After the frequency components are obtained, \( \{d_k, h_k\} \) can be recovered by solving a least squares system. Moreover, due to the multiplicative form of \( d_k \) and \( h_k \) in (I.7), we only expect to recover them separately up to a scaling ambiguity. Thus, we will assume that \( \|h_k\|_2 = 1 \) without loss of generality.

Note that the formulations in (I.4) and (I.9) are by no means new and they have been utilized in [62]. Moreover, when the point spread function \( g \) is shared among all point sources (i.e., the stationary case), (I.4) reduces to the blind sparse spikes deconvolution model considered in [18]. To recover the target matrix \( X^2 \) from the linear measurements \( y \), following the approach developed in [54] for spectrally sparse signal recovery, a similar atomic norm minimization method is proposed in [62].

\[
\min_X \|X\|_B \text{ subject to } y = A(X),
\]

where the atomic norm \( \|X\|_B \) is defined as

\[
\|X\|_B := \inf \{ t > 0 : X \in t \cdot \text{conv}(B) \} = \inf_{d_k, r_k, \|h_k\|_2 = 1} \left\{ \sum_{k=1}^{r} d_k : X = \sum_{k=1}^{r} d_k h_k a_{r_k}^T, d_k > 0 \right\},
\]

The successful recovery guarantee of (I.10) is studied in [62], while the robust analysis is provided separately in [33]. Note that for spectrally sparse signal recovery, in addition to atomic norm minimization, there are also methods which exploit the low rank property of the structured matrix formed from the signal [6], [7], [15]. This motivates us to develop a low rank approach for blind super-resolution.

**B. Exploiting the Low Rank Structure: Vectorized Hankel Lift**

We start with a brief view of spectrally sparse signal recovery based on the hidden low rank structure. Let \( x(t) \) be a spectrally sparse signal consisting of \( r \) complex sinusoids,

\[
x(t) = \sum_{k=1}^{r} d_k e^{-2\pi i t r_k}.
\]

Let \( x = [x(0), \cdots, x(n - 1)]^T \) be a vector of length \( n \) which is obtained by sampling \( x(t) \) at \( n \) contiguous, equally-spaced points. In a nutshell, spectrally sparse signal recovery is about reconstructing the signal \( x \) from its partial samples. Recalling the definition of \( a_r \) in (I.6), we can represent \( x \) as

\[
x = \sum_{k=1}^{r} d_k a_{r_k}^T.
\]

Let \( \mathcal{H} \) be a linear operator which maps a vector \( x \) into an \( n_1 \times n_2 \) Hankel matrix,

\[
\mathcal{H}(x) = \begin{bmatrix} x_0 & x_1 & \cdots & x_{n_2-1} \\ x_1 & x_2 & \cdots & x_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1-1} & x_{n_1} & \cdots & x_{n_1+n_2-1} \end{bmatrix} \in \mathbb{C}^{n_1 \times n_2},
\]

where \( x_i \) is the \( i \)th entry of \( x \) and \( n_1 + n_2 = n + 1 \). Without loss of generality, we assume \( n_1 = n_2 = (n + 1)/2 \) in this paper. Due to the particular expression of \( x \) in (I.11), it is not hard to see that the rank of \( \mathcal{H}(x) \) is at most \( r \) according to the Vandermonde decomposition of \( \mathcal{H}(x) \) [15].

Note that the expression for the data matrix \( X^2 \) in (I.7) is overall similar to that for the spectrally sparse vector \( x \) in (I.11), except that the weights \( d_k h_k \) in front of \( a_{r_k}^T \) in (I.7) are vectors and consequently \( X^2 \) is a matrix rather than a vector. Intuitively, if we treat each column of \( X^2 \) as a single element and form a matrix in the same fashion as in (I.12), it can be expected that the resulting matrix is also low rank. This is indeed true. Specifically, let \( \mathcal{H} \) be the vectorized Hankel lifting operator which maps a matrix \( X \in \mathbb{C}^{n \times n} \) with columns \( \{x_j\} \) into an \( s n_1 \times n_2 \) matrix,

\[
\mathcal{H}(X) = \begin{bmatrix} x_0 & x_1 & \cdots & x_{n_2-1} \\ x_1 & x_2 & \cdots & x_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1-1} & x_{n_1} & \cdots & x_{n_1+n_2-1} \end{bmatrix} \in \mathbb{C}^{sn_1 \times n_2},
\]

where \( n_1 + n_2 = n + 1 \). To distinguish the matrix \( \mathcal{H}(X) \) in (I.13) from the one in (I.12), we refer to \( \mathcal{H}(X) \) as the vectorized Hankel matrix associated with \( X \). Then a simple algebra yields that the vectorized Hankel matrix \( \mathcal{H}(X^2) \) associated with \( X^2 \) appearing in the blind super-resolution problem admits the following decomposition:

\[
\mathcal{H}(X^2) = E_{h, L} \text{diag}(d_1, \cdots, d_r) E_R^T,
\]

where the matrices \( E_{h, L} \in \mathbb{C}^{sn_1 \times n_2} \) and \( E_R \in \mathbb{C}^{n_2 \times r} \) are given by

\[
E_{h, L} = \begin{bmatrix} h_1 & \cdots & h_r \\ \vdots & \ddots & \vdots \\ h_1 e^{-2\pi i r_1 (n_1-1)} & \cdots & h_r e^{-2\pi i r_r (n_1-1)} \end{bmatrix}
\]

and

\[
E_R = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ e^{-2\pi i r_1 (n_2-1)} & \cdots & e^{-2\pi i r_r (n_2-1)} \end{bmatrix}.
\]

It follows immediately that the rank of \( \mathcal{H}(X^2) \) is at most \( r \) and thus it is a low rank matrix when \( r \) is smaller than \( \min(s n_1, n_2) \).
In this paper we adopt the popular nuclear norm minimization to exploit the low rank structure of $\mathcal{H}(X^3)$, yielding a convex approach for the reconstruction of $X^3$ which is also referred to Vectorized Hankel Lift. Exact recovery guarantee will be established based on certain assumptions on the subspace matrix $B$ in (1.5).

### C. Other Related Work

In this section, we give a brief introduction of other related work in addition to [18], [33], [62]. When the point spread functions are known and do not depend on the locations of the point sources, the measurement model (1.4) reduces to

$$y[j] = \sum_{k=1}^{r} d_k e^{-2\pi i \gamma_k j} \text{ for } j = 0, \ldots, n - 1. \quad (1.17)$$

In this case, estimating the locations $\gamma_k$ and amplitudes $d_k$ from $y$ is typically known as super-resolution or line spectrum estimation. This problem arises in many areas of science and engineering, such as array imaging [31], [53], Direction-of-Arrival (DOA) estimation [52], and inverse scattering [26]. The solution to this problem can date back to Prony [46]. In the Prony’s method, the locations are retrieved from the roots of a polynomial whose coefficients form an annihilating filter for the observation vector. Nevertheless, the Prony’s method is numerical unstable despite that in the noiseless setting successful retrieval is guaranteed in exact arithmetic.

As alternatives, several subspace methods have been developed, including MUSIC [51], ESPRIT [49], and the matrix pencil method [29]. In the absence of noise, the subspace methods are also able to identify the locations of the point sources. When there is noise, the stability of these methods has been studied in [35], [40], [41], [45] in the regime when $\Delta > C/n$, where $\Delta$ is the minimum (wraparound) separation between any two locations, and $C > 1$ is a proper numerical constant. The analysis essentially relies on the lower bound on the smallest singular value of the Vandermonde matrix. The super-resolution limits of MUSIC and ESPRIT have been discussed in [34], [35], which is about the noise level that can be tolerated in order for the algorithms to achieve super-resolution when $\Delta < 1/n$. In this regime, it is difficult to obtain a general and nontrivial lower bound on the smallest singular value of the Vandermonde matrix. Thus, the super-resolution limits in [34], [35] are established for point sources whose locations obey certain configurations.

Inspired by compressed sensing and low rank matrix reconstruction, various optimization based methods have also been developed for super-resolution and related problems. In [9], the total variation (TV) minimization method is used to resolve the locations of the point sources. It is shown that when $\Delta > C/n$, exact recovery of the locations can be guaranteed. Moreover, the solution to the TV minimization problem can be computed by solving a semidefinite programming (SDP). Note, in the discrete setting, super-resolution can be interpreted within the framework of compressed sensing. However, since the measurement model in super-resolution considers the low end spectrum, and hence is deterministic, the typical successful recovery guarantee for compressed sensing [11] cannot sufficiently explain the success of the TV norm minimization method for super-resolution. The robustness of TV norm minimization is studied in [8], and the super-resolution problem of non-negative point sources is considered in [20]–[23], [50]. Moreover, super-resolution from time domain samples has been investigated in [2], [4], [23].

When only partial entries of $y$ are observed in (1.17), filling in the missing entries is indeed the spectrally sparse signal recovery problem. Motivated by the work in [13], an atomic norm minimization method (ANM) is proposed for this problem. It is shown that $y$ can be reconstructed from $O(r \log r \log n)$ random samples provided the frequencies are well separated. ANM has been extended in [39], [65] to handle the case when multiple measurement vector (MMV) are available. In the setting of MMV, multiple snapshots of observations are collected and they share the same frequencies information. As already mentioned previously, the Hankel matrix corresponding to $y$ is a low rank matrix. Consequently, spectrally sparse signal recovery can be reformulated as a low rank Hankel matrix completion problem, and replacing the rank objective with the nuclear norm yields a recovery method known as EMaC. It has been shown that EMaC is able to reconstruct a spectrally sparse signal with high probability provided the number of observed entries is $O(r \log^4 n)$. In [63], a formulation of EMaC for the multi-snapsots scenario is presented. Additionally, based on the low rank property of the Hankel matrix, provable non-convex algorithms have been developed in [6], [7] to reconstruct spectrally sparse signals. Later, Zhang et al. [67] extend one of the non-convex algorithms to complete an MMV matrix, and in this work the same vectorized Hankel lift technique is used to exploit the hidden low rank structure. Recently, a matrix completion problem based on the low dimensional structure in the transform domain is studied in [14]. More precisely, it is assumed that after applying the Fourier transform to each column of the target matrix, each row of the resulting matrix will be a spectrally sparse signal. Since it does not require the spectrally signals share the same frequency information, a block-diagonal low rank structure is adopted to exploit the low dimensional structure. Exact recovery guarantee is also established provided the sampling complexity is nearly optimal.

Apart from super-resolution and spectrally sparse signal recovery, our work is also related to blind deconvolution. After the reparametrization of the signal and blurring kernel under the subspace assumption [1], the goal in blind deconvolution is to recover the vectors $x^2$ and $h^2$ simultaneously from the measurement vector in the form of

$$y = \text{diag}(Bh^2)Ax^2.$$  

Noting that the above measurement model can be reformulated as a linear operation on a rank-1 matrix, a nuclear norm minimization method is proposed for blind deconvolution. The performance guarantee of the method has been established in the case when $B$ is a partial Fourier matrix and $A$ is a Gaussian matrix. A non-convex gradient descent approach for blind deconvolution is developed and analyzed in [37], and the identifiability problem is studied in [19], [38].
D. Notation and Organization

Throughout this work, vectors, matrices and operators are denoted by bold lowercase letters, bold uppercase letters and calligraphic letters, respectively. Note that vectors and matrices are indexed starting with zero. The letter \( I \) denotes the identity operator. We use \( G_i \) to denote the matrix defined by

\[
G_i = \frac{1}{\sqrt{w_i}} \sum_{j+k=i} e_j e_k^T, \tag{I.18}
\]

where \( w_i \) is a constant defined as

\[
w_i = \#\{(j,k) | j + k = i, 0 \leq j \leq n_1 - 1, 0 \leq k \leq n_2 - 1\}. \tag{I.19}
\]

In fact, \( \{G_i\}_{i=0}^{n-1} \) forms an orthonormal basis of the space of \( n_1 \times n_2 \) Hankel matrices.

We use \( x[i] \) to denote the \( i \)th entry of \( x \) and \( X_{j,k} \) or \( X[j,k] \) to denote the \( (j,k) \)th entry of \( X \). Additionally, the \( i \)th row and \( j \)th column of \( X \) are denoted by \( X_i \) and \( X_j \), respectively. Furthermore, we use the MATLAB notation \( X(i:j,k) \) to denote a vector of size \( j - i + 1 \), with entries \( X_{i,k}, \ldots, X_{j,k} \), i.e.,

\[
X(i:j,k) = [X_{i,k}, \ldots, X_{j,k}]^T.
\]

For any matrix \( X \), \( \text{trace}(X) \), \( X^* \), \( X^T \) and \( \text{vec}(X) \) are used to denote the trace, conjugate transpose, transpose and column vectorization of \( X \), respectively. Also, \( \|X\| \), \( \|X\|_F \) and \( \|X\|_\diamond \) denote its spectral norm, Frobenius norm and nuclear norm, respectively.

We use \( \text{diag}(a) \) to denote the diagonal matrix specified by the vector \( a \). For a natural number \( n \), we use \([n]\) to denote the set \( \{0, \ldots, n-1\} \). For any two matrices \( A, B \) of the same size, their inner product is defined as \( \langle A, B \rangle = \text{trace}(A^*B) \). Moreover, we will refer to \( A \circ B, A \otimes B, A \odot B \) as the Hadamard, Kronecker product and Khatri-Rao product respectively. More precisely, the Hadamard product is the element-wise product of two matrices and the Kronecker product between \( A \) and \( B \) is given by

\[
A \otimes B = \begin{bmatrix}
A_{11}B & A_{12}B & \cdots & A_{1r}B \\
A_{21}B & A_{22}B & \cdots & A_{2r}B \\
\vdots & \vdots & \ddots & \vdots \\
A_{s1}B & A_{s2}B & \cdots & A_{sr}B
\end{bmatrix} \in \mathbb{C}^{sn_1 \times rn_2},
\]

and the Khatri-Rao product is given by

\[
A \odot B = [a_1 \otimes b_1 \ldots a_r \otimes b_r] \in \mathbb{C}^{sn_1 \times r},
\]

where \( a_i, b_i \) denote the \( i \)th column of \( A \) and \( B \), respectively. By the application of the Khatri-Rao product, we can rewrite \( E_{h,L} \) in (I.15) as \( E_{h,L} = E_L \odot H \), where \( E_L \) and \( H \) are matrices given by

\[
E_L = \begin{bmatrix}
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
e^{-2\pi \imath (n_1-1)} & \cdots & e^{-2\pi \imath (n_1-1)}
\end{bmatrix} \in \mathbb{C}^{n_1 \times r}, \tag{I.20}
\]

and \( H = [h_1 \ldots h_r] \in \mathbb{C}^{s \times r} \).

Throughout this paper, \( c, c_1, c_2, \ldots \) denote absolute positive numerical constants whose values may vary from line to line. The notation \( n \gtrless f(m) \) means that there exists an absolute constant \( c > 0 \) such that \( n \geq c \cdot f(m) \). Similarly, the notation \( n \lesssim f(m) \) means that there exists an absolute constant \( c > 0 \) such that \( n \leq c \cdot f(m) \).

The rest of this paper is organized as follows. Section II begins with the presentation of Vectorized Hankel Lift and its recovery guarantee, followed by the retrieval of the point source locations. Numerical results to demonstrate the performance of Vectorized Hankel Lift is presented at the end of Section II. The proofs of the main result are provided from Section III to Section VI. Finally, we conclude this paper with a few future directions in Section VII.

II. VECTORIZED HANKEL LIFT AND FREQUENCY RETRIEVAL

A. Vectorized Hankel Lift and Recovery Guarantee

Under the assumption that \( \mathcal{H}(X^\natural) \) is a low rank matrix, it is natural to reconstruct \( X^\natural \) by solving the affine rank minimization problem

\[
\min \text{rank}(\mathcal{H}(X)) \text{ s.t. } y = A(X). \tag{II.1}
\]

However, the problem (II.1) is computational intractable due to the rank objective. Since the nuclear norm of a matrix is the tightest convex envelope of the matrix rank, seeking a solution with a small nuclear norm is also able to enforce the low rank structure. Therefore, instead of solving (II.1) directly, we consider the following nuclear norm minimization problem for the recovery of \( X^\natural \):

\[
\min_{X \in \mathbb{C}^{s \times n}} \|\mathcal{H}(X)\|_\diamond \text{ s.t. } A(X) = y. \tag{II.2}
\]

In this paper, we refer to (II.2) as Vectorized Hankel Lift. There are many existing software packages that can be used to solve this problem. Thus we restrict our attention on the theoretical recovery guarantee of Vectorized Hankel Lift and investigate when the solution of (II.2) coincides with \( X^\natural \).

We need to reformulate (II.2) in order to facilitate the analysis. Let \( Z \) be an \( sn_1 \times n_2 \) matrix which can be expressed as

\[
Z = \begin{bmatrix}
z_{0,0} & \cdots & z_{0,n_2-1} \\
\vdots & \ddots & \vdots \\
z_{n_1-1,0} & \cdots & z_{n_1-1,n_2-1}
\end{bmatrix} \in \mathbb{C}^{sn_1 \times n_2},
\]

where \( z_{j,k} = Z(js : (j+1)s-1, k) \) for \( j = 0, \ldots, n_1-1 \) and \( k = 0, \ldots, n_2-1 \). Recall that \( \mathcal{H} \) is the vectorized Hankel lift operator defined in (I.13). The adjoint of \( \mathcal{H} \), denoted \( \mathcal{H}^* \), is a linear mapping from \( sn_1 \times n_2 \) matrices to matrices of size \( s \times n \). In particular, for any matrix \( Z \in \mathbb{C}^{s \times n} \), the \( i \)th column of \( \mathcal{H}^*(Z) \) is given by

\[
\mathcal{H}^*(Z)e_i = \sum_{j+k=i} z_{j,k}, \text{ for } i = 0, \ldots, n_1-1.
\]

Letting \( D^2 = \mathcal{H}^* \mathcal{H} \), we have

\[
D^2(X) = [w_0 x_0 \cdots w_{n-1} x_{n-1}],
\]
for any \( X \in \mathbb{C}^{s \times n} \), where the scalar \( w_i \) is defined as
\[
 w_i = \#\{(j, k)| j + k = i, 0 \leq j \leq n_1 - 1, 0 \leq k \leq n_2 - 1\}
\]
for \( i = 0, \cdots, n - 1 \). Moreover, we define \( \mathcal{G} = \mathcal{H}^{D^{-1}} \). Then
\[
 \mathcal{G}(X) = \sum_{i=0}^{n-1} \mathcal{G}(x_i e_i^T) = \sum_{i=0}^{n-1} G_i \otimes x_i, \quad (II.3)
\]
where the set of matrices \( \{G_i\}_{i=0}^{n-1} \) defined in (I.18) forms an orthonormal basis of the space of \( n_1 \times n_2 \) Hankel matrices. The adjoint of \( \mathcal{G} \), denoted \( \mathcal{G}^* \), is given by \( \mathcal{G}^* = \mathcal{D}^{-1} \mathcal{H}^* \). Additionally, \( \mathcal{G} \) and \( \mathcal{G}^* \) satisfy
\[
 \mathcal{G}^* \mathcal{G} = \mathcal{I} \quad ||\mathcal{G}|| = 1, \quad \text{and} \quad ||\mathcal{G}^*|| \leq 1.
\]
Letting \( Z = \mathcal{H}(X) = \mathcal{G}D(X) \), it can be readily verified that
\[
 D(X) = \mathcal{G}^*(Z) \quad \text{and} \quad (\mathcal{I} - \mathcal{G} \mathcal{G}^*)(Z) = 0.
\]
Furthermore, define \( D = \text{diag}(\sqrt{w_0}, \cdots, \sqrt{w_{n-1}}) \). We have \( \mathcal{A}D(X) = D\mathcal{A}(X) \) for any matrix \( X \). Therefore, the optimization problem (II.2) can be reformulated as
\[
 \min_{Z \in \mathbb{C}^{s,n}} ||Z||_*
\]
\[
 \text{s.t.} \quad Dy = A\mathcal{G}^*(Z) \quad \text{and} \quad (\mathcal{I} - \mathcal{G} \mathcal{G}^*)(Z) = 0. \quad (II.4)
\]
Due to the equivalence between (II.2) and (II.4), it suffices to investigate the recovery guarantee of (II.4). To this end, we make two assumptions.

**Assumption II.1:** The column vectors \( \{b_j\}_{j=0}^{n-1} \) of the subspace matrix \( B^* \) are independent and identically sampled from a distribution \( F \) which obeys the following properties:

- **Isotropy property.** A distribution \( F \) obeys the isotropy property if for \( b \sim F \),
  \[
  \mathbb{E}[bb^T] = I_s, \quad (II.5)
  \]
- **Incoherence property.** A distribution \( F \) satisfies the incoherence property with parameter \( \mu_0 \) if for \( b \sim F \),
  \[
  \max_{0 \leq j \leq n_1-1} |b[j]|^2 \leq \mu_0 \quad (II.6)
  \]
  holds, where \( b[j] \) denotes the \( j \)th entry of \( b \).
- **Lower bounded property.** A distribution \( F \) obeys the lower bounded property if for \( b \sim F \),
  \[
  ||b||_2^2 \geq 1. \quad (II.7)
  \]

The first two conditions (II.5) and (II.6) in Assumption II.1 are first introduced in [10] in the context of compressed sensing and these two properties are also made in [18], [33], [62] for the blind super-resolution problem. If \( F \) has mean zero, the isotropy condition states that the entries of \( b \) have unit variance and are uncorrelated, which implies \( \mu_0 \geq 1 \) in the incoherence property. The lower bound \( \mu_0 = 1 \) is achievable by several examples, for instance, when the components of \( b \) are Rademacher random variables taking the values \( \pm 1 \) with equal probability or \( b \) is uniformly sampled from the rows of a Discrete Fourier Transform (DFT) matrix. In addition to (II.5) and (II.6), we also need (II.7) to establish our main result. However, we would like to point out that (II.7) is not a stringent condition, but holds by many common random ensembles.

- If the components of \( b \) are Rademacher random variables or \( b \) is uniformly sampled from the rows of a DFT matrix, it is trivial that for any fixed \( j \in [n] \), \( ||b||_2^2 \geq s \geq 1 \).
- Note that what we really need is \( \min_{0 \leq j \leq n-1} ||b||_2^2 \geq 1 \).

Suppose the components of \( b \) are independently and identically sampled from a distribution with mean zero and unit variance, such as the uniform distribution on the interval \([-\sqrt{3}, \sqrt{3}]\). In such case, we can apply the bounded difference inequality to show that (II.7) holds with high probability, see Lemma III.1.

**Assumption II.2:** There exists a constant \( \mu_1 > 0 \) such that
\[
 \sigma_{\min}(E_L^* E_L) \geq \frac{n_1}{\mu_1} \quad \text{and} \quad \sigma_{\min}(E_R^* E_R) \geq \frac{n_2}{\mu_1}, \quad (II.8)
\]
where \( E_L \) and \( E_R \) are given in (I.20) and (I.16) and \( \sigma_{\min}(\cdot) \) denotes the smallest singular value of a matrix.

Assumption II.2 is the same as the one made in [6], [7], [15] for spectrally sparse signal recovery. Later, we will show that \( \sigma_{\min}(E_{hL}^* E_{hL}) \geq \frac{\mu_1}{\mu_0} \) also holds when \( \sigma_{\min}(E_j^* E_j) \geq \frac{\mu_1}{\mu_0} \), see Lemma III.3. Recalling the definition of \( E_L \) and \( E_R \), this assumption is essentially about the conditioning property of the Vandermonde matrix. This property is studied in [41] through the discrete Ingham inequality [30] and in [45] through the discrete large sieve inequality [57]. In particular, it follows from [45] that Assumption II.2 holds when the minimum wrap-around distance between the frequencies, denoted \( \Delta \), satisfies
\[
 \Delta \geq \frac{2\mu_1/(\mu_1 - 1)}{n}. \quad (II.9)
\]

We are in position to present the main result of this paper.

**Theorem II.1 (Exact recovery guarantee of Vectorized Hankel Lift):** Under Assumptions II.1 and II.2, \( \mathcal{G}(X^2) \) is the unique optimal solution to (II.4) with probability exceeding \( 1 - c_0(sn)^{-c_1} \), provided that \( n \gtrsim \mu_0 \mu_1 \cdot sr \log^3(sn) \), where \( c_0, c_1 \gtrsim 1 \) are absolute constants.

**Remark II.1:** Indeed, the constant \( c_1 \) in the successful probability can be chosen to be sufficiently large.

**Remark II.2:** The sampling complexity established in [62] for the atomic norm minimization method is \( n \gtrsim \mu_0 \cdot sr \log^3(sn) \). While this is slightly better than the sampling complexity for Vectorized Hankel Lift, our analysis is based on less stringent assumptions. In our analysis, the coefficients \( \{h_k\}_{k=1}^s \) are not required to be i.i.d. samples from the uniform distribution on the complex unit sphere, but can be any unit norm vectors. In addition, noting that the right-hand side of (II.9) is about \( 2/n \) for moderately large \( \mu_1 \), which is smaller than \( 4/n \), the separation required in the main result of [62]. Moreover, the numerical results in Section II-D justify that Vectorized Hankel Lift is less sensitive to frequency separation than atomic norm minimization. It is worth noting that the robust analysis of the atomic norm minimization method has been studied in [33] and we will leave the robust analysis of Vectorized Hankel Lift for future work.

The proof of Theorem II.1 follows a well established route that has been widely used for compressed sensing and low
B. Variants of MUSIC for Frequency Retrieval

In this section, we discuss the subspace method, particularly the MUSIC algorithm [51], for computing the frequency parameters \( \{ \tau_k \}_{k=1}^r \) from the matrix \( X^k \). Note that once \( \{ \tau_k \}_{k=1}^r \) are obtained, the weights \( \{ d_k, h_k \} \) can be computed by solving an overdetermined linear system. As can be seen later, applying the idea of the single snapshot MUSIC to \( H(X^k) \) yields a variant which is equivalent to the existing spatial smoothing technique proposed to improve the performance of the Multiple Measurement Vector (MMV) MUSIC.

The careful reader may notice that every single row of \( X^k \) is a spectrally sparse signal of the form (I.11), and moreover, all the rows share the same frequency parameters \( \{ \tau_k \}_{k=1}^r \). Thus we can apply the single snapshot MUSIC algorithm to a row of \( X^k \) for frequency retrieval. Let \( x_t = \sum_{k=1}^r d_k h_k[e^T] \). Recall that \( H(x_t) \) is the Hankel matrix of rank \( r \) and it admits the Vandermonde decomposition

\[
H(x_t) = E_L \text{diag}(d_1 h_1[e], \ldots, d_r h_r[e]) E_R^T.
\] (II.10)

Moreover, letting

\[
H(x_t)^T = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & V^* \end{bmatrix} \begin{bmatrix} V^* \\ \Sigma \end{bmatrix} \] (II.11)

be the SVD of \( H(x_t)^T \), where \( U \in \mathbb{C}^{n_2 \times r}, U_\perp \in \mathbb{C}^{n_2 \times (n_2-r)}, \Sigma \in \mathbb{R}^{r \times r}, V \in \mathbb{C}^{n_1 \times r} \) and \( V_\perp \in \mathbb{C}^{n_1 \times (n_1-r)} \), it is evident that \( U \) and \( E_R \) span the same column space. Note that \( E_R = [a_{\tau_1} \cdots a_{\tau_r}] \), where \( a_{\tau_k} = [1, \ldots, e^{-2\pi i x_{\tau_k}(n_2-r)}] \). It follows from the property of the Vandermonde matrix that

\[ a_\tau \in \text{Range}(E_R) \] if and only if \( \tau \in \{ \tau_1, \ldots, \tau_r \} \).

Therefore we conclude that \( \tau \in \{ \tau_1, \ldots, \tau_r \} \) if and only if \( 1/\|U_\perp a_\tau\|_2^2 = \infty \). The single snapshot MUSIC algorithm utilizes this idea to identify the frequencies, and it consists of the following two steps:

1) Compute the SVD of \( H(x_t)^T \) as in (II.11);

2) Identify \( \{ \tau_k \}_{k=1}^r \) as the \( r \) largest local maxima of the pseudospectrum: \( f(\tau) = 1/\|U_\perp a_\tau\|_2^2 \).

Here we present the single snapshot MUSIC algorithm directly based on the Vandermonde matrix \( H(x_t) \). Equivalently, it can be interpreted from the autocorrelation matrix model for signals, see for example [32] and references therein. In the noiseless setting, it is easy to see that the single snapshot MUSIC algorithm is able to compute \( \{ \tau_k \}_{k=1}^r \) exactly. When noise exists in \( x_t \), the procedure of the algorithm remains unchanged, but with the SVD of \( H(x_t)^T \) being replaced by the SVD of the noisy Hankel matrix and with \( U_\perp \) being the left singular vectors corresponding to the \( n_2-r \) smallest singular values. The stability analysis of the single snapshot algorithm is discussed in [41].

To motivate the new variant of the MUSIC algorithm for estimating the frequencies from \( X^k \), we note that \( E_R \) appears as a separate component both in the Vandermonde decomposition of \( H(x_t) \) and that of \( H(X^k) \), see (II.10) and (I.14). Therefore, we can replace the SVD of \( H(x_t)^T \) with the SVD of \( H(X^k)^T \) in the first step of the single snapshot MUSIC algorithm. This gives the following variant:

1) Compute the SVD of \( H(X^k)^T \):

\[
H(X^k)^T = \begin{bmatrix} U & U_\perp \end{bmatrix} \Sigma \begin{bmatrix} V^* \\ 0 \end{bmatrix}, \quad \text{where } U \in \mathbb{C}^{n_2 \times r} \text{ and } U_\perp \in \mathbb{C}^{n_2 \times (n_2-r)}.
\]

2) Identify \( \{ \tau_k \}_{k=1}^r \) as the \( r \) largest local maxima of the pseudospectrum: \( f(\tau) = 1/\|U_\perp a_\tau\|_2^2 \).

The following lemma establishes a connection between this variant and the single snapshot MUSIC, showing that the former one actually utilizes the SVD of the matrix formed by stacking all \( H(x_t) \) (\( t = 1, \ldots, s \)) together.

**Lemma II.2:** Let \( H(X^k) \) be a matrix constructed by stacking all \( H(x_t) \) on top of one another:

\[
\tilde{H}(X^k) = \begin{bmatrix} H(x_1) \\ \vdots \\ H(x_s) \end{bmatrix} \in \mathbb{C}^{sn_1 \times n_2}.
\]

There exists a permutation matrix \( P \in \mathbb{R}^{sn_1 \times sn_1} \) such that \( \tilde{H}(X^k) = P H(X^k) \).

**Proof:** Following the Vandermonde decomposition, the \( \ell \)-th block of \( \tilde{H}(X^k) \) can be rewritten as

\[
\tilde{H}(e_{\ell}^T X^k) = E_L \begin{bmatrix} d_1 \cdot h_1[e] \\ \vdots \\ d_r \cdot h_r[e] \end{bmatrix} E_R^T = (E_L \odot e_{\ell}^T H) \begin{bmatrix} d_1 \\ \vdots \\ d_r \end{bmatrix} E_R^T,
\]

where \( h_i \) is the \( i \)-th column of \( H \) and \( h_i[e] \) is the \( \ell \)-th entry of \( h_i \). Thus \( \tilde{H}(X^k) \) has the following decomposition

\[
\tilde{H}(X^k) = (H \odot E_L) D E_R^T.
\]
According to the commutative law in [68, Section 1.10.3], there exists a permutation matrix $P$ such that $H \otimes E_L = P(E_L \otimes H)$.

Based on Lemma II.2, we will see that the variant obtained by applying the single snapshot MUSIC idea to $\mathcal{H}(X^2)$ corresponds to the spatial smoothing technique (more precisely the forward only spatial smoothing technique). First, treating the rows of $X^2$ as i.i.d samples of a random signal whose covariance matrix can be used to compute the signal space $U$ as in (II.11), MMV MUSIC [53] uses the principal eigenspace of the empirical covariance matrix (up to a scaling factor $1/s$)

$$R = \sum_{i=1}^{s} \mathcal{H}(x_i)\mathcal{H}(x_i)^*.$$ 

to compute $U$. However, when the signal comes from coherence sources, the performance of MMV MUSIC will degrade. To deal with this difficulty, the forward only spatial smoothing technique proposes to increase the number of samples by partitioning each $x_i$ into $n_2$ overlapped short samples (with each short sample being of length $n_1$, where $n_1 + n_2 = n + 1$), and then construct the empirical covariance matrix from all the $s \cdot n_2$ short samples. A simple algebra yields that the new empirical covariance matrix is indeed given by (up to a scaling factor $1/(s n_2)$)

$$\hat{R} = \sum_{i=1}^{s} \mathcal{H}(x_i)\mathcal{H}(x_i)^*.$$ 

It is not hard to see that the principal eigenspace of $\hat{R}$ is the same as the principal singular vector space of $\mathcal{H}(X^2)$. Thus, by Lemma II.2, we know that the variant obtained by applying the single snapshot MUSIC idea to $\mathcal{H}(X^2)$ is equivalent to the spatial smoothing MUSIC. For more details about spatial smoothing, see [24], [25], [64].

**C. Extension to Higher Dimension**

Vectorized Hankel Lift and the analysis are easily extended to higher dimensional array recovery problem. For ease of exposition, we give a brief discussion of the two-dimensional (2D) case but emphasize that the situation in higher dimensions is similar. For the 2D blind super-resolution problem, the data matrix can be expressed as

$$Y_{j,\ell} = \sum_{k=1}^{r} d_k e^{-2\pi i (j \tau_{1k} + \ell \tau_{2k})} G_k[j, \ell],$$

where $d_k$ is the amplitude, $\tau_k := (\tau_{1k}, \tau_{2k})$ is the 2D frequency and $G_k$ corresponds to the Fourier samples of the unknown 2D point spread function. Letting $a_{\tau_k} = \begin{bmatrix} e^{-2\pi i \tau_{1k} - 1} & \cdots & e^{-2\pi i \tau_{1k} - (n-1)} \end{bmatrix}^T \in \mathbb{C}^n$ for $s = 1, 2$, the 2D data array can be rewritten in a more compact form:

$$Y = \sum_{k=1}^{r} d_k (a_{\tau_k} a_{\tau_k}^T) \circ G_k.$$ 

Likewise, we assume that there exists a subspace matrix $B \in \mathbb{C}^{n_2 \times s}$ such that $\text{vec}(G_k) = Bh_k$ for any $k = 1, \ldots, r$. Then

$$y := \text{vec}(Y) = \sum_{k=1}^{r} d_k \text{vec}(a_{\tau_k} a_{\tau_k}^T) \circ \text{vec}(G_k) = \sum_{k=1}^{r} d_k (a_{\tau_k} \circ a_{\tau_k}) \circ (Bh_k).$$

For any $0 \leq j, \ell \leq n - 1$, the $(jn + \ell)$th entry of $y$ is given by

$$y_{jn+\ell} = \sum_{k=1}^{r} d_k (a_{\tau_k} \circ a_{\tau_k})^T e_{jn+\ell}^T (b_{jn+\ell}^* h_k)$$

$$= \sum_{k=1}^{r} \text{trace} \left( d_k (a_{\tau_k} \circ a_{\tau_k})^T e_{jn+\ell}^T (b_{jn+\ell}^* h_k) \right)$$

$$= \sum_{k=1}^{r} \text{trace} \left( e_{jn+\ell} (b_{jn+\ell}^* a_{\tau_k} \circ a_{\tau_k})^T \right),$$

where $b_{jn+\ell}$ is the $(jn + \ell)$th column of $B^*$. Therefore, we have $y = A(X^2)$, where $X^2 = \sum_{k=1}^{r} d_k a_{\tau_k}^T \otimes (h_k a_{\tau_k}^T)$, and $A : \mathbb{C}^{s \times n^2} \rightarrow \mathbb{C}^{n^2 \times n^2}$ is a linear operator given by

$$[A(X)]_{jn+\ell} = \langle b_{jn+\ell} e^T_{jn+\ell}, X \rangle.$$ 

As in the 1D case, the blind super-resolution problem is essentially about recovering the target matrix $X^2$ from the observation vector $y$.

Note that the target matrix $X^2$ can be rewritten as the following block form:

$$X^2 = \begin{bmatrix} \sum_{k=1}^{r} d_k (h_k a_{\tau_k}^T)^T & & & \\ \vdots & \ddots & & \\ \sum_{k=1}^{r} d_k e^{-2\pi i \tau_{1k} - (n-1)} (h_k a_{\tau_k}^T)^T & & & \\ \\ \end{bmatrix},$$

Letting $X^2 := \sum_{k=1}^{r} d_k e^{-2\pi i \tau_{1k} - \ell} (h_k a_{\tau_k}^T)$, we define the two-fold vectorized Hankel lift of $X^2$ as follows:

$$\mathcal{H}(X^2) = \begin{bmatrix} \mathcal{H}(X^2_0) & \mathcal{H}(X^2_1) & \cdots & \mathcal{H}(X^2_{n_2-1}) \\ \mathcal{H}(X^2_1) & \mathcal{H}(X^2_2) & \cdots & \mathcal{H}(X^2_{n_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}(X^2_{n_1-1}) & \mathcal{H}(X^2_{n_1}) & \cdots & \mathcal{H}(X^2_{n_1+n_2-1}) \end{bmatrix},$$

where $\mathcal{H}(X^2)$ is the vectorized Hankel matrix defined in (1.13). It can be readily shown that $\mathcal{H}(X^2)$ has the following decomposition

$$\mathcal{H}(X^2) = \begin{bmatrix} (E_L \otimes H) Y^0 \\ (E_L \otimes H) Y^1 \\ \vdots \\ (E_L \otimes H) Y^{n_1-1} \end{bmatrix} D \begin{bmatrix} E_R Y^0^T \\ E_R Y^1^T \\ \vdots \\ E_R Y^{n_2-1}^T \end{bmatrix}^T,$$

where $E_L, E_R$ are two matrices defined in (1.20) and (1.16) but with the frequencies $\tau_{1k}$, $H = \begin{bmatrix} h_1 & \cdots & h_r \end{bmatrix} \in \mathbb{C}^{s \times r}$, $D = \text{diag}(d_1, \ldots, d_r)$ and $Y = \text{diag}(e^{-2\pi i \tau_{11}}, \ldots, e^{-2\pi i \tau_{1r}})$. 

If all frequencies $\tau_{1k}, \tau_{2k}$ are distinct and all $d_k$ are non-zeros, it is not hard to see that $\mathcal{H}(X^*)$ is a low rank matrix. Therefore, we can recover $X^*$ by solving the following convex programming

$$
\min_{X \in \mathbb{C}^r \times \mathbb{R}^2} \| \mathcal{H}(X) \|_s \quad \text{s.t.} \quad A(X) = y. \quad (II.13)
$$

The recovery guarantee of (II.13) can be similarly established in the following theorem. The proof details are overall similar to that for Theorem II.1, and thus are omitted.

**Theorem II.3:** Under Assumption II.1 and suppose $\sigma_{\min}(L^*L) \geq \frac{\mu_1^2}{\mu_2}$ and $\sigma_{\min}(R^*R) \geq \frac{\mu_2^2}{\mu_1}$, the data matrix $X^* \in \mathbb{C}^r \times \mathbb{R}^2$ is the unique optimal solution to (II.13) with probability at least $1 - c_0(sn)^{-c_1}$ for absolute constants $c_0, c_1 \geq 1$, provided that $n^2 \gtrsim \mu_1 \mu_2 \cdot sr \log^4(sn)$.

After the matrix $X^*$ is recovered, the frequency $\{\tau_k = (\tau_{1k}, \tau_{2k})\}_{k=1}^r$ can be estimated by a 2D-MUSIC algorithm [3], [40], [69] based on the two-fold vectorized Hankel matrix $\mathcal{H}(X^*)$ in (II.12), followed by the recovery of $\{d_k h_k\}_{k=1}^r$ through least-squares.

**D. Numerical Experiments**

In this section, we empirically evaluate the performance of Vectorized Hankel Lift for the recovery of $X^*$ in the blind super-resolution problem. Vectorized Hankel Lift is solved by SDPT3 [55] based on CVX [27]. The recovery ability of Vectorized Hankel Lift will be evaluated via the framework of empirical phase transition and we compare it with the atomic norm minimization method [62]. The locations $\{\tau_k\}_{k=1}^r$ of the point source signals are generated randomly from $[0, 1)$, while the amplitudes $\{d_k\}_{k=1}^r$ are generated via $d_k = (1 + 10^{c_k}) e^{-i\psi_k}$ with $\psi_k$ being uniformly sampled from $[0, 2\pi)$ and $c_k$ being uniformly sampled from $[0, 1]$. The subspace matrix $B$ are sampled from two random ensembles which all satisfy the conditions in Assumption II.1. The first one is the random submatrix sampled from the DFT matrix, and the other one is the random matrix whose entries satisfy the uniform distribution over $[-\sqrt{3}, \sqrt{3}]$. The coefficients $\{h_k\}_{k=1}^r$ are i.i.d. standard Gaussian random vectors followed by normalization. In our tests, 20 Monte Carlo trails are repeated for each problem instance and we report the probability of successful recovery out of those trials. A trail is declared to be successful if the relative reconstruction error of $X^*$ in terms of the Frobenius norm is less than $10^{-3}$.

We first fix $n = 64$ and vary the values of $r$ and $s$. Figure 1(a) and Figure 1(b) show the phase transitions of Vectorized Hankel Lift and atomic norm minimization method when the subspace matrix $B$ is randomly sampled from the DFT matrix and the locations of point sources are randomly generated without imposing the separation condition, and Figure 1(c) illustrates the phase transition of the atomic minimization method when the separation condition $\Delta := \min_{k \neq j} |\tau_k - \tau_j| \geq \frac{1}{n}$ is imposed. Here we omit the phase transition plot of Vectorized Hankel Lift for the frequency separation case because the plot is similar to Figure 1(a). It can be observed that the atomic norm minimization method has a higher phase transition curve when the separation condition is satisfied. However, in contrast to Vectorized Hankel Lift, its performance degrades severely when there is no frequency separation requirement. That is, Vectorized Hankel Lift is less sensitive to the separation condition. We also conduct the phase transition tests when the entries of $B$ are i.i.d. sampled from the uniform distribution over $[-\sqrt{3}, \sqrt{3}]$. The phase transition diagrams are presented in Figure 2, and similar observations can be made. Note that the phase transition plot of Vectorized Hankel Lift for the frequency separation case is still omitted due to the high similarity with Figure 2(a).

In the above phase transition tests, the coefficients $\{h_k\}_{k=1}^r$ are sampled from random Gaussian with normalization. In order to test whether the choice of $\{h_k\}_{k=1}^r$ matters, we also test another two cases for the coefficients. One is the Identical Gaussian, where $\{h_k\}_{k=1}^r$ are the same across $r$ (sampled from random Gaussian with normalization). The other one is QR where $\{h_k\}_{k=1}^r$ are obtained from the Q matrix in the QR decomposition of an $s \times r$ random Gaussian matrix. Tests are conducted for fixed $s = 4$ and $n = 64$, and the plots of successful recovery probability against the number of spikes $r$ are presented in Figure 3. It can be clearly seen that no significant differences over different types of $\{h_k\}_{k=1}^r$ are observed from the plots. Therefore, the numerical results
Fig. 2. The phase transitions of Vectorized Hankel Lift and the atomic norm minimization method when the entries of $B$ are i.i.d. sampled from the uniform distribution over $[-\sqrt{3}, \sqrt{3}]$. (a) Vectorized Hankel Lift for randomly generated frequencies, (b) atomic norm minimization for randomly generated frequencies, and (c) atomic norm minimization for frequencies obeying the separation condition $\Delta := \min_{k \neq j} |\tau_k - \tau_j| \geq \frac{1}{n}$. The number of measurements is fixed to be $n = 64$. The red curve plots the hyperbola curve $rs = 20$.

Fig. 3. The probability of successful recovery of Vectorized Hankel Lift against $r$ with three different subspace coefficients $\{h_k\}_{k=1}^r$. (a): The subspace matrix $B$ are randomly sampled from the DFT matrix. (b): The entries of $B$ are i.i.d. sampled from the uniform distribution over $[-\sqrt{3}, \sqrt{3}]$.

validate that our main result can hold without any conditions of $\{h_k\}_{k=1}^r$.

In order to examine the effect of the separation condition more carefully, we further conduct tests for fixed $s = 3, r = 3$, and vary the number of samples $n$. In the tests, we impose that there are at least two spikes with separation equal to $1.0/n$ and $0.5/n$, respectively. For each problem instance, we repeat 50 Monte Carlo trails and report the probability of successful recovery out of those trials. The numerical results are presented in Figure 4. It is evident that Vectorized Hankel Lift presents a better performance when the minimum separation is $0.5/n$. When the spikes are well separated (i.e., the minimum separation is $\Delta = 1.0/n$), the atomic norm minimization method performs better. In addition, the results confirm that Vectorized Hankel Lift is overall not affected by the separation condition.

We also plot the locations of the point sources $\{\tau_k\}_{k=1}^r$ and the unknown point spread function samples $\{g_k\}_{k=1}^r$ computed from $X^*$ for a random instance corresponding to $n = 64, s = 3$ and $r = 4$. We apply the MUSIC variant introduced in Section II-B (i.e., the spatial smoothing MUSIC) to localize the $\{\tau_k\}_{k=1}^r$. Figure 5(a) shows the pseudospectrum $f(\tau)$ on a set of points on $[0, 1]$ with equal distance $10^{-4}$. As can be seen from this figure, the function $f(\tau)$ peaks at the locations of true frequencies. After the $\{\tau_k\}_{k=1}^r$ are identified, the coefficients $\{h_k\}_{k=1}^r$ are computed by solving a least squares problem and $\{g_k\}_{k=1}^r$ are estimated as $Bh_k$. Figure 5(b) includes the plots of the estimates of $\{|g_k|\}_{k=1}^r$ against the true values which clearly show that $\{g_k\}_{k=1}^r$ can be recovered.

III. PROOF ARCHITECTURE OF MAIN RESULT

A. Preliminaries

We first apply the bounded difference inequality to show that for the column vectors $\{b_j\}_{j=0}^{n-1}$ with independent entries, the condition (II.7) in Assumption II.1 holds with high probability given (II.5) and (II.6).

**Lemma III.1:** The column vectors $\{b_j\}_{j=0}^{n-1}$ of the subspace matrix $B^*$ are independently and identically sampled from a distribution $F$ which obeys the conditions (II.5) and (II.6) in
Fig. 4. The probability of successful recovery of Vectorized Hankel Lift and the atomic norm minimization method under two separation conditions, $\Delta = \frac{0.5}{n}$ and $\Delta = \frac{1}{n}$. The dimension of subspace and the number of spikes are both fixed to be $s = 3$ and $r = 3$. The number of samples $n$ is varied. (a): The subspace matrix $B$ are randomly sampled from the DFT matrix. (b): The entries of $B$ are i.i.d. sampled from the uniform distribution over $[-\sqrt{3}, \sqrt{3}]$.

Assumption II.1. Assume the components of $b$ are independent, the event

$$\min_{0 \leq j \leq n-1} \|b_j\|^2 \geq 1 \quad \text{(III.1)}$$

occurs with probability at least $1 - n \exp\left(-\frac{s}{16\mu_0^2}\right)$.

Proof: Since $b_j$ satisfies (II.5), we first have

$$\mathbb{E}\left[\|b_j\|^2\right] = \mathbb{E}\left[\text{trace}(b_j^*b_j)\right] = \mathbb{E}\left[\text{trace}(b_jb_j^*)\right] = s.$$ 

Define $f(x_1, \ldots, x_s) = \sum_{i=1}^s |x_i|^2$. It is evident that

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_s) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_s)| \leq |x_i|^2 + |x_i'|^2 \leq 2\mu_0$$

when $|x_i|^2 \leq \mu_0$ and $|x_i'|^2 \leq \mu_0$. Because $b_j$ also satisfies (II.6), the application of the bounded difference inequality yields that

$$\mathbb{P}\left[\|b_j\|^2 - s \geq \frac{s}{2}\right] \geq 1 - \exp\left(-\frac{s^2}{16\mu_0}\right).$$

Consequently, we can take $t = \frac{s}{2}$ to obtain

$$\mathbb{P}\left[\|b_j\|^2 \geq \frac{s}{2}\right] \geq 1 - \exp\left(-\frac{s^2}{16\mu_0}\right).$$

Taking the uniform bound yields that for all $j \in [n]$, with probability at least $1 - n \exp\left(-\frac{s}{16\mu_0}\right)$, $\|b_j\|^2 \geq \frac{s}{2} \geq 1$ when $s \geq 2$. \qed
Next, we present a lemma about the basic properties of the linear operator $A$.

**Lemma III.2:** Under Assumption II.1, the following properties hold:

$$\langle y, AA^*(y) \rangle \geq \|y\|_2^2$$

for any fixed vector $y \in \mathbb{C}^n$. (III.2)

$$\|AA^* - I\| \leq s_{\mu_0} \text{ and } \|A\| \leq \sqrt{s_{\mu_0}}.$$  (III.3)

**Proof:** Since

$$AA^*(y) = A \left( \sum_{i=0}^{n-1} y[i] b_i e_i^T \right) = \left[ \begin{array}{c} \langle b_0 e_0^T, \sum_{i=0}^{n-1} y[i] b_i e_i^T \rangle \\ \vdots \\ \langle b_{n-1} e_{n-1}^T, \sum_{i=0}^{n-1} y[i] b_i e_i^T \rangle \end{array} \right] = \left[ \begin{array}{c} \|b_0\|_2^2 \cdot y[0] \\ \vdots \\ \|b_{n-1}\|_2^2 \cdot y[n-1] \end{array} \right] \in \mathbb{C}^n,$$

(III.2) follows immediately from (II.7). The properties in (III.3) follows directly from the definition of $A$. For the left inequality, we have

$$\|AA^* - I\| = \sup_{y \in \mathbb{C}^n: \|y\|_2 = 1} \|AA^*(y) - y\|_2 = \sup_{y \in \mathbb{C}^n: \|y\|_2 = 1} \left( \sum_{i=0}^{n-1} \left( \|b_i\|_2^2 - 1 \right)^2 \cdot |y[i]|^2 \right) \leq \max_{0 \leq i \leq n-1} \left( \|b_i\|_2^2 - 1 \right) \leq s_{\mu_0}.$$

The right one can be proved as follows

$$\|A\| = \sup_{X \in \mathbb{C}^{s \times n}: \|X\|_F = 1} \|A(X)\|_2 = \sup_{X \in \mathbb{C}^{s \times n}: \|X\|_F = 1} \left( \sum_{i=0}^{n-1} \|b_i^T X e_i\|_2^2 \right) \leq \sup_{X \in \mathbb{C}^{s \times n}: \|X\|_F = 1} \left( \sum_{i=0}^{n-1} \|b_i\|_2^2 \cdot \|X e_i\|_2^2 \right) \leq \max_{0 \leq i \leq n-1} \|b_i\|_2 \cdot \sup_{X \in \mathbb{C}^{s \times n}: \|X\|_F = 1} \left( \sum_{i=0}^{n-1} \|X e_i\|_2^2 \right) \leq \sqrt{s_{\mu_0}}.$$

The proof is now complete.

The following lemma suggests that the smallest singular value of $E_{h, L}$ can be lower bounded by the smallest singular value of $E_L$.

**Lemma III.3:** Recall that $H = \begin{bmatrix} h_1 & \cdots & h_r \end{bmatrix} \in \mathbb{C}^{s \times r}$ and suppose all columns of $H$ are of unit norm. Under the incoherence condition (II.8), we have

$$\sigma_{\min}(E_{h, L}) \geq \frac{n_1}{\mu_1},$$

where $E_{h, L}$ is the matrix defined in (I.15).

**Proof:** Let $a_{\tau_i} = [1 \quad e^{-2\pi i r_1} \cdots e^{-2\pi i r_{(n-1)}}]^T \in \mathbb{C}^{n_1}$ be the $i$th column of $E_L$. Since $E_{h, L} = E_L \odot H$, it can be easily seen that

$$E_{h, L} = \begin{bmatrix} a_{\tau_1} \otimes h_1^* \\ \vdots \\ a_{\tau_{n_1}} \otimes h_{r}^* \end{bmatrix} = \begin{bmatrix} (a_{\tau_1}^* \otimes h_1^*)(a_{\tau_1} \otimes h_1) & \cdots & (a_{\tau_1}^* \otimes h_1^*)(a_{\tau_r} \otimes h_r) \\ \vdots & \ddots & \vdots \\ (a_{\tau_{n_1}}^* \otimes h_{r}^*)(a_{\tau_1} \otimes h_1) & \cdots & (a_{\tau_{n_1}}^* \otimes h_{r}^*)(a_{\tau_r} \otimes h_r) \end{bmatrix}.$$

Recall that a selection matrix $P \in \mathbb{R}^{n_2 \times n}$ is the unique matrix such that

$$Pz = \text{vec} (\text{diag}(z)) \quad \text{for all } z \in \mathbb{C}^n,$$

and it has the remarkable property that $P^T (A \otimes B) P = A \circ B$ [59, Corollary 2]. Thus we have

$$\sigma_{\min}(E_{h, L}) = \inf_{\beta \geq 1} \left| \beta^* \left( (E_L^* E_L) \circ (H^* H) \right) \beta \right| = \inf_{\beta \geq 1} \left| \beta^* P^T \left( (E_L^* E_L) \circ (H^* H) \right) P \beta \right| = \inf_{\beta \geq 1} \left| \beta^* P^T \left( E_L^* \otimes H^* \right) (E_L \otimes H) P \beta \right| = \inf_{\beta \geq 1} \left\| \left( E_L \otimes H \right) \text{vec} (\text{diag}(\beta)) \right\|_2^2 = \inf_{\beta \geq 1} \left\| \left( (E_L \otimes H) \text{vec} (\text{diag}(\beta)) \right)^T \right\|_2^2 = \inf_{\beta \geq 1} \left\| H \text{diag}(\beta) E_L^T \right\|_F^2 \geq \sigma_{\min}(E_L) \cdot \inf_{\beta \geq 1} \left\| H \text{diag}(\beta) \right\|_F^2$$

which completes the proof.

A straightforward application of Lemma III.3 yields the following result, which can be regarded as a variant of [7, Lemma 1].
Lemma III.4: Suppose $\mathcal{H}(X^2)$ obeys the incoherence condition (II.8) with parameter $\mu_1$. Let $\mathcal{H}(X^2) = USV^*$ be the singular value decomposition of $\mathcal{H}(X^2)$, where $U \in \mathbb{C}^{n_1 \times r}$, $S \in \mathbb{R}^{r \times r}$ and $V \in \mathbb{C}^{n_2 \times r}$. If we rewrite $U$ as

$$U = \begin{bmatrix} U_0 & \cdots & U_{n_1-1} \end{bmatrix},$$

where the $\ell$th block is $U_\ell = U(\ell s : (\ell + 1)s - 1, :)$ for $\ell = 0, \ldots, n_1 - 1$, then

$$\max_{0 \leq \ell s \leq n_1 - 1} \left\| U_\ell \right\|_F^2 \leq \frac{\mu_1 r}{n}, \quad \max_{0 \leq j \leq n_2 - 1} \left\| e_j^T V \right\|_2^2 \leq \frac{\mu_1 r}{n}. \quad \text{(III.4)}$$

Proof: We only need to prove the left inequality in (III.4) as the right one can be similarly established. Recall that $\mathcal{H}(X^2) = E_h L \operatorname{diag}(d_1, \ldots, d_r) E_h^T$. Since $U \in \mathbb{C}^{n_1 \times r}$ and $E_h$ span the same subspace and $U$ is orthogonal, there exists an orthonormal matrix $Q \in \mathbb{C}^{r \times r}$ such that $U = E_h L (E_h L E_h)_{1/2}^{-1} Q$. Therefore,

$$\left\| U_\ell \right\|_F^2 = \sum_{j=0}^{(\ell + 1)s - 1} \left\| e_j^T E_h L (E_h L E_h)_{1/2}^{-1} \right\|_2^2 \leq \sum_{j=0}^{(\ell + 1)s - 1} \left\| e_j^T E_h L \right\|_2 \left\| (E_h L E_h)_{1/2}^{-1} \right\|_2^2 \leq \frac{\mu_1 r}{n},$$

$$\sum_{j=0}^{(\ell + 1)s - 1} \left\| e_j^T E_h L \right\|_2 \left\| (E_h L E_h)_{1/2}^{-1} \right\|_2^2 \leq \frac{\mu_1 r}{n},$$

$$\sum_{j=0}^{(\ell + 1)s - 1} \left\| e_j^T E_h L \right\|_2 \left\| (E_h L E_h)_{1/2}^{-1} \right\|_2^2 \leq \frac{\mu_1 r}{n},$$

$$\sum_{j=0}^{(\ell + 1)s - 1} \left\| e_j^T E_h L \right\|_2 \left\| (E_h L E_h)_{1/2}^{-1} \right\|_2^2 \leq \frac{\mu_1 r}{n},$$

where the second inequality is due to Lemma III.3.

The following corollary is a direct consequence of Lemma III.4 and will be frequently used in the sequel.

Corollary III.5: Suppose $\mathcal{H}(X^2)$ obeys the incoherence condition (II.8) with parameter $\mu_1$. Then,

$$\max_{0 \leq i \leq n_1 - 1} \frac{1}{w_i} \sum_{0 \leq j \leq n_2 - 1} \left\| U_{i,j} \right\|_F^2 \leq \frac{\mu_1 r}{n}. \quad \text{(III.5)}$$

The matrix Bernstein inequality, stated below, will be used frequently in our analysis.

Lemma III.6 ([56], [58]): Let $\{X_\ell\}_{\ell=1}^n$ be a set of independent random matrices of dimension $n_1 \times n_2$, which satisfy $\mathbb{E} \left[ X_\ell \right] = 0$ and $\left\| X_\ell \right\| \leq B$. Define

$$\sigma^2 := \max \left\{ \mathbb{E} \left[ n \sum_{\ell=1}^n X_\ell^T X_\ell \right], \mathbb{E} \left[ n \sum_{\ell=1}^n X_\ell^T X_\ell \right] \right\}.$$

Then for all $t > 0$,

$$\mathbb{P} \left( \sum_{\ell=1}^n X_\ell \geq t \right) \leq (n_1 + n_2) \cdot \exp \left( \frac{-t^2}{2 \sigma^2 + 2Bt/3} \right).$$

We can directly obtain the following corollary from Lemma III.6 by choosing

$$t = c \left( \frac{2}{3} B \log a + \sqrt{2\sigma^2 \log a} \right),$$

where $c > 1$ and $a > 0$.

Corollary III.7: Under the same condition of Lemma III.6, for any $c > 1$ and $a > 0$, one has

$$\sum_{\ell=1}^n X_\ell \leq c \left( \sqrt{2\sigma^2 \log a} + \frac{2}{3} B \log a \right). \quad \text{(III.7)}$$

with probability at least $1 - (n_1 + n_2) a^{-c}$.

B. Deterministic Optimality Condition

As is typical in the analysis of low rank matrix recovery, in order to show that $Z^2$ is the unique optimal solution to the convex program (II.4), we need to construct a dual certificate which satisfies a set of sufficient conditions. These conditions can be viewed as a variant of the KKT condition for the optimality of $Z^2$. Recall that the singular value decomposition (SVD) of $\mathcal{H}(X^2)$ is $\mathcal{H}(X^2) = USV^*$. The tangent space $T$ of the nuclear norm at $\mathcal{H}(X^2)$ can be defined as

$$T = \{ UA^* + BV^* : A \in \mathbb{C}^{n_2 \times r}, B \in \mathbb{C}^{n_1 \times r} \}.$$

The projections $P_T(Z)$ onto the tangent space can be defined as

$$P_T(Z) := UU^* Z + ZV V^* - UU^* Z V V^*, \quad \text{(III.8)}$$

and the corresponding projector onto the orthogonal complement of $T$ is given by $P_{T^\perp}(Z) = Z - P_T(Z)$.

Theorem III.8: Suppose $\| A A^* \| \geq 1$ and

$$\| P_T G A^* A G^* P_T - P_T G G^* P_T \| \leq \frac{1}{2}. \quad \text{(III.9)}$$

If there exists a dual certificate $A \in \mathbb{C}^{n_1 \times n_2}$ such that

$$\| P_{T^\perp}(A) \| \leq \frac{1}{2},$$

$$G^*(A) \in \text{Range}(A^*),$$

then $Z^2$ is the unique solution to (II.4).

Proof: The structure of the proof is overall similar to those in [14]–[16]. Consider any feasible solution $Z^2 + M$, where the perturbation $M \in \mathbb{C}^{n_1 \times n_2}$ satisfies

$$A G^*(M) = 0,$$

$$\| I - G^* G^* \| = 0.$$
In the meantime, we have $UV^* + S \in \partial \left\| Z^2 \right\|_*$. Thus,

$$\Delta : = \left\| Z^2 + M \right\|_* - \left\| Z^2 \right\|_* \geq \langle UV^* + S, M \rangle$$

$$= \langle UV^*, M \rangle + \| P_{T^\perp} (M) \|_*$$

$$\geq \| P_{T^\perp} (M) \|_* - \| \langle UV^* - \Lambda, M \rangle \| - \| \Lambda \|. \quad \text{(III.15)}$$

The condition (III.12) directly implies that there exists a vector $p \in \mathbb{C}^n$ such that

$$G^* (\Lambda) = A^* (p).$$

Therefore, combining (III.12) and (III.14), we obtain

$$\| A, M \| = \| \langle A, G^* (M) \rangle \| = \| G^* (A), G^* (M) \| = \| (A^* (p), G^* (M)) \| = \langle p, A^* (M) \rangle = 0.$$ 

Moreover, the second term of (III.15) can be upper bounded as follows:

$$\| \langle UV^* - \Lambda, M \rangle \| \leq \| \langle P_{T^\perp} (UV^* - \Lambda), M \rangle \| + \| P_{T^\perp} (UV^* - \Lambda), M \rangle \| \| P_{T^\perp} (M) \|_F$$

$$\leq \| P_{T^\perp} (UV^* - \Lambda) \|_F \cdot \| P_{T^\perp} (M) \|_F + \| P_{T^\perp} (\Lambda) \| \cdot \| P_{T^\perp} (M) \|_*$$

$$\leq \frac{1}{16 \mu_0} \cdot \| P_{T^\perp} (M) \|_F + \frac{1}{2} \cdot \| P_{T^\perp} (M) \|_*,$$

where the last step is due to (III.10) and (III.11). Consequently,

$$\Delta \geq \| P_{T^\perp} (M) \|_* - \| \langle UV^* - \Lambda, M \rangle \| - \| \Lambda \|$$

$$\geq \frac{1}{2} \cdot \| P_{T^\perp} (M) \|_* - \frac{1}{16 \mu_0} \cdot \| P_{T^\perp} (M) \|_F$$

$$\geq \frac{1}{2} \cdot \| P_{T^\perp} (M) \|_F - \frac{1}{16 \mu_0} \cdot \| P_{T^\perp} (M) \|_F$$

$$\geq \left( \frac{1}{2} - \frac{1}{16 \mu_0} \right) \| P_{T^\perp} (M) \|_F$$

$$= \frac{1}{4} \| P_{T^\perp} (M) \|_F,$$

where the fourth line is due to Lemma VI.1 in Section VI. It follows that $\Delta > 0$ unless $\| P_{T^\perp} (M) \|_F = 0$.

Note that $\Delta = 0$ requires $P_{T^\perp} (M) = 0$, which in turn requires $M = P_T (M)$. In this case, we have

$$\| P_T (M) \|_F^2 = \langle P_T (M), M \rangle$$

$$= \langle P_T (M), G^* (M) \rangle$$

$$= \langle M, P_T G^* P_T (M) - P_T G^* P_T (M) \rangle$$

$$+ \langle M, P_T G^* A^* P_T (M) \rangle$$

$$= \langle M, P_T G^* P_T (M) - P_T G^* A^* P_T (M) \rangle$$

$$\leq \| P_T G^* A^* P_T - P_T G^* P_T \| \cdot \| P_T (M) \|_F^2$$

$$\leq \frac{1}{2} \| P_T (M) \|_F^2,$$

which implies that $P_T (M) = 0$. Thus $Z^2$ is the unique minimizer.

C. Constructing the Dual Certificate

It is intuitively clear that we may construct a dual certificate $\Lambda \in \mathbb{C}^{n_1 \times n_2}$ obeying the conditions (III.10), (III.11) and (III.12) by solving the following constrained least squares problem:

$$\min_{\Lambda} \| P_T (UV^* - \Lambda) \|_F^2 \text{ s.t. } G^* (\Lambda) \in \text{Range}(A^*).$$

Here only the conditions (III.10) and (III.12) are taken into account because once $\| P_T (UV^* - \Lambda) \|_F$ is small, the projection of $\Lambda$ onto $T^\perp$ can be simultaneously small.

Applying the projected gradient method to solve the above optimization problem, we obtain the following update rule:

$$Y^k = Y^{k-1} + (GA^* A G^* + I - GG^*) P_T (UV^* - Y^{k-1}).$$

However, due to the statistical dependence among the iterations, the convergence analysis of the vanilla gradient iteration is difficult. Therefore, the golfsing scheme [28] proposes to break the statistical dependence by dividing all the linear measurements into a few disjoint partitions and use a fresh partition in each iteration.

Assume we divide the linear measurements in (I.8) into $k_0$ partitions, denoted $\Omega_k \subseteq \Omega_{k-1}$, and let $m = \frac{n}{k_0}$. Define

$$A_k (X) = \{ (b, e_i^T, X) \}_{i \in \Omega_k} \subseteq \mathbb{C}^{[\Omega_k]} \quad \text{(III.16)}$$

and

$$A_k^* A_k (X) = \sum_{i \in \Omega_k} \langle b, e_i^T, X \rangle b_i e_i^T$$

$$= \sum_{i \in \Omega_k} b_i b_i^T X e_i e_i^T \in \mathbb{C}^{s \times n}. \quad \text{(III.17)}$$

Then the golf scheme for the construction of $\Lambda$ satisfying the conditions in Theorem III.8 can be formally expressed as

$$Y^0 = 0 \in \mathbb{C}^{n_1 \times n_2},$$

$$Y^k = Y^{k-1} + \left( \frac{n}{m} G A_k^* A_k G^* + I - GG^* \right) P_T (UV^* - Y^{k-1}),$$

for $k = 1, \ldots, k_0$, \quad \text{(III.18)}

$$\Lambda := Y^{k_0}.$$ 

Evidently, the property of $\Lambda$ relies on the partitions $\{\Omega_k\}_{k=1}^{k_0}$. In order to construct the desirable $\Lambda$, we require $\{\Omega_k\}_{k=1}^{k_0}$ to satisfy the first order conditions of the following lemma, in which we have

$$\| Z \|_{G,F} = \left( \sum_{i=0}^{n-1} \frac{\| G^* (Z) e_i \|_2^2}{w_i} \right)^{\frac{1}{2}}, \quad \text{(III.19)}$$

$$\| Z \|_{G,\infty} = \max_{0 \leq i \leq n-1} \frac{\| G^* (Z) e_i \|_2}{\sqrt{w_i}} \quad \text{(III.20)}$$

for any $Z \in \mathbb{C}^{n_1 \times n_2}$. The proof of this lemma will be presented in Section IV.
Lemma III.9: Let \( k_0 \in \{1, \ldots, n\} \) and set \( m = \frac{n}{k_0} \). If \( n \geq k_0 \) then \( \max \{ \mu_1 r \log(sn), \log(k_0) \} \), there exists a partition \( \{ \Omega_k \}_{k=1}^{k_0} \) such that the following properties hold:

\begin{equation}
\frac{m}{2} \leq |\Omega_k| \leq \frac{3m}{2}, \quad k = 1, \ldots, k_0, \tag{III.21}
\end{equation}

\begin{equation}
\max_{1 \leq k \leq k_0} \left\| P_T G \left( I - \frac{n}{m} E [A_k^* A_k] \right) G^* P_T \right\| \leq \frac{1}{4}, \tag{III.22}
\end{equation}

\begin{equation}
\max_{1 \leq k \leq k_0} \left\| G \left( I - \frac{n}{m} E [A_k^* A_k] \right) G^* (Z) \right\| \leq \sqrt{\frac{n \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F} + \frac{n \log(sn)}{m} \left\| Z \right\|_{\dot{g}, \infty}, \tag{III.23}
\end{equation}

\begin{equation}
\max_{1 \leq k \leq k_0} \left\| P_T G \left( I - \frac{n}{m} E [A_k^* A_k] \right) G^* (Z) \right\|_{\dot{g}, F} \leq \sqrt{\frac{\mu_1 r \log(sn)}{n}} \left( \sqrt{\frac{n \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F} + \frac{n \log(sn)}{m} \left\| Z \right\|_{\dot{g}, \infty} \right), \tag{III.24}
\end{equation}

\begin{equation}
\max_{1 \leq k \leq k_0} \left\| P_T G \left( I - \frac{n}{m} E [A_k^* A_k] \right) G^* (Z) \right\|_{\dot{g}, \infty} \leq \frac{\mu_1 r}{n} \left( \sqrt{\frac{n \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F} + \frac{n \log(sn)}{m} \left\| Z \right\|_{\dot{g}, \infty} \right). \tag{III.25}
\end{equation}

Here \( Z \in \mathbb{C}^{sn_1 \times n_2} \) is fixed. Recalling the definition of the operator \( A_k^* A_k \) in (III.17), the expectation is taken with respect to \( \{ b_i \}_{i \in \Omega_k} \).

D. Validating the Dual Certificate and Completing the Proof

In this section we show that the dual certificate \( \Lambda \) constructed from the iteration (III.18) satisfies the conditions in Theorem III.8. The result follows from several lemmas that will be proved in Section V. In these lemmas, \( \{ \Omega_k \}_{k=1}^{k_0} \) is a partition of \( \{1, \ldots, n\} \) satisfying the conditions in Lemma III.9, and \( \{ A_k \}_{k=1}^{k_0} \) are the associated linear operators defined in (III.16). Note that we assume (III.4) holds in the remainder of this paper, which follows from Assumption II.2 and Lemma III.4.

Lemma III.10: Assume \( n \geq k_0 s \mu_0 \cdot \mu_1 r \log(sn) \). Under the condition (III.22) of Lemma III.9, the event

\begin{equation}
\max_{1 \leq k \leq k_0} \left\| P_T G \left( I - \frac{n}{m} E [A_k^* A_k] \right) G^* P_T \right\| \leq \frac{1}{2} \tag{III.26}
\end{equation}

occurs with probability at least \( 1 - (sn)^{-c_1} \) for a universal constant \( c_1 \geq 1 \).

The following corollary is the special case of Lemma III.10 when \( k_0 = 1 \) and \( n = m \).

Corollary III.11: Assume \( n \geq s \mu_0 \cdot \mu_1 r \log(sn) \). The event

\begin{equation}
\left\| P_T G A_k^* A_k^* P_T - P_T G G^* P_T \right\| \leq \frac{1}{2} \tag{III.27}
\end{equation}

occurs with probability at least \( 1 - (sn)^{-c_1} \) for a universal constant \( c_1 \geq 1 \).

Lemma III.12: Under the condition (III.23) of Lemma III.9, for any \( 1 \leq k \leq k_0 \) and fixed \( Z \in \mathbb{C}^{sn_1 \times n_2} \), the event

\begin{equation}
\left\| \left( \frac{n}{m} G A_k^* A_k G^* - G G^* \right) (Z) \right\| \leq \sqrt{\frac{4nk_0 s \mu_0 \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F},
\end{equation}

occurs with probability at least \( 1 - (sn)^{-c_1} \) for a universal constant \( c_1 \geq 1 \).

Lemma III.13: Under the condition (III.24) of Lemma III.9, for any \( 1 \leq k \leq k_0 \) and fixed \( Z \in \mathbb{C}^{sn_1 \times n_2} \), the event

\begin{equation}
\left\| P_T G \left( I - \frac{n}{m} A_k^* A_k \right) G^* (Z) \right\|_{\dot{g}, F} \leq \sqrt{\frac{\mu_1 r \log(sn)}{n}} \left( \sqrt{\frac{4nk_0 s \mu_0 \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F} + \frac{2ns \mu_0 \log(sn)}{m} \left\| Z \right\|_{\dot{g}, \infty} \right), \tag{III.29}
\end{equation}

occurs with probability at least \( 1 - (sn)^{-c_1} \) for a universal constant \( c_1 \geq 1 \).

Lemma III.14: Under the condition (III.25) of Lemma III.9, for any \( 1 \leq k \leq k_0 \) and fixed \( Z \in \mathbb{C}^{sn_1 \times n_2} \), the event

\begin{equation}
\left\| P_T G \left( I - \frac{n}{m} A_k^* A_k \right) G^* (Z) \right\|_{\dot{g}, \infty} \leq \frac{\mu_1 r}{n} \left( \sqrt{\frac{4nk_0 s \mu_0 \log(sn)}{m}} \left\| Z \right\|_{\dot{g}, F} + \frac{2ns \mu_0 \log(sn)}{m} \left\| Z \right\|_{\dot{g}, \infty} \right), \tag{III.30}
\end{equation}

occurs with probability at least \( 1 - (sn)^{-c_1} \) for a numerical constant \( c_1 \geq 1 \).

Lemma III.15: Recalling that \( U \) and \( V \) satisfy (III.4), we have

\begin{equation}
\| UV^* \|^2_{\dot{g}, F} \leq \frac{\mu_1 r \log(sn)}{n}, \tag{III.31}
\end{equation}

Equipped with these lemmas, we are in position to validate the conditions in Theorem III.8. Note that \( \| A A^* \| \geq 1 \) holds due to (III.2) in Lemma III.2, and (III.9) is proved in Corollary III.11. As for (III.12), it follows immediately from the construction of \( \Lambda \). Thus, it remains to validate (III.10) and (III.11).

a) Validating (III.10): A simple calculation yields that

\begin{equation}
E^k := P_T \left( UV^* - Y^k \right)
= P_T \left( UV^* - Y^{k-1} \right) - \left( \frac{n}{m} G A_k^* A_k G^* + I - G G^* \right) P_T (E^{k-1})
= P_T (E^{k-1}) - P_T \left( \frac{n}{m} G A_k^* A_k G^* + I - G G^* \right) P_T (E^{k-1})
= P_T \left( G G^* - \frac{n}{m} G A_k^* A_k G^* \right) P_T (E^{k-1}),
\end{equation}

where the second line is due to (III.18). By the construction of \( \Lambda \), we can obtain

\begin{equation}
\| P_T (UV^* - \Lambda) \|_F = \left\| E^{k_0} \right\|_F
\end{equation}
where the second line follows from the fact that

Applying Lemma III.13 and Lemma III.14 yields that

Then it follows that

where step (a) is due to Lemma III.10 and the last inequality holds when \( k_0 = \lceil \log_2(16s\mu_0) \rceil \).

b) Validating (III.11): First recall that \( E^k := P_T \left( UV^* - Y^k \right) \). According to (III.18), we have

\[
\Lambda = \sum_{k=1}^{k_0} \left( \frac{n}{m} G A_k^* A_k G^* + I - GG^* \right) (E^{k-1}).
\]

Then it follows that

where the second line follows from the fact that \( E^{k-1} \in T \).

For any \( 1 \leq k \leq k_0 \), Lemma III.12 implies that

Recalling from the equality (III.32), we have

\[
E^{k-1} = P_T \left( GG^* - \frac{n}{m} G A_{k-1}^* A_{k-1} G^* \right) P_T (E^{k-2}).
\]

Applying Lemma III.13 and Lemma III.14 yields that

\[
\left\| P_T (E^{k-1}) \right\|_{G,F} = \left\| P_T (I - \frac{n}{m} A_{k-1}^* A_{k-1}) G^* P_T (E^{k-2}) \right\|_{G,F} \leq \left\| P_T (I - \frac{n}{m} A_{k-1}^* A_{k-1}) G^* (E^{k-2}) \right\|_{G,F}
\]

\[
\leq \left( \frac{1}{2} \frac{1}{k_0} \right) \left\| E^{k-1} \right\|_{G,F} \leq \frac{1}{k_0} \left\| E^k \right\|_{G,F} \leq \frac{1}{16s\mu_0},
\]

where step (a) holds provided \( m > k_0 s\mu_0 \log^2 (sn) \).

Finally, noting that \( E^0 = UV^* \), the application of Lemma III.15 gives

After substituting (III.35) and (III.36) into (III.34), we have

\[
\left\| \left( \frac{n}{m} G A_k^* A_k G^* - GG^* \right) (E^{k-1}) \right\|_{G,F} \leq \left( \frac{1}{2} \frac{1}{k_0} \right) \left\| E^0 \right\|_{G,F} \leq \frac{1}{16s\mu_0},
\]

where the final inequality holds when \( k_0 = \lceil \log_2(16s\mu_0) \rceil \).

Therefore, we have completed the proof.
\[ \begin{aligned}
&\leq \frac{1}{2} \left( \sqrt{\frac{4k_0 \mu_0 \rho_1 r \log^2(sn)}{m}} + \frac{2 \mu_0 \rho_1 r \log(sn)}{m} \right) \\
&\leq \frac{1}{2}
\end{aligned} \]

when \( m \geq k_0 \mu_0 \rho_1 r \log^2(sn) \), where the first inequality follows from (III.33).

Thus we have shown that the dual certificate \( \Lambda \) constructed from the iteration (III.18) satisfies the conditions in Theorem III.8 with probability at least \( 1 - c_0(sn)^{-c_1} \) provided that \( n = m k_0 \geq \mu_0 \mu_1 \cdot s r \log^2(sn) \). Corollary III.11 implies (III.9) holds with probability at least \( 1 - (sn)^{-c_1} \) if \( n \geq \mu_0 \mu_1 \cdot s r \log(sn) \). Taking an upper bound on the number of measurements completes the proof of Theorem II.1.

IV. PROOF OF LEMMA III.9

In this section, we will use probabilistic argument to show that the events (III.21) - (III.25) occur with high probability if we construct \( \{\Omega_k\}_{k=1}^{k_0} \) in a random manner and thus conclude that there at least exists a partition satisfying (III.21) - (III.25).

Let \( \{\epsilon_i\}_{i=0}^{n-1} \) be \( n \) independent random variables, each of which takes in \( \{1, \cdots, k_0\} \) uniformly at random. For any \( k \in \{1, \cdots, k_0\} \), we construct \( \{\Omega_k\}_{k=1}^{k_0} \) as follows:

\[ \Omega_k = \{i \in [n] : \epsilon_i = k\} \]

Clearly, \( \{\Omega_k\}_{k=1}^{k_0} \) form a partition of \([n]\). For any fixed \( k \in \{1, \cdots, k_0\} \), we also have

\[ P\{i \in \Omega_k\} = P\{\epsilon_i = k\} = \frac{1}{k_0} \]

for all \( i = 0, \cdots, n - 1 \). Therefore \( |\Omega_k| \) can be viewed as the sum of Bernoulli random variables, i.e.,

\[ |\Omega_k| = \sum_{i=0}^{n-1} 1\{i \in \Omega_k\} = \sum_{i=0}^{n-1} \delta_i \]

where \( \{\delta_i\}_{i=0}^{n-1} \) are i.i.d. Bernoulli random variables with parameter \( p = \frac{1}{k_0} = \frac{m}{n} \). The application of the Hoeffding inequality yields that \( \frac{m}{n} \leq |\Omega_k| \leq \frac{2m}{n} \) holds with probability at least \( 1 - 2 \exp(-cm) \) for a universal constant \( c > 0 \). Then we can take the uniform bound to obtain

\[ P\left( \frac{m}{2} \leq |\Omega_k| \leq \frac{3m}{2} \right) \text{ for all } k \]

\[ \geq 1 - 2k_0 \exp(-cm) \]

\[ \geq \frac{1}{2} \]

where the last inequality is due to \( m = \frac{m}{k_0} \geq \log(k_0) \).

Our next goal is to show that the events (III.22) - (III.25) occur with high probability. We will first apply the matrix Bernstein inequality (III.7) to obtain the desired upper bounds for fixed \( k \), and then take the uniform bound analysis to complete the proof.

A. Proof of (III.22)

For any \( Z \in \mathbb{C}^{n_1 \times n_2} \), by the definition of \( \mathcal{A}_k^{*} A_k \) in (III.17), we have

\[ E[\mathcal{A}_k^{*} A_k] G^* P_T(Z) = E \left[ \sum_{i \in \Omega_k} \langle b_i, e_i^T, G^* P_T(Z) \rangle b_i e_i^T \right] \]

\[ = \sum_{i \in \Omega_k} E[b_i b_i^T] G^* P_T(Z) e_i e_i^T \]

\[ = G^* P_T(Z) \sum_{i \in \Omega_k} e_i e_i^T, \]

where the third line follows from the isotropy property (II.5) of \( \{b_i\} \).

As a result, one has the following equality

\[ \left\| P_T G \left( I - \frac{1}{p} E[\mathcal{A}_k^{*} A_k] \right) G^* P_T \right\| = \sup_{\|W\|_F = 1} \left\| P_T G \left( I - \frac{1}{p} E[\mathcal{A}_k^{*} A_k] \right) G^* P_T(W) \right\|_F \]

\[ = \sup_{\|W\|_F = 1} \left\| \frac{1}{p} P_T G G^* P_T(W) - P_T G G^* P_T(W) \right\|_F \]

\[ = \sup_{\|W\|_F = 1} \left\| \sum_{i = 0}^{n-1} \left( \frac{1}{p} - 1 \right) P_T G (G^* P_T(W)e_i e_i^T) \right\|_F \]

\[ = \left\| \sum_{i = 0}^{n-1} \left( \frac{1}{p} - 1 \right) \chi_i \right\|, \]

where \( \delta_i \) is the Bernoulli random variable defined in (IV.1) and \( \chi_i \) is the operator defined as

\[ \chi_i(W) = P_T G (G^* P_T(W)e_i e_i^T) \]

for any \( W \in \mathbb{C}^{n_1 \times n_2} \). It is easy to verify that \( \chi_i \) is self-adjoint and positive semi-definite.

In order to apply the matrix Bernstein inequality (III.7) to bound \( \left\| \sum_{i = 0}^{n-1} \left( \frac{1}{p} - 1 \right) \chi_i \right\| \), one needs to bound \( \left\| \left( \frac{1}{p} - 1 \right) \chi_i \right\| \) and \( E \left[ \sum_{i = 0}^{n-1} \left( \frac{1}{p} - 1 \right)^2 \chi_i^2 \right] \).

For the upper bound of \( \left\| \left( \frac{1}{p} - 1 \right) \chi_i \right\| \), a simple calculation yields that

\[ \left\| \left( \frac{1}{p} - 1 \right) \chi_i \right\| \leq \frac{1}{p} \left\| \chi_i \right\| \]

\[ \leq \frac{1}{p} \left\| \chi_i \right\| \]

\[ = \frac{1}{p} \sup_{\|W\|_F = 1} \left\| P_T G (G^* P_T(W)e_i e_i^T) \right\|_F \]

\[ \leq \frac{1}{p} \sup_{\|W\|_F = 1} \left\| W \right\|_F 2 \mu_1 r \]

\[ \leq \frac{2 \mu_1 r}{np}, \]

where the third line follows from Corollary VI.5.
To bound \( \| \mathbb{E} \left[ \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right)^2 \mathcal{X}_i^2 \right] \| \), we have

\[
\left\| \sum_{i=0}^{n-1} \mathbb{E} \left[ \left( \frac{\delta_i}{p} - 1 \right)^2 \mathcal{X}_i^2 \right] \right\| \\
\leq \frac{1}{p} \left\| \sum_{i=0}^{n-1} \mathcal{X}_i^2 \right\|
\]

\[
\leq \frac{1}{p} \max_{0 \leq i \leq n-1} \| \mathcal{X}_i \| \cdot \left\| \sum_{i=0}^{n-1} \mathcal{X}_i \right\|
\]

\[
\leq \frac{2 \mu_1 r}{np} \sup_{\| W \|_F = 1} \left\| \sum_{i=0}^{n-1} \mathcal{X}_i(W) \right\|_F
\]

\[
= \frac{2 \mu_1 r}{np} \sup_{\| W \|_F = 1} \left\| \sum_{i=0}^{n-1} \mathcal{P}_T \mathcal{G} (\mathcal{G}^* \mathcal{P}_T(W)e_i e_i^T) \right\|_F
\]

\[
= \frac{2 \mu_1 r}{np} \sup_{\| W \|_F = 1} \left\| \mathcal{P}_T \mathcal{G} \mathcal{G}^* \mathcal{P}_T \right\|_F
\]

\[
\leq \frac{2 \mu_1 r}{np}
\]

where the second line is due to the positive semi-definite property of \( \mathcal{X}_i \), the third line follows from (IV.2), and the last line follows from the fact that \( \| \mathcal{G} \| = 1, \| \mathcal{G}^* \| \leq 1 \) and \( \mathcal{P}_T \) is the projection operator.

Applying the matrix Bernstein inequality (III.7) and taking \( a = sn \), \( c_2 \geq 3 \) imply that

\[
\mathbb{E} \left[ \sum_{i=0}^{n-1} \mathcal{X}_i A_i \mathcal{A}_i^* \mathcal{A}_i^* \right]
\]

\[
\leq \frac{\mu_1 r \log(sn)}{np} + \frac{\mu_1 r \log(sn)}{np}
\]

\[
\leq \frac{1}{4}
\]

holds with probability at least \( 1 - (sn)^{-c_2} \), where the second and third lines are due to \( p \geq \frac{\mu_1 r \log(sn)}{n} \). Finally, we take the uniform bound to obtain that

\[
P \left\{ \max_{1 \leq k \leq k_0} \left\| \mathcal{P}_T \mathcal{G} \left( I - \frac{n}{m} \mathbb{E} [A_i A_i^*] \right) \mathcal{G}^* \mathcal{P}_T \right\| \leq \frac{1}{4} \right\}
\]

\[
\geq 1 - k_0(sn)^{-c_2} \geq 1 - (sn)^{-c_2 - 2},
\]

where the last inequality follows from the fact that \( k_0 \ll sn \).

**B. Proof of (I.1)**

Following the definition of \( A_i A_i^* \) in (II.17) and the isotropy property of \( \{ b_i \} \) in (II.5), we have

\[
\left\| \mathcal{G} \left( I - \frac{1}{p} \mathbb{E} [A_i A_i^*] \right) \mathcal{G}^* \right\|
\]

\[
= \left\| \frac{1}{p} \mathcal{G} \mathcal{G}^*(Z) \sum_{i \in \Omega_k} e_i e_i^T - \mathcal{G} \mathcal{G}^*(Z) \right\|
\]

\[
= \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) \right\|
\]

\[
\leq \left\| \sum_{i=0}^{n-1} X_i \right\|
\]

where \( \delta_i \) is defined in (IV.1) and

\[
X_i := \left( \frac{\delta_i}{p} - 1 \right) \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) \in \mathbb{C}^{sn_1 \times n_2}
\]

are independent random matrices with zero mean.

Firstly, \( \| X_i \| \) can be bounded as follows:

\[
\| X_i \| \leq \frac{1}{p} \| \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) \|
\]

\[
= \frac{1}{p} \| \mathcal{G}_i \| \cdot \| \mathcal{G}^*(Z)e_i \|_2
\]

\[
\leq \frac{1}{p} \| \mathcal{G}_i \| \cdot \| \mathcal{G}^*(Z)e_i \|_2
\]

\[
\leq \frac{1}{p} \| \mathcal{G}^*(Z)e_i \|_2
\]

where the second line is due to (II.3), the third line follows from the fact that \( \| A \otimes B \| \leq \| A \| \cdot \| B \| \), and the last line directly follows from the definition of \( \| \cdot \|_{G, \infty} \) in (III.20).

Secondly, we have

\[
\mathbb{E} \left[ \sum_{i=0}^{n-1} X_i X_i^* \right]
\]

\[
= \sum_{i=0}^{n-1} \mathbb{E} \left[ \left( \frac{\delta_i}{p} - 1 \right)^2 \right]
\]

\[
= \sum_{i=0}^{n-1} \left( \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) \right) (\mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T))^* \left( \mathcal{G}_i \otimes (\mathcal{G}^*(Z)e_i) \right) \left( \mathcal{G}_i \otimes (\mathcal{G}^*(Z)e_i)^* \right)
\]

\[
\leq \frac{1}{p} \sum_{i=0}^{n-1} \left\| (\mathcal{G}_i \otimes (\mathcal{G}^*(Z)e_i)) \left( \mathcal{G}_i \otimes (\mathcal{G}^*(Z)e_i)^* \right) \right\|
\]

\[
\leq \frac{1}{p} \sum_{i=0}^{n-1} \left\| (\mathcal{G}_i \mathcal{G}_i^*) \otimes (\mathcal{G}^*(Z)e_i) (\mathcal{G}^*(Z)e_i)^* \right\|
\]

\[
\leq \frac{1}{p} \sum_{i=0}^{n-1} \left\| \mathcal{G}_i \mathcal{G}_i^* \right\| \cdot \| \mathcal{G}^*(Z)e_i \|_2^2
\]

\[
\leq \frac{1}{p} \sum_{i=0}^{n-1} \frac{1}{w_i} \| \mathcal{G}^*(Z)e_i \|_2^2
\]

\[
= \frac{1}{p} \| Z \|_{G,F}^2,
\]

Since \( \mathbb{E} \left[ \sum_{i=0}^{n-1} X_i X_i^* \right] \) can be bounded by the same quantity, using the matrix Bernstein inequality (III.7) and
taking \( a = s_n, c_2 \geq 3 \) imply that
\[
\|G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \|_{\mathcal{G}, F}^2 = \left\| \sum_{i=0}^{n-1} X_i \right\|_{\mathcal{G}, F}^2 \leq \left( \frac{1}{p} \log(s_n) + \frac{\log(s_n)}{p} \right) \|Z\|_{\mathcal{G}, \infty},
\]
holds with probability at least \( 1 - (s_n)^{-c_3} \).

By the uniform bound we conclude that the event (III.23) occurs with probability at least \( 1 - (s_n)^{-c_3} \).

### C. Proof of (III.24)

By the definition of \( \|\cdot\|_{\mathcal{G}, F} \) in (III.19) and the isotropy property of \( \{b_i\} \) in (II.5), it follows that
\[
\left\| \mathcal{P}_T G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \right\|_{\mathcal{G}, F}^2 = \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) \mathcal{P}_T G (G^*(Z)e_i e_i^T) \right\|_{\mathcal{G}, F}^2 = \sum_{j=0}^{n-1} \frac{1}{w_j} \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) G^* \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_j \right\|_{2}^2.
\]
If we construct a new vector \( z_i \in \mathbb{C}^{s \times 1} \) as
\[
z_i := \left( \frac{\delta_i}{p} - 1 \right) \left[ \begin{array}{c}
\frac{1}{\sqrt{w_i}} G^* \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_0 \\
\vdots \\
\frac{1}{\sqrt{w_i}} G^* \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_{n-1}
\end{array} \right],
\]
then it can be easily seen that
\[
\left\| \mathcal{P}_T G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \right\|_{\mathcal{G}, F}^2 =: \left\| \sum_{i=0}^{n-1} z_i \right\|_{2}^2.
\]
For the upper bound of \( \|z_i\|_2 \), a direct calculation yields that
\[
\|z_i\|_2 \leq \frac{1}{p} \left\| \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) \right\|_{\mathcal{G}, F} \leq \frac{1}{p} \left\| G^*(Z)e_i \right\|_2 \leq \frac{1}{p} \sqrt{\frac{\mu_1 r \log(s_n)}{n}} ||G^*(Z)e_i||_2 \leq \frac{1}{p} \sqrt{\frac{\mu_1 r \log(s_n)}{n}} ||Z||_{\mathcal{G}, \infty},
\]
where the third line follows from Lemma VI.9 and the last line is due to the definition of \( \|\cdot\|_{\mathcal{G}, \infty} \) in (III.20).

In addition,
\[
\mathbb{E} \left\| \sum_{i=0}^{n-1} z_i e_j^* \right\|_2 \leq \frac{1}{p} \left\| \mathcal{P}_T G \left( (G^*(Z)e_i e_i^T) \right) \right\|_{\mathcal{G}, F} \leq \frac{1}{p} \mu_1 r \log(s_n) ||Z||_{\mathcal{G}, \infty} \leq \frac{1}{p} \mu_1 r \log(s_n) ||Z||_{\mathcal{G}, \infty},
\]
where the third inequality is due to Lemma VI.9, and the same bound can be obtained for \( \mathbb{E} \left[ \sum_{i=0}^{n-1} z_i^* z_i \right] \).

Therefore, by the matrix Bernstein inequality (III.7) and taking \( a = s_n, c_2 \geq 3 \), we can show that
\[
\left\| \sum_{i=0}^{n-1} z_i \right\|_2 \leq \sqrt{\frac{\mu_1 r \log(s_n)}{n}} \left( \frac{\log(s_n)}{p} \right) \|Z\|_{\mathcal{G}, \infty},
\]
holds with probability at least \( 1 - (s_n)^{-c_3} \). Taking the uniform bound completes the proof.

### D. Proof of (III.25)

The definition of \( \|\cdot\|_{\mathcal{G}, \infty} \) in (III.20) allows us to express
\[
\left\| \mathcal{P}_T G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \right\|_{\mathcal{G}, \infty} = \left\| \mathcal{P}_T G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \right\|_{\mathcal{G}, \infty} = \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) \right\|_{\mathcal{G}, \infty} = \max_{0 \leq j \leq n-1} \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_j \right\|_{\mathcal{G}, \infty}.
\]
Define \( z^*_j \) to be the \( s \)-dimensional vector
\[
z^*_j := \left( \frac{\delta_j}{p} - 1 \right) \frac{1}{\sqrt{w_j}} G^* \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_j
\]
for \((i, j) \in [n] \times [n]\). Then one can easily see that
\[
\left\| \mathcal{P}_T G \left( I - \frac{1}{p} \mathbf{E} [A_k^* A_k] \right) G^*(Z) \right\|_{\mathcal{G}, \infty} \leq \left\| \sum_{i=0}^{n-1} \left( \frac{\delta_i}{p} - 1 \right) \mathcal{P}_T G \left( G^*(Z)e_i e_i^T \right) e_j \right\|_{\mathcal{G}, \infty}.
\]
For any fixed $j \in [n]$, $\|z_i^j\|_2^2$ can be bounded as follows:

$$
\|z_i^j\|_2^2 \leq \frac{1}{p} \sum_{i=0}^{n-1} \left\| \frac{1}{\sqrt{w_i}} \left( \mathcal{G}^* \mathcal{P} \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) e_j \right) \right\|_2^2
$$

$$
= \frac{1}{p} \sum_{i=0}^{n-1} \sup_{\|\beta\|_2=1} \left\| \mathcal{G}^* \mathcal{P} \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) e_j, \beta \right\|_2
$$

$$
= \frac{1}{p} \sum_{i=0}^{n-1} \mathbb{E} \left[ \mathcal{G}^* (Z) e_i e_i^T \right] \|z_i^j\|_2^2
$$

$$
\leq \frac{3 \mu r}{p n} \frac{1}{n} \mathbb{E} \left[ \mathcal{G}^* (Z) e_i e_i^T \right] \|z_i^j\|_2^2
$$

$$
\leq \frac{3 \mu r}{p n} \mathbb{E} \left[ \mathcal{G}^* (Z) e_i e_i^T \right] \|Z\|_{g, \infty}, \quad \text{(IV.3)}
$$

where the fourth line follows from Lemma VI.6 and the last line is due to the definition of $\|\cdot\|_{g, \infty}$ in (III.20).

Moreover, we have

$$
\mathbb{E} \left[ \sum_{i=0}^{n-1} \left( z_i^j (z_i^j)^* \right) \right]
$$

$$
\leq \sum_{i=0}^{n-1} \mathbb{E} \left[ \|z_i^j\|_2^2 \right]
$$

$$
\leq \frac{1}{p} \sum_{i=0}^{n-1} \left\| \frac{1}{\sqrt{w_i}} \mathcal{G}^* \mathcal{P} \mathcal{G} (\mathcal{G}^*(Z)e_i e_i^T) e_j \right\|_2^2
$$

$$
\leq \frac{1}{p} \left( \frac{3 \mu r}{n} \right)^2 \sum_{i=0}^{n-1} \left( \frac{\|\mathcal{G}^*(Z) e_i e_i^T\|_2}{\sqrt{w_i}} \right)^2
$$

$$
= \frac{1}{p} \left( \frac{3 \mu r}{n} \right)^2 \cdot \frac{1}{2} \|Z\|_{g, F}^2,
$$

where the third inequality follows from (IV.3). The same bound can be obtained for $\sum_{i=0}^{n-1} \mathbb{E} \left[ (z_i^j)^* z_i^j \right]$ as well.

If we take $a = sn$, $c_2 \geq 3$, the matrix Bernstein inequality (III.7) taken collectively with the uniform bound yields that

$$
\mathcal{P} \mathcal{G} \left( \mathcal{I} - \frac{1}{p} \mathbb{E} \left[ A_k^\alpha A_k \right] \mathcal{G}^*(Z) \right) \|_{g, \infty}
$$

$$
= \sup \left\{ \| \mathcal{P} \mathcal{G} (b_i e_i^T) \|_2 \right\}
$$

$$
= \sup \left\{ \mathbb{E} \left[ (b_i e_i^T)^* (b_i e_i^T)^* \right] \right\}
$$

$$
= \sup \left\{ \mathbb{E} \left[ (b_i e_i^T)^* (b_i e_i^T)^* \right] \right\}
$$

$$
\leq \frac{3 \mu r}{p n} \mathbb{E} \left[ \mathcal{G}^* (Z) e_i e_i^T \right] \|Z\|_{g, \infty},
$$

$$
\text{where it is obvious that } z_i z_i^* \text{ are independent and positive semi-definite random matrices. Hence,}
$$

$$
\mathcal{P} \mathcal{G} \left( A_k^\alpha A_k - \mathbb{E} \left[ A_k^\alpha A_k \right] \right) \|_{g, \infty}
$$

$$
= \mathbb{E} \left[ (z_i^j)^* (z_i^j)^* \right] \|Z\|_{g, \infty},
$$

$$
\text{Firstly, } \|z_i z_i^* - \mathbb{E} \left[ z_i z_i^* \right] \|_2 \text{ can be bounded as follows:}
$$

$$
\|z_i z_i^* - \mathbb{E} \left[ z_i z_i^* \right] \|_2 \leq \max \left\{ \|z_i z_i^*\|_2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]\right\}
$$

$$
\leq \max \left\{ \|z_i z_i^*\|_2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]\right\}
$$

$$
\leq \max \left\{ \|z_i z_i^*\|_2^2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]^2\right\},
$$

$$
\text{This section presents the proofs of Lemmas III.10 to III.15, which have been used to verify (III.10) and (III.11).}
$$

A. Proof of Lemma III.10

Note that

$$
\| \mathcal{P} \mathcal{G} (I - \frac{1}{m} A_k^\alpha A_k) \mathcal{G}^* \mathcal{P} \mathcal{G} \|_F
$$

$$
\leq \| \mathcal{P} \mathcal{G} (I - \frac{1}{m} A_k^\alpha A_k) \mathcal{G}^* \mathcal{P} \mathcal{G} \|_F
$$

$$
+ \frac{1}{m} \| \mathcal{P} \mathcal{G} (A_k^\alpha A_k - \mathbb{E} \left[ A_k^\alpha A_k \right]) \mathcal{G}^* \mathcal{P} \mathcal{G} \|_F.
$$

According to (III.22) in Lemma III.9, the first term is upper bounded by $\frac{1}{4}$. We will bound the second term via the matrix Bernstein inequality (III.7).

For any $Z \in \mathbb{C}^{m_1 \times n_2}$, by the definition of $A_k^\alpha A_k$ in (III.17), we have

$$
\mathcal{P} \mathcal{G} A_k^\alpha A_k \mathcal{G}^* \mathcal{P} \mathcal{G} (Z)
$$

$$
= \sum_{i \in \Omega_k} \langle b_i e_i^T, \mathcal{G}^* \mathcal{P} \mathcal{G} (Z) \rangle b_i e_i^T
$$

$$
= \sum_{i \in \Omega_k} \langle \mathcal{P} \mathcal{G} (b_i e_i^T), Z \rangle \mathcal{P} \mathcal{G} (b_i e_i^T).
$$

If we define $z_i := \text{vec}(\mathcal{P} \mathcal{G} (b_i e_i^T)) \in \mathbb{C}^{m_1 n_2 \times 1}$, then it follows that

$$
\mathcal{P} \mathcal{G} A_k^\alpha A_k \mathcal{G}^* \mathcal{P} \mathcal{G} (Z)
$$

$$
= \sup \left\{ \| \mathcal{P} \mathcal{G} A_k^\alpha A_k \mathcal{G}^* \mathcal{P} \mathcal{G} (W) \|_F \right\}
$$

$$
= \sup \left\{ \| \mathcal{P} \mathcal{G} (b_i e_i^T), W \rangle \mathcal{P} \mathcal{G} (b_i e_i^T) \|_F \right\}
$$

$$
= \sup \left\{ \| \text{vec}(W) \|_2 \right\}
$$

$$
= \| \text{vec}(W) \|_2,
$$

$$
\text{where it is obvious that } z_i z_i^* \text{ are independent and positive semi-definite random matrices. Hence,}
$$

$$
\mathcal{P} \mathcal{G} (A_k^\alpha A_k - \mathbb{E} \left[ A_k^\alpha A_k \right]) \mathcal{G}^* \mathcal{P} \mathcal{G} (Z)
$$

$$
= \left\| (z_i^j)^* (z_i^j)^* \right\|_2.
$$

$$
\text{Firstly, } \|z_i z_i^* - \mathbb{E} \left[ z_i z_i^* \right] \|_2 \text{ can be bounded as follows:}
$$

$$
\|z_i z_i^* - \mathbb{E} \left[ z_i z_i^* \right] \|_2 \leq \max \left\{ \|z_i z_i^*\|_2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]\right\}
$$

$$
\leq \max \left\{ \|z_i z_i^*\|_2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]\right\}
$$

$$
\leq \max \left\{ \|z_i z_i^*\|_2^2, \mathbb{E} \left[ \mathbb{E} \left[ z_i z_i^* \right] \right]^2\right\},
$$

holds with probability at least $1 - (sn)^{-(c_2 - 1)}$.
where the second line is due to the Jensen inequality. By the definition of $z_i$, we have $\|z_i\|_2^2 = \|\mathcal{P}_T G(b_i e_i^T)^*\|_F^2$. Then applying (VI.6) in Corollary VI.3 implies that

$$\|z_i z_i^* - \mathbb{E}[z_i z_i^*]\| \leq \max\{\|z_i\|_2^2, \mathbb{E}[|z_i|^2]\}\leq \frac{2\mu_1 r s \mu_0}{n}.$$ 

Secondly,

$$\left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i z_i^* - \mathbb{E}[z_i z_i^*])^2\right]\right\| = \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i z_i^*)^2\right] - \mathbb{E}\left[|z_i|^2\right]^2\right\| \leq \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i z_i^*)^2\right]\right\| \leq \max_{i \in \Omega_k} \|z_i z_i^*\| \cdot \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i z_i^*)^2\right]\right\| \leq \frac{2\mu_1 r s \mu_0}{n} \cdot \frac{5m}{4n}.$$ 

Here the last line follows from a direct calculation:

$$\left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i z_i^*)^2\right]\right\| = \sup_{\|\mathbf{w}\|_F = 1} \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[\mathbf{w}^T \mathcal{P}_T G(b_i e_i^T)^* \mathbf{w}\right]\right\|_2 = \sup_{\|\mathbf{w}\|_F = 1} \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[\langle \mathcal{P}_T G(b_i e_i^T), \mathbf{w} \rangle \mathcal{P}_T G(b_i e_i^T)^*\right]\right\|_2 = \sup_{\|\mathbf{w}\|_F = 1} \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[\langle \mathcal{P}_T G(b_i e_i^T), \mathbf{w} \rangle \mathcal{P}_T G(b_i e_i^T)^*\right]\right\|_F = \sup_{\|\mathbf{w}\|_F = 1} \left\|\sum_{i \in \Omega_k} \mathbb{E}\left[\langle \mathcal{P}_T G(b_i e_i^T), \mathbf{w} \rangle \mathcal{P}_T G(b_i e_i^T)^*\right]\right\|_F = \left\|\sum_{i \in \Omega_k} \mathbb{P}_T G\left((b_i^*)^* \mathcal{P}_T (g) e_i e_i^T\right)\right\| = \frac{5m}{4n}.$$ 

where in the last inequality we have utilized (III.22) in the following way,

$$\frac{1}{4} \geq \left\|\mathcal{P}_T G \left(1 - \frac{n}{m} \mathbb{E}[A_k^* A_k]\right) G^* \mathcal{P}_T\right\| \geq \frac{n}{m} \left\|\mathcal{P}_T G \left(A_k^* A_k\right) G^* \mathcal{P}_T - \mathcal{P}_T G G^* \mathcal{P}_T\right\|.$$ 

Since we can obtain the same bound for $\left\|\sum_{i \in \Omega_k} \mathbb{E}\left[(z_i^* z_i - \mathbb{E}[z_i z_i^*])^2\right]\right\|$, applying the matrix Bernstein inequality (III.7) and taking $a = sn$, $c_2 \geq 3$ imply that with probability at least $1 - (sn)^{-c_2}$,

$$\frac{n}{m} \left\|\mathcal{P}_T G \left(A_k^* A_k - \mathbb{E}[A_k^* A_k]\right) G^* \mathcal{P}_T\right\| \geq \frac{n}{m} \left\|\sum_{i \in \Omega_k} \left(z_i z_i^* - \mathbb{E}[z_i z_i^*]\right) \left(G^* \mathcal{P}_T\right)\right\| \leq \frac{n}{m} \left\|\sum_{i \in \Omega_k} \left(z_i z_i^* - \mathbb{E}[z_i z_i^*]\right)\right\| \leq \frac{n}{m} \left(\sqrt{\frac{5m}{4n}} \cdot \frac{2\mu_1 r s \mu_0}{n} \cdot \log(sn) + \frac{2\mu_1 r s \mu_0 \log(sn)}{n}\right) \leq \frac{1}{m} \left(\sqrt{\frac{5m}{4n}} \cdot \frac{2\mu_1 r s \mu_0 \log(sn)}{2} + \frac{2\mu_1 r s \mu_0 \log(sn)}{2}\right) \leq \frac{1}{4}.$$ 

Finally, combining the two terms together completes the proof.

B. Proof of Lemma III.12

Notice that

$$\mathbb{G}\left(G \left(1 - \frac{n}{m} E[A_k^* A_k]\right) G^* (Z)\right)\leq \mathbb{E}\left[G \left(1 - \frac{n}{m} E[A_k^* A_k]\right) G^* (Z)\right] + \frac{n}{m} \left\|G \left(A_k^* A_k - E[A_k^* A_k]\right) G^* (Z)\right\| \leq \sqrt{\frac{n \log(sn)}{m}} \|Z\|_{g, F} + \frac{n \log(sn)}{m} \|Z\|_{g, \infty} + \frac{n}{m} \left\|G \left(A_k^* A_k - E[A_k^* A_k]\right) G^* (Z)\right\|, \quad (V.1)$$

where the second line follows from (III.23). In order to prove (III.28), it suffices to bound the last term.

Recalling the definition of $A_k^* A_k$ in (III.17) and using the isotropy property of $\{b_i\}$ in (II.5), we can rewrite the last term as

$$\frac{n}{m} \left\|G \left(A_k^* A_k - E[A_k^* A_k]\right) G^* (Z)\right\| = \frac{n}{m} \left\|\sum_{i \in \Omega_k} G \left((b_i b_i^* - I) G^* (Z) e_i e_i^T\right)\right\| = n \left\|\sum_{i \in \Omega_k} G \left((b_i b_i^* - I) G^* (Z) e_i e_i^T\right)\right\| = n \left\|\sum_{i \in \Omega_k} G \left((b_i b_i^* - I) G^* (Z) e_i e_i^T\right)\right\| = n \left\|\sum_{i \in \Omega_k} G \left((b_i b_i^* - I) G^* (Z) e_i e_i^T\right)\right\|,$$

where $X_i = G \left((b_i b_i^* - I) G^* (Z) e_i e_i^T\right) \in \mathbb{C}^{sn_1 \times n_2}$. It can be easily seen that $X_i$ are independent random matrices with zero mean.
The upper bound of $\|X_i\|$ can be established as follows:

$$\|X_i\| = \|G((b_ib_i^* - I)G^*(Z)e_i)e_i^T)\|$$

$$= \|G_i \otimes ((b_ib_i^* - I)G^*(Z)e_i))\|$$

$$\leq \|G_i\| \cdot \|(b_ib_i^* - I)G^*(Z)e_i)\|$$

$$\leq \frac{1}{\sqrt{\nu_i}} \max\left\{\|b_i\|_2^2, 1\right\} \cdot \|G^*(Z)e_i\|_2$$

$$\leq s\mu_0 \|Z\|_{G,\infty},$$

where the second line follows from (II.3), the third line is due to $\|A \otimes B\| \leq \|A\| \cdot \|B\|$, and the last line follows from the definition of $\|\|_{G,\infty}$ in (III.20).

To bound $\|E \left[ \sum_{i \in \Omega_k} X_i^* X_i \right]\|$, we first define $z_i = (b_ib_i^* - I)G^*(Z)e_i \in \mathbb{C}^n$. Then a simple calculation yields that

$$\mathbb{E} \left[ \|z_i\|_2^2 \right]$$

$$= \mathbb{E} \left[ e_i^T (G^*(Z))^*(b_ib_i^* - I)^2G^*(Z)e_i \right]$$

$$= e_i^T (G^*(Z))^*(\mathbb{E} \left[ (b_ib_i^*)^2 - 2b_ib_i^* + I \right])G^*(Z)e_i$$

$$\leq e_i^T (G^*(Z))^*(s\mu_0 \mathbb{E} \left[ (b_ib_i^*)^2 - I \right])G^*(Z)e_i$$

$$\leq s\mu_0 \cdot \|G^*(Z)e_i\|_2^2,$$

(V.2)

where the last two inequalities follow from the incoherence property (II.6) and the isotropy property (II.5) of $\{b_i\}$. Furthermore, it follows that

$$\mathbb{E} \left[ \sum_{i \in \Omega_k} X_i^* X_i \right]$$

$$= \sum_{i \in \Omega_k} \mathbb{E} \left[ (G_i \otimes z_i)^*(G_i \otimes z_i) \right]$$

$$= \sum_{i \in \Omega_k} \mathbb{E} \left[ (G_i^T G_i) \otimes (z_i^* z_i) \right]$$

$$= \sum_{i \in \Omega_k} (G_i^T G_i) \mathbb{E} \left[ \|z_i\|_2^2 \right]$$

$$\leq s\mu_0 \cdot \left\| \sum_{i \in \Omega_k} \|G^*(Z)e_i\|_2^2 \right\| \cdot \|G_i^T G_i\|$$

$$\leq s\mu_0 \cdot \left\| \sum_{i \in \Omega_k} \|G^*(Z)e_i\|_2^2 \right\| \cdot \|G_i^T G_i\|$$

$$\leq s\mu_0 \cdot \left\| \sum_{i \in \Omega_k} \|G^*(Z)e_i\|_2^2 \right\| \cdot \|G_i^T G_i\|$$

$$\leq s\mu_0 \cdot \left\| \sum_{i \in \Omega_k} \|G^*(Z)e_i\|_2^2 \right\| \cdot \|G_i^T G_i\|$$

$$\leq s\mu_0 \cdot \left\| \sum_{i \in \Omega_k} \|G^*(Z)e_i\|_2^2 \right\| \cdot \|G_i^T G_i\|$$

$$= s\mu_0 \cdot \|Z\|_{G,F},$$

where the fourth line follows from (V.2), and $\mathbb{E} \left[ \sum_{i \in \Omega_k} X_i X_i^* \right]$ can be similarly bounded.

Therefore, by the matrix Bernstein inequality (III.16) and taking $a = s\mu$, $c_3 \geq 2$,

$$\frac{n}{m} \|G((A_k^* A_k - \mathbb{E} [A_k^* A_k])G^*(Z))\|$$

$$= \frac{n}{m} \left\| \sum_{i \in \Omega_k} X_i \right\|$$

$$\leq \frac{n}{m} \left\{ \sqrt{s\mu_0 \log\left(\frac{sn}{m}\right)} \|Z\|_{G,F} + s\mu_0 \log\left(\frac{sn}{m}\right) \|Z\|_{G,\infty} \right\}$$

holds with probability at least $1 - (sn)^{-(c_3 - 1)}$. Inserting this bound into (V.1), we conclude that

$$\|G((I - \frac{n}{m} A_k^* A_k)G^*(Z))\|$$

$$\leq \frac{n}{m} \left\{ \sqrt{k \mu_0 \log\left(\frac{sn}{m}\right)} + \frac{n}{m} \log\left(\frac{sn}{m}\right) \right\} \|Z\|_{G,F}$$

$$+ \frac{n}{m} \left\{ \left(2n \mu_0 \log\left(\frac{sn}{m}\right) + \frac{n}{m} \log\left(\frac{sn}{m}\right) \right) \right\} \|Z\|_{G,\infty}$$

holds with probability exceeding $1 - (sn)^{-(c_3 - 1)}$.

C. Proof of Lemma III.13

Notice that

$$\|P_T G((I - \frac{n}{m} A_k^* A_k)G^*(Z))\|_{G,F}$$

$$\leq \|P_T G((I - \frac{n}{m} \mathbb{E} [A_k^* A_k])G^*(Z))\|_{G,F}$$

$$+ \frac{n}{m} \|P_T G((A_k^* A_k - \mathbb{E} [A_k^* A_k])G^*(Z))\|_{G,F}$$

$$\leq \frac{n}{m} \left\{ \sqrt{\mu_1 r \log\left(\frac{sn}{m}\right)} \right\} \|Z\|_{G,F} + \frac{n}{m} \log\left(\frac{sn}{m}\right) \|Z\|_{G,\infty}$$

$$+ \frac{n}{m} \|P_T G((A_k^* A_k - \mathbb{E} [A_k^* A_k])G^*(Z))\|_{G,F},$$

where the second line follows from (III.24). We will adopt the matrix Bernstein inequality (III.7) to bound the second term.

Recalling the definition of $A_k^* A_k$ in (III.17) and letting $z_i := (b_i b_i^* - I)G^*(Z)e_i \in \mathbb{C}^n$, we have

$$\frac{n}{m} \|P_T G((A_k^* A_k - \mathbb{E} [A_k^* A_k])G^*(Z))\|_{G,F}$$

$$= \frac{n}{m} \left\| P_T G((b_i b_i^* - \mathbb{E} [b_i b_i^*])G^*(Z)e_i e_i^T) \right\|_{G,F}$$

$$= \frac{n}{m} \left\| \sum_{i \in \Omega_k} P_T G((b_i b_i^* - I)G^*(Z)e_i e_i^T) \right\|_{G,F}$$

$$= \frac{n}{m} \left\| \sum_{i \in \Omega_k} P_T G(z_i e_i^T) \right\|_{G,F}.$$
where the second equality is due to the isotropy property of \(\{b_i\} \) in (II.5). Furthermore, denoting by \(y_i \in \mathbb{C}^{sn \times 1} \) the vector

\[
y_i := \begin{bmatrix}
\sqrt{w_i} \mathcal{G}^* \mathcal{P}_T \mathcal{G}(z_i e_i^T) e_0 \\
\vdots \\
\sqrt{w_i} \mathcal{G}^* \mathcal{P}_T \mathcal{G}(z_i e_i^T) e_l \\
\vdots \\
\sqrt{w_{n-1}} \mathcal{G}^* \mathcal{P}_T \mathcal{G}(z_i e_i^T) e_{n-1}
\end{bmatrix},
\]

the second term can be expressed as

\[
\frac{n}{m} \| \mathcal{P}_T \mathcal{G} (A_k^* A_k - \mathbb{E}[A_k^* A_k]) \mathcal{G}^*(Z) \|_{\mathcal{G}, \mathcal{F}}^2 \\
= \frac{n}{m} \left\| \sum_{i \in \Omega_k} y_i \right\|_2^2.
\]

Clearly, \(y_i\) are independent random vectors with zero mean.

A direct calculation yields that

\[
\left\| y_i \right\|_2^2 \\
= \frac{1}{m} \sum_{j=0}^{n-1} \left\| \mathcal{G}^* \mathcal{P}_T \mathcal{G}(z_i e_i^T) e_j \right\|_2^2 \\
= \left\| \mathcal{P}_T \mathcal{G}(z_i e_i^T) \right\|_{\mathcal{G}, \mathcal{F}}^2 \\
= \frac{1}{\sqrt{w_i}} \left\| \mathcal{P}_T \mathcal{G}(\sqrt{w_i} z_i e_i^T) \right\|_{\mathcal{G}, \mathcal{F}} \\
\lesssim \frac{1}{\sqrt{w_i}} \left\| \mathcal{G}^*(\mathbf{b}_i^{*} - I) \right\|_{\mathcal{G}, \mathcal{F}} \|e_i\|_2 \\
\leq \frac{1}{\sqrt{w_i}} \sqrt{\mu_1 r \log(sn)} \cdot s \mu_0 \cdot \|Z\|_{\mathcal{G}, \infty},
\]

where the fourth line follows from Lemma VI.9 and the last line is due to the definition of \(\|\cdot\|_{\mathcal{G}, \infty}\) in (III.20).

Additionally, we have

\[
\mathbb{E} \left[ \sum_{i \in \Omega_k} y_i^* y_i \right] \\
\leq \sum_{i \in \Omega_k} \mathbb{E} \left[ \left\| y_i \right\|_2^2 \right] \\
= \sum_{i \in \Omega_k} \mathbb{E} \left[ \left\| \mathcal{P}_T \mathcal{G}(z_i e_i^T) \right\|_{\mathcal{G}, \mathcal{F}}^2 \right] \\
\lesssim \sum_{i \in \Omega_k} \frac{\mu_1 r \log(sn)}{n} \cdot \mathbb{E} \left[ \|z_i\|_2^2 \right] \\
\lesssim s \mu_0 \frac{\mu_1 r \log(sn)}{n} \cdot \sum_{i \in \Omega_k} \frac{1}{w_i} \|\mathcal{G}^*(Z)e_i\|_2^2 \\
\lesssim s \mu_0 \frac{\mu_1 r \log(sn)}{n} \cdot \|Z\|_{\mathcal{G}, \mathcal{F}}^2,
\]

where the third line is due to Lemma VI.9 and the fourth line follows from

\[
\mathbb{E} \left[ \|z_i\|_2^2 \right] \\
= \mathbb{E} \left[ \| (\mathbf{b}_i^{*} - I) \mathcal{G}^*(Z)e_i \|_2^2 \right] \\
= \mathbb{E} \left[ \|e_i^T (\mathcal{G}^*(Z))^{*} (\mathbf{b}_i^{*} - I)^2 \mathcal{G}^*(Z)e_i \|_2^2 \right] \\
= e_i^T (\mathcal{G}^*(Z))^{*} \left( \mathbb{E} \left[ (\mathbf{b}_i^{*} - I)^2 \mathcal{G}^*(Z)e_i \right] \right) \mathcal{G}^*(Z)e_i \\
\leq s \mu_0 \|\mathcal{G}^*(Z)e_i\|_2^2.
\]

The same upper bound can be obtained for

\[
\mathbb{E} \left[ \sum_{i \in \Omega_k} y_i^* y_i \right].
\]

Applying the matrix Bernstein inequality (III.7) and taking

\(a = sn, \ c_3 \geq 2 \) yield that

\[
\frac{1}{n} \left[ \sum_{i \in \Omega_k} y_i \right] \right) \right) \\
\lesssim \frac{1}{n} \left( \sqrt{\frac{\mu_1 r \log(sn)}{n} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \mathcal{F}}^2 \right]} \\
+ \sqrt{\frac{\mu_1 r \log(sn)}{n} \cdot s \mu_0 \log(sn) \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \infty} \right]} \\
+ \frac{n \log(sn)}{m} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \mathcal{F}}^2 \right] \\
+ \frac{n \log(sn)}{m} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \infty} \right] \\
+ \frac{ns \mu_0 \log(sn)}{m} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \mathcal{F}}^2 \right] \\
+ \frac{ns \mu_0 \log(sn)}{m} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \infty} \right] \\
\right)
\]

holds with probability at least \(1 - (sn)^{-c_3 - 1}\). Noting (V.3) and (V.4), it follows immediately that

\[
\left\| \mathcal{P}_T \mathcal{G} (I - \frac{n}{m} A_k^* A_k) \mathcal{G}^*(Z) \right\|_{\mathcal{G}, \mathcal{F}} \\
\lesssim \sqrt{\frac{\mu_1 r \log(sn)}{n} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \mathcal{F}}^2 \right]} \\
+ \sqrt{\frac{\mu_1 r \log(sn)}{n} \cdot \mathbb{E} \left[ \|Z\|_{\mathcal{G}, \infty} \right]} \\
\right)
\]

holds with probability greater than \(1 - (sn)^{-c_3 - 1}\).
D. Proof of Lemma III.14

By the triangle inequality, we have

\[
\begin{align*}
\left\| \mathcal{P}_T \mathcal{G} \left( \mathcal{I} - \frac{n}{m} A_k^* A_k \right) G^*(Z) \right\|_{G, \infty}^2 & \leq \left\| \mathcal{P}_T \mathcal{G} \left( \mathcal{I} - \frac{n}{m} [A_k^* A_k] \right) G^*(Z) \right\|_{G, \infty}^2 \\
& \quad + \frac{n}{m} \left\| \mathcal{P}_T \mathcal{G} \left( A_k^* A_k - \mathbb{E} [A_k^* A_k] \right) G^*(Z) \right\|_{G, \infty}^2 \\
& \quad + \frac{n}{m} \left( \frac{\| G^* \mathcal{P}_T \mathcal{G} (z_i e_i^T) \|_2}{\sqrt{m}} \right)^2 \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty}^2.
\end{align*}
\]

(III.6)

where the second line is due to (III.25). In the following proof, we will bound the second term by the matrix Bernstein inequality (III.7) and the uniform bound argument.

If we define \( z_i = (b_i, b_i^* - I) G^*(Z) e_i \in \mathbb{C}^n \) and \( y_i^T = \frac{1}{\sqrt{w_i}} G^* \mathcal{P}_T \mathcal{G} (z_i e_i^T) e_i \), the second term can be rewritten as

\[
\begin{align*}
\frac{n}{m} \left\| \mathcal{P}_T \mathcal{G} \left( A_k^* A_k - \mathbb{E} [A_k^* A_k] \right) G^*(Z) \right\|_{G, \infty} & \leq \frac{n}{m} \left\| \sum_{i \in \Omega_k} \mathcal{P}_T \mathcal{G} ( (b_i, b_i^* - I) G^*(Z) e_i e_i^T ) \right\|_{G, \infty} \\
& = \frac{n}{m} \left\| \sum_{i \in \Omega_k} \mathcal{P}_T \mathcal{G} (z_i e_i^T) \right\|_{G, \infty} \\
& = \frac{n}{m} \left\| \sum_{i \in \Omega_k} \mathcal{P}_T \mathcal{G} (z_i e_i^T) \right\|_{G, \infty} \\
& = \frac{n}{m} \sup_{0 \leq j \leq n-1} \left\| \sum_{i \in \Omega_k} G^* \left( \mathcal{P}_T \mathcal{G} (z_i e_i^T) \right) e_j \right\|_2 \\
& \leq \frac{n}{m} \sup_{0 \leq j \leq n-1} \left\| \sum_{i \in \Omega_k} y_i^T e_j \right\|_2.
\end{align*}
\]

(III.7)

where the first equation follows from (III.17) and the isotropy property of \( \{b_i\} \) in (II.5).

For any fixed \( j \in [n] \), \( \left\| y_i^T \right\|_2 \) can be bounded as follows:

\[
\begin{align*}
\left\| y_i^T \right\|_2 & = \frac{1}{\sqrt{w_i}} \left\| G^* \mathcal{P}_T \mathcal{G} (z_i e_i^T) e_j \right\|_2 \\
& = \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} \left| \langle G^* \mathcal{P}_T \mathcal{G} (z_i e_i^T) e_j, \beta \rangle \right| \\
& \leq \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} \left\| G^* \mathcal{P}_T \mathcal{G} (z_i e_i^T), \mathcal{G} (\beta e_j^T) \right\| \\
& \leq \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} \left\| (b_i, b_i^* - I) G^*(Z) e_i \right\|_2 \\
& \leq \frac{1}{\sqrt{w_i}} \left\| (b_i, b_i^* - I) \cdot \| G^*(Z) e_i \|_2 \right\| \leq \mu_1 r \frac{n}{m} \| Z \|_{G, \infty}.
\end{align*}
\]

(III.8)

where the fourth line follows from Lemma VI.6 and the last line is due to the incoherence property of \( \{b_i\} \) in (II.6) and the definition of \( \| \cdot \|_{G, \infty} \) in (III.20).

Moreover,

\[
\begin{align*}
E \left[ \sum_{i \in \Omega_k} y_i^T (y_i^T)^T \right] & \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty} \\
& \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty} \\
& \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty}.
\end{align*}
\]

(III.9)

where the second line is due to (V.8) and the third line follows from (V.5). It also holds that

\[
\begin{align*}
E \left[ \sum_{i \in \Omega_k} (y_i^T)^* y_i \right] & \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty} \\
& \leq \frac{\mu_1 r}{n} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G,F} \left( \frac{n \log(sn)}{m} \right) \| Z \|_{G, \infty}.
\end{align*}
\]

(III.10)

where the second line is due to (V.8) and the third line follows from (V.5).

E. Proof of Lemma III.15

According to (III.4), a simple algebra yields that

\[
\max_{0 \leq i \leq n-1} \| U_i V^* \|_F^2 \leq \max_{0 \leq i \leq n-1} \| U_i \|_F^2 \leq \frac{\mu_1 r}{n}.
\]
Then the application of Corollary VI.8 implies that
\[ \| UV^* \|_{G,F}^2 \lesssim_{G,F} \frac{\mu_1 r \log (sn)}{n}. \]
The upper bound of \( \| UV^* \|_{G,\infty} \) can be established as follows. Note that
\[ \| UV^* \|_{G,\infty} = \max_{0 \leq i \leq n-1} \frac{\| G^* (UV^*) e_i \|_2}{\sqrt{w_i}}. \]
For any fixed \( i \in [n] \), we have
\[
\frac{\| G^* (UV^*) e_i \|_2^2}{\sqrt{w_i}} = \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} | \langle G^* (UV^*) e_i, \beta \rangle | \\
= \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} \left| \langle UV^*, (\sum_{0 \leq j \leq n_1 - 1} e_j e_k^T) \otimes \beta \rangle \right| \\
\leq \frac{1}{\sqrt{w_i}} \sup_{\| \beta \|_2 = 1} \left( \sum_{0 \leq j \leq n_1 - 1} \| \langle e_j \otimes \beta \rangle^* U \|_2 \right) \| e_k^T V \|_2 \\
\leq \sup_{\| \beta \|_2 = 1} \left( \sum_{0 \leq j \leq n_1 - 1} \| (e_j \otimes \beta)^* U \|_2 \right) \| e_k^T V \|_2 \\
= \sup_{\| \beta \|_2 = 1} \left( \sum_{0 \leq j \leq n_1 - 1} \| \beta^* U_j \|_2 \right) \\
= \sup_{\| \beta \|_2 = 1} \left( \sum_{0 \leq j \leq n_1 - 1} \| e_k^T V \|_2 \right) \\
\leq \frac{1}{w_i} \sum_{0 \leq j \leq n_1 - 1} \| (e_j \otimes \beta)^* U \|_2 \\
\leq \frac{1}{w_i} \sum_{0 \leq j \leq n_1 - 1} \| e_k^T V \|_2 \]
where the fourth line is due to the definition of \( G \) in (I.18) and the last line follows from (III.5) and (III.6). Therefore, \( \| UV^* \|_{G,\infty} \leq \frac{\mu_1 r}{n} \).

VI. Auxiliary Results

In this section, we present some necessary results which have been used in the previous proofs. The following lemma is used in the proof of Theorem III.8.

Lemma VI.1: Suppose \( \| AA^* \| \geq 1 \) and
\[ \| P_T G A^* G^* P_T - P_T G G^* P_T \| \leq \frac{1}{2}. \]
For any \( M \in \mathbb{C}^{n_1 \times n_2} \) which obeys
\[ A^* G^* (M) = 0 \quad \text{and} \quad (I - G G^*) (M) = 0, \]
we have
\[ \| P_T (M) \|_F \leq 4 s_0 \| P_T \cdot (M) \|_F. \]

Proof: It follows from (III.13) and (III.14) that
\[
0 = \| (G A^* A^* G^* + (I - G G^*)) (M) \|_F \\
\geq \| (G A^* A^* G^* + (I - G G^*)) P_T (M) \|_F \\
- \| (G A^* A^* G^* + (I - G G^*)) P_{T^\perp} (M) \|_F.
\]
For the first term,
\[
\| (G A^* A^* G^* + (I - G G^*)) P_T (M) \|_F^2 \\
= \| G A^* A^* G^* P_T (M) \|_F^2 + \| (I - G G^*) P_T (M) \|_F^2 \\
= \| G A^* A^* G^* P_T (M) \|_F^2 \\
+ \| P_T (M) \|_F^2 \\
= \| G A^* A^* G^* P_T (M) \|_F^2 \\
+ \| P_T (M) \|_F^2 \\
\geq \| P_T (M) \|_F^2 \\
- \| P_T (G A^* A^* G^* - G G^*) P_T (M) \|_F \\
\geq \frac{1}{2} \| P_T (M) \|_F^2.
\]
where the fourth step is due to (III.2) in Lemma III.2. For the second term,
\[
\| (G A^* A^* G^* + (I - G G^*)) P_{T^\perp} (M) \|_F \\
\leq \| (G A^* A^* G^*) P_{T^\perp} (M) \|_F + \| (I - G G^*) P_{T^\perp} (M) \|_F \\
\leq \| G \| \cdot \| A^* A \| \cdot \| G^* \| \cdot \| P_{T^\perp} (M) \|_F \\
+ \| I - G G^* \| \cdot \| P_{T^\perp} (M) \|_F \\
\leq (1 + s_0) \| P_{T^\perp} (M) \|_F \\
\leq 2 s_0 \| P_{T^\perp} (M) \|_F.
\]
where the third line is due to \( \| G \| = 1, \| G^* \| \leq 1 \) and (III.3) in Lemma III.2.
Combining these two terms together completes the proof.

The following lemmas play an important role in the proofs of Lemmas III.9 to III.15.

**Lemma VI.2:** Recall that $U$ and $V$ obey (III.4). For any fixed $z \in \mathbb{C}^n$, there holds

$$
\max_{0 \leq i \leq n-1} \left\| U^* G(z e_i^T) \right\|_F^2 \leq \frac{\mu_{1R}}{n}, \quad (VI.1)
$$

$$
\max_{0 \leq i \leq n-1} \left\| G(z e_i^T) V \right\|_F^2 \leq \frac{\mu_{1R}}{n}, \quad (VI.2)
$$

$$
\max_{0 \leq i \leq n-1} \left\| P_T G(z e_i^T) \right\|_F^2 \leq 2 \left\| z \right\|_2^2 \frac{\mu_{1R}}{n}. \quad (VI.3)
$$

**Proof:** To show (VI.1), note that for any $0 \leq i \leq n-1$,

$$
G(z e_i^T) = G_i \otimes z
$$

$$
= \left( \sum_{0 \leq j \leq n_i-1} \frac{1}{\sqrt{w_i}} e_j e_k^T \right) \otimes z
$$

$$
= \sum_{0 \leq j \leq n_i-1} \frac{1}{\sqrt{w_i}} (e_j \otimes z) e_k^T,
$$

where the second equality is due to the definition of $G_i$ in (I.18). It follows that

$$
\left\| U^* G(z e_i^T) \right\|_F^2
$$

$$
= \left\langle U^* G(z e_i^T), U^* G(z e_i^T) \right\rangle
$$

$$
= \frac{1}{w_i} \left\langle \sum_{0 \leq j \leq n_i-1} \frac{1}{\sqrt{w_i}} U^* (e_j \otimes z) e_k^T, \sum_{0 \leq k \leq n_i-1} \frac{1}{\sqrt{w_i}} U^* (e_p \otimes z) e_q^T \right\rangle
$$

$$
= \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \left\| U^* (e_j \otimes z) \right\|_2^2
$$

$$
= \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \left\| U^* z \right\|_2^2
$$

$$
\leq \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \left\| z \right\|_2^2 \cdot \left\| U_j \right\|_F^2
$$

$$
\leq \left\| z \right\|_2^2 \cdot \frac{\mu_{1R}}{n},
$$

where the last step follows from (III.5).

As for (VI.2), note that

$$
\left\| G(z e_i^T) V \right\|_F^2
$$

$$
= \left\langle G(z e_i^T) V, G(z e_i^T) V \right\rangle
$$

$$
= \frac{1}{w_i} \left\langle \sum_{0 \leq j \leq n_i-1} \frac{1}{\sqrt{w_i}} (e_j \otimes z) e_k^T V, \sum_{0 \leq k \leq n_i-1} \frac{1}{\sqrt{w_i}} (e_p \otimes z) e_q^T V \right\rangle
$$

$$
= \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \sum_{0 \leq k \leq n_i-1} \left\langle (e_j \otimes z) e_k^T V, (e_p \otimes z) e_q^T V \right\rangle
$$

$$
= \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \sum_{0 \leq k \leq n_i-1} \left\langle e_j^T V, e_k^T V \right\rangle \left\langle e^T e_j, e^T e_p \right\rangle
$$

$$
= \frac{1}{w_i} \sum_{0 \leq j \leq n_i-1} \left\| e_j^T V \right\|_2^2 \left\| e_j \right\|_2^2
$$

$$
\leq \left\| e \right\|_2^2 \frac{\mu_{1R}}{n},
$$

where the last step is also due to (III.6).

For the inequality (VI.3), by the definition of $P_T$ in (III.8), we have

$$
\left\| P_T G(z e_i^T) \right\|_F^2
$$

$$
= \left\langle P_T G(z e_i^T), P_T G(z e_i^T) \right\rangle
$$

$$
= \left\langle P T G(z e_i^T), G(z e_i^T) \right\rangle
$$

$$
= \langle U U^* G(z e_i^T) + G(z e_i^T) V V^*, G(z e_i^T) \rangle
$$

$$
= \left\| U^* G(z e_i^T) \right\|_F^2 + \left\| G(z e_i^T) V \right\|_F^2
$$

$$
- \left\| U^* G(z e_i^T) V + G(z e_i^T) V \right\|_F^2
$$

$$
\leq \left\| U^* G(z e_i^T) \right\|_F^2 + \left\| G(z e_i^T) V \right\|_F^2
$$

$$
\leq 2 \left\| z \right\|_2^2 \frac{\mu_{1R}}{n},
$$

which completes the proof.

After replacing $z$ with $b_i$ in Lemma VI.2, we obtain the following corollary based on the incoherence property (II.6) of $b_i$, where $b_i$ is the $i$th column of $B^*$.

**Corollary VI.3:** Under the condition (III.4), there holds

$$
\max_{0 \leq i \leq n-1} \left\| U^* G(b_i e_i^T) \right\|_F^2 \leq \frac{\mu_{01R} \mu_{1}}{n}, \quad (VI.4)
$$

$$
\max_{0 \leq i \leq n-1} \left\| G(b_i e_i^T) V \right\|_F^2 \leq \frac{\mu_{01R} \mu_{1}}{n}, \quad (VI.5)
$$

$$
\max_{0 \leq i \leq n-1} \left\| P_T G(b_i e_i^T) \right\|_F^2 \leq \frac{2 \mu_{01R} \mu_{1}}{n}. \quad (VI.6)
$$

**Lemma VI.4:** Under the condition (III.4), for any fixed matrix $W \in \mathbb{C}^{n_1 \times n_2}$,

$$
\left\| G^* P_T(W) e_i \right\|_2 \leq \left\| W \right\|_F \sqrt{\frac{2 \mu_{1R}}{n}}. \quad (VI.7)
$$
Proof: The result follows from a direct calculation:
\[
\|G^* \mathcal{P}_T(W)e_i\|_2 = \sup_{\|\beta\|_2 = 1} |\langle G^* \mathcal{P}_T(W)e_i, \beta \rangle| \\
= \sup_{\|\beta\|_2 = 1} |\langle G^* \mathcal{P}_T(W), \beta e_i^T \rangle| \\
= \sup_{\|\beta\|_2 = 1} |\langle \mathcal{P}_T(W), G(\beta e_i^T) \rangle| \\
\leq \|W\|_F \cdot \sup_{\|\beta\|_2 = 1} \|\mathcal{P}_T(G(\beta e_i^T))\|_F \\
\leq \|W\|_F \cdot \frac{2\mu_1 r}{n},
\]
where the last line follows from (VI.3) in Lemma VI.2.

By combining Lemmas VI.2 and VI.4, the following corollary can be established, which is used in the proof of (III.22).

**Corollary VI.5:** For any fixed matrix $W \in \mathbb{C}^{n \times n}$, under the condition (III.3), there holds
\[
\max_{0 \leq i \leq n-1} \|\mathcal{P}_T(G(\mathcal{P}_T(W)e_i e_i^T))\|_2^2 \\
\leq \|W\|_F^2 \cdot \frac{(2\mu_1 r)^2}{n},
\]
where the last line is due to Lemma VI.4.

**Lemma VI.6:** For any two fixed vectors $\beta, \gamma \in \mathbb{C}^n$,
\[
\sqrt{\frac{w_i}{w_j}} |\langle \mathcal{P}_T(G(\beta e_i^T), G(\gamma e_i^T))\rangle| \leq 3\mu_1 r \cdot \|\beta\|_2 \cdot \|\gamma\|_2
\]
holds for any $(i, j) \in [n] \times [n]$.

**Proof:** Recall that
\[
G(\beta e_i^T) = G_i \otimes \beta = \left( \frac{1}{w_i} \sum_{k+t=i}^{0 \leq k \leq n-1, 0 \leq t \leq n-1} e_k e_i^T \right) \otimes \beta, \\
G(\gamma e_i^T) = G_j \otimes \gamma = \left( \frac{1}{w_j} \sum_{p+q=j}^{0 \leq p \leq n-1, 0 \leq q \leq n-2} e_p e_q^T \right) \otimes \gamma.
\]
By the definition of $\mathcal{P}_T$ in (III.8), we have
\[
\sqrt{\frac{w_i}{w_j}} |\langle \mathcal{P}_T(G(\beta e_i^T), G(\gamma e_i^T))\rangle| \\
\leq \sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta), G_j \otimes \gamma \rangle| \\
+ \sqrt{\frac{w_i}{w_j}} |\langle (G_i \otimes \beta) VV^*, G_j \otimes \gamma \rangle| \\
+ \sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta) VV^*, G_j \otimes \gamma \rangle|.
\]
It suffices to bound each of the three terms separately. For the first term, we have
\[
\sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta), G_j \otimes \gamma \rangle| \\
= \sqrt{\frac{w_i}{w_j}} \left| \left( \sum_{0 \leq k \leq n-1, 0 \leq t \leq n-2} \frac{1}{w_i} e_k e_t^T \right) \otimes \beta \right| \\
\leq \sqrt{\frac{w_i}{w_j}} \left| \left( \sum_{0 \leq p \leq n-1, 0 \leq q \leq n-2} (e_p \otimes \gamma) e_q^T \right) \right| \\
= \frac{1}{\sqrt{w_j}} \left| \sum_{0 \leq p \leq n-1, 0 \leq q \leq n-2} \langle UU^* (e_k \otimes \beta) e_t^T, (e_p \otimes \gamma) e_q^T \rangle \right| \\
= \frac{1}{\sqrt{w_j}} \left| \sum_{0 \leq p \leq n-1, 0 \leq q \leq n-2} \langle U^*(e_{i-q} \otimes \beta), U^*(e_p \otimes \gamma) \rangle \right| \\
\leq \frac{1}{\sqrt{w_j}} \sum_{0 \leq p \leq n-1, 0 \leq q \leq n-2} \|U^*(e_{i-q} \otimes \beta)\|_2 \cdot \|U^*(e_p \otimes \gamma)\|_2
\]
holds for any $(i, j) \in [n] \times [n]$.

The second term can be bounded in a similar way. For the last term, we have
\[
\sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta) VV^*, G_j \otimes \gamma \rangle| \\
= \sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta), (G_j \otimes \gamma) VV^* \rangle| \\
= \sqrt{\frac{w_i}{w_j}} |\langle UU^* (G_i \otimes \beta), G_j \otimes \gamma \rangle |.
\]
\[
\begin{align*}
\|\beta\|_2 & \leq \frac{\|\beta\|_2 \cdot \|\gamma\|_2}{n} \leq \frac{\|\beta\|_2 \cdot \|\gamma\|_2}{n}, \\
\text{where the last step is due to (III.6).} \\
\text{Combining the three bounds together completes the proof.} \quad \Box
\end{align*}
\]

The following lemma is established in [15] and the proof will be omitted here.

**Lemma VI.7:** Suppose a matrix \( F \in \mathbb{C}^{n_1 \times n_2} \) satisfies
\[
\max_{0 \leq i \leq n_1-1} \|e_i^T F\|_2^2 \leq B. \tag{VI.8}
\]
We have
\[
\sum_{i=0}^{n-1} \frac{1}{w_i} \|\langle F, G_i \rangle\|_2^2 \lesssim B \log(n). \tag{VI.9}
\]
We will apply this lemma to upper bound \( \|Z\|_{G,F} \) for \( Z \in \mathbb{C}^{n_1 \times n_2} \). Note that \( Z \) can be written as
\[
Z = \begin{bmatrix}
  z_{0,0} & \cdots & z_{0,n_2-1} \\
  \vdots & \ddots & \vdots \\
  z_{n_1-1,0} & \cdots & z_{n_1-1,n_2-1}
\end{bmatrix},
\]
where \( z_{i,j} \in \mathbb{C}^{n_1 \times n_2} \) is the \((i,j)\)th block of \( Z \).

**Corollary VI.8:** For any matrix \( Z \in \mathbb{C}^{n_1 \times n_2} \) satisfying
\[
\max_{0 \leq i \leq n_1-1} \sum_{j=0}^{n_2-1} \|z_{i,j}\|_2 \leq B, \tag{VI.10}
\]
we have
\[
\|Z\|_{G,F}^2 \lesssim B \log(n). \tag{VI.11}
\]

**Proof:** Define the matrix
\[
\tilde{Z} = \begin{bmatrix}
  \|z_{0,0}\|_2 & \cdots & \|z_{0,n_2-1}\|_2 \\
  \vdots & \ddots & \vdots \\
  \|z_{n_1-1,0}\|_2 & \cdots & \|z_{n_1-1,n_2-1}\|_2
\end{bmatrix} \in \mathbb{R}^{n_1 \times n_2}.
\]
The definition of \( G^* \) implies that the \( i \)th column of \( G^*(Z) \) is given by
\[
G^*(Z)e_i = \frac{1}{\sqrt{w_i}} \sum_{j+k=i, 0 \leq j \leq n_1-1, 0 \leq k \leq n_2-1} z_{j,k}.
\]
It follows that
\[
\|Z\|_{G,F}^2 = \sum_{i=0}^{n-1} \frac{1}{w_i} \|G^*(Z)e_i\|_2^2 \\
= \sum_{i=0}^{n-1} \frac{1}{w_i} \left( \sum_{j+k=i, 0 \leq j \leq n_1-1, 0 \leq k \leq n_2-1} \|z_{j,k}\|_2^2 \right) \\
\leq \sum_{i=0}^{n-1} \frac{1}{w_i} \left( \frac{1}{\sqrt{w_i}} \sum_{j+k=i, 0 \leq j \leq n_1-1, 0 \leq k \leq n_2-1} \|z_{j,k}\|_2^2 \right) \\
= \sum_{i=0}^{n-1} \frac{1}{w_i} \left( \frac{1}{\sqrt{w_i}} \sum_{j+k=i, 0 \leq j \leq n_1-1, 0 \leq k \leq n_2-1} \langle \tilde{Z} e_k, e_j \rangle \right)^2
\]
It has been used in the proofs of (III.24) and (III.29).

The proof is completed after combining the three bounds together.

\[ \max_{0 \leq i \leq n_1-1} \left\| \epsilon_i^T \mathbf{Z} \right\|_2^2 \leq B, \] applying Lemma VI.7 completes the proof. \( \square \)

The following lemma can be established based on Corollary VI.8. It has been used in the proofs of (III.24) and (III.29).

**Lemma VI.9:** For any fixed \( z \in \mathbb{C}^r \),

\[ \left\| \mathcal{P}_T \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \right\|_{G,F}^2 \lesssim \left\| z \right\|_2^2 \frac{\mu_1 r \log(sn)}{n}. \]

**Proof:** Recalling the definition of \( \mathcal{P}_T \) in (III.8), we have

\[ \mathcal{P}_T \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) = \mathbf{U}^* \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) + \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \mathbf{V}^* - \mathbf{U} \mathbf{U}^* \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \mathbf{V}^*. \]

It suffices to bound the three terms separately. For the first term, recall that \( \mathbf{U} \in \mathbb{C}^{sn \times n_1^r} \) can be rewritten as

\[ \mathbf{U} = \begin{bmatrix} \mathbf{U}_0 \\ \vdots \\ \mathbf{U}_{n_1-1} \end{bmatrix}, \]

where \( \mathbf{U}_t \in \mathbb{C}^{s \times r} \) is the \( t \)-th block. Since

\[ \left\| \mathbf{U}_t \mathbf{U}^* \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \right\|_{F}^2 \]

\[ = w_i \mathbf{U}_t \mathbf{G}_i \left( \mathbf{G}_i \otimes z \right) - \mathbf{G}_i \left( \mathbf{G}_i \otimes z \right) \mathbf{U}_t \mathbf{U}^* \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \mathbf{V}^* - \mathbf{U} \mathbf{U}^* \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \mathbf{V}^*. \]

The same bound can be obtained for \( \mathcal{G} \left( \mathbf{w}_i \epsilon^T \right) \mathbf{V}^* \).

The proof is completed after combining the three bounds together. \( \square \)

**VII. Conclusion**

A convex approach called Vectorized Hankel Lift is proposed for blind super-resolution. It is based on the observation that the corresponding vectorized Hankel matrix is low rank if the Fourier samples of the unknown PSFs lie in a low dimensional subspace. Theoretical guarantee has been established for Vectorized Hankel Lift, showing that exact resolution can be achieved provided the number of samples is nearly optimal. We leave the robust analysis of the method to the future work. In particular, we would like to see whether the technique that bridges convex and nonconvex programs in [17] may yield an optimal error bound for the blind super-resolution problem.

For low rank matrix recovery and spectrally sparse signal recovery, many simple yet efficient nonconvex iterative algorithms have been developed and analyzed based on inherent low rank structures of the problems [5]–[7], [60], [61]. Thus, it is also interesting to develop nonconvex optimization methods for blind super-resolution based on the low rank structure of the vectorized Hankel matrix. Actually, this has been done in a recent released work [44] where a nonconvex first order method has been proposed.

For the single snapshot MUSIC and the MMV MUSIC, the super-resolution effect has been studied in [34], [36], [41]. Since the spatial smoothing MUSIC is designed to improve the performance of the MMV MUSIC, it is also interesting to investigate the super-resolution effect of this variant. The equivalence between it and MUSIC through Vectorized Hankel Lift (i.e., Lemma II.2) may provide a new perspective to approach this problem.

**Acknowledgment**

Ke Wei would like to thank Wenjing Liao for fruitful discussions on the subspace methods for line spectrum estimation, and would like to thank Zai Yang for pointing out that the MUSIC variant arising naturally from the vectorized Hankel lift framework is indeed equivalent to the spatial smoothing technique proposed to improve the performance of the MMV MUSIC.

## References

[1] A. Ahmed, B. Recht, and J. Romberg, “Blind deconvolution using convex programming,” IEEE Trans. Inf. Theory, vol. 60, no. 3, pp. 1711–1732, Mar. 2014.

[2] T. Bendory, D. Dekel, and A. Feuer, “Robust recovery of stream of pulses using convex optimization,” J. Math. Anal. Appl., vol. 442, no. 2, pp. 511–536, Oct. 2016.

[3] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, “Signal processing for passive radar using OFDM waveforms,” IEEE J. Sel. Topics Signal Process., vol. 4, no. 1, pp. 226–238, Feb. 2010.

[4] B. Bernstein and C. Fernandez-Granda, “Deconvolution of point sources: A sampling theorem and robustness guarantees,” Commun. Pure Appl. Math., vol. 72, no. 6, pp. 1152–1230, 2019.

[5] J.-F. Cai, S. Liu, and W. Xu, “A fast algorithm for reconstruction of spectrally sparse signals in super-resolution,” Proc. SPIE, vol. 9597, Aug. 2015, Art. no. 95970A.

[6] J.-F. Cai, T. Wang, and K. Wei, “Spectral compressed sensing via projected gradient descent,” SIAM J. Optim., vol. 28, no. 3, pp. 2625–2653, Jan. 2018.

[7] 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., vol. 46, pp. 94–121, Jan. 2019.

[8] E. J. Candès and C. Fernandez-Granda, “Super-resolution from noisy data,” J. Fourier Anal. Appl., vol. 19, no. 6, pp. 1229–1254, 2013.
Jinchi Chen received the B.S. degree in electronic engineering from the Nanjing Institute of Technology, Nanjing, China, in 2012, and the Ph.D. degree in physics from the Beijing Institute of Technology, Beijing, China, in 2019. Since 2019, he has been a Post-Doctoral Researcher with the School of Data Science, Fudan University. His research interests include compressed sensing and mathematical of data science.

Ke Wei received the B.Sc. and M.Sc. degrees in mathematics from Shandong University in 2007 and 2010, respectively, and the Ph.D. degree in applied and computational mathematics from the University of Oxford in 2014. He was a Post-Doctoral Research Associate with The Hong Kong University of Science and Technology from 2014 to 2015 and a Post-Doctoral Fellow with the University of California at Davis from 2015 to 2017. He is currently an Associate Professor with the School of Data Science, Fudan University. His research interests include high-dimensional signal processing and data analysis, and mathematical optimization.