Abstract

Suppose that we have \( r \) sensors and each one intends to send a function \( z_i \) (e.g. a signal or an image) to a receiver common to all \( r \) sensors. During transmission, each \( z_i \) gets convolved with a function \( g_i \). The receiver records the function \( y \), given by the sum of all these convolved signals. 

When and under which conditions is it possible to recover the individual signals \( z_i \) and the blurring functions \( g_i \) from just one received signal \( y \)? This challenging problem, which intertwines blind deconvolution with blind demixing, appears in a variety of applications, such as audio processing, image processing, neuroscience, spectroscopy, and astronomy. It is also expected to play a central role in connection with the future Internet-of-Things. We will prove that under reasonable and practical assumptions, it is possible to solve this otherwise highly ill-posed problem and recover the \( r \) transmitted functions \( z_i \) and the impulse responses \( g_i \) in a robust, reliable, and efficient manner from just one single received function \( y \) by solving a semidefinite program. We derive explicit bounds on the number of measurements needed for successful recovery and prove that our method is robust in presence of noise. Our theory is actually a bit pessimistic, since numerical experiments demonstrate that, quite remarkably, recovery is still possible if the number of measurements is close to the number of degrees of freedom.

Keywords— blind deconvolution, demixing, semidefinite programming, nuclear norm minimization, channel estimation, low-rank matrix.

1 Introduction

Suppose we are given \( r \) sensors, each one sends a function \( z_i \) (e.g. a signal or image) to a receiver common to all \( r \) sensors. During transmission each \( z_i \) gets convolved with a function \( g_i \) (the \( g_i \) may all differ from each other). The receiver records the function \( y \), given by the sum of all these convolved signals. More precisely, 

\[
y = \sum_{i=1}^{r} g_i \ast z_i + w, \tag{1}
\]

where \( w \) is additive noise. Assume that the receiver knows neither \( \{z_i\}_{i=1}^{r} \) nor \( \{g_i\}_{i=1}^{r} \). When and under which conditions is it possible to recover all the individual signals \( z_i \) and \( g_i \) from just one received signal \( y \)?

*This research was partially supported by the NSF via grant DTRA-DMS 1322393.
Blind deconvolution (when \( r = 1 \)) by itself is already a hard problem to solve. Here we deal with the even more difficult situation of a mixture of blind deconvolution problems. Thus we need to correctly blindly deconvolve and demix at the same time. This challenging problem appears in a variety of applications, such as audio processing [27], image processing [31, 32], neuroscience [37], spectroscopy [38], astronomy [12]. It also arises in wireless communications [42] and is expected to play a central role in connection with the future Internet-of-Things [45]. Common to almost all approaches to tackle this problem is the assumption that we have multiple received signals at our disposal, often at least as many received signals as there are transmitted signals. Indeed, many of the existing methods fail if the assumption of multiple received signals is not fulfilled.

In this paper, we consider the rather difficult case, where only one received signal is given, as shown in (1). Of course, without further assumptions, this problem is highly underdetermined and not solvable. We will prove that under reasonable and practical conditions, it is indeed possible to recover the \( r \) transmitted signals and the associated channels in a robust, reliable, and efficient manner from just one single received signal. Our theory has important implications for applications, such as the Internet-of-Things, since it paves the way for an efficient multi-sensor communication strategy with minimal signaling overhead.

To provide a glimpse of the kind of results we will prove, let us assume that each of the \( z_i \in \mathbb{R}^N \) lies in a known subspace of dimension \( N \), i.e., there exists matrices \( A_i \) of size \( L \times N \) such that \( z_i = A_i x_i \). In addition the matrices \( A_i \) need to satisfy a certain “local” mutual incoherence condition described in detail in [27]. This condition can be satisfied if the \( A_i \) are, e.g., Gaussian random matrices. We will prove a formal and slightly more general version (see Theorem 3.1 and Theorem 3.3 of the following informal theorem. For simplicity for the moment we consider a noise-free scenario, that is, \( w = 0 \). Below and throughout the paper “∗” denotes circular convolution.

**Theorem 1.1 (Informal version).** Let \( x_i \in \mathbb{R}^N \) and let the \( A_i \) be \( L \times N \) i.i.d. Gaussian random matrices. Furthermore, assume that the impulse responses \( g_i \in \mathbb{C}^N \) have maximum delay spread \( K \), i.e., for each \( g_i \) there holds \( g_i(k) = 0 \) if \( k > K \). Let \( \mu_2^2 \) be a certain “incoherence parameter” related to the measurement matrices, defined in [13]. Suppose we are given

\[
y = \sum_{i=1}^{r} g_i * (A_i x_i).
\]  

Then, as long as the number of measurements \( L \) satisfies

\[
L \gtrsim C r^2 \max \{ K, \mu_2^2 N \} \log^3 L \log(r + 1),
\]

(where \( C \) is a numerical constant), all \( x_i \) (and thus \( z_i = A_i x_i \)) as well as all \( g_i \) can be recovered from \( y \) with high probability by solving a semidefinite program.

Recovering \( \{x_i\}_{i=1}^{r} \) and \( \{g_i\}_{i=1}^{r} \) is only possible up to a constant, since we can always multiply each \( x_i \) with \( c_i \neq 0 \) and each \( g_i \) with \( 1/c_i \) and still get the same result. Hence, here and throughout the paper, recovery of the vectors \( x_i \) and \( g_i \) always means recovery modulo constants \( c_i \).

We point out that the emphasis of this paper is on developing a theoretical and algorithmic framework for joint blind deconvolution and blind demixing. A detailed discussion of applications is beyond the scope of this paper. There are several aspects, such as time synchronization, that do play a role in some applications and need further attention. We postpone such details to a forthcoming paper, in which we plan to elaborate on the proposed framework in connection with specific applications.

---

\[1\] In wireless communications this is also known as “multiuser joint channel estimation and equalization.”
1.1 Related work

Problems of the type (1) or (2) are ubiquitous in many applied scientific disciplines and in applications, see e.g [17, 42, 27, 34, 33, 23, 32, 37, 38, 12, 45]. Thus, there is a large body of works to solve different versions of these problems. Most of the existing works however require the availability of multiple received signals \( y_1, \ldots, y_m \). And indeed, it is not hard to imagine that for instance an SVD-based approach will succeed if \( m \geq r \) (and must fail if \( m = 1 \)). A sparsity-based approach can be found in [36]. However, in this paper we are interested in the case where we have only one single received signal \( y \) – a single snapshot, in the jargon of array processing. Hence, there is little overlap between these methods heavily relying on multiple snapshots (many of which do not come with any theory) and the work presented here.

The setup in (2) is reminiscent of a single-antenna multi-user spread spectrum communication scenario [40]. There, the matrix \( A_i \) represents the spreading matrix assigned to the \( i \)-th user and \( g_i \) models the associated multipath channel. There are numerous papers on blind channel estimation in connection with CDMA, including the previously cited articles [17, 12, 23]. Our work differs from the existing literature on this topic in several ways: As mentioned before, we do not require that we have multiple received signals, we allow all multipath channels \( g_i \) to differ from each other, and do not impose a particular channel model. Moreover, we provide a rigorous mathematical theory, instead of just empirical observations.

The special case \( r = 1 \) (one unknown signal and one unknown convolving function) reduces (1) to the standard blind deconvolution problem, which has been heavily studied in the literature, cf. [13] and the references therein. Many of the techniques for “ordinary” blind deconvolution do not extend (at least not in any obvious manner) to the case \( r > 1 \). Hence, there is essentially no overlap with this work – with one notable exception. The pioneering paper [2] has definitely inspired our work and also informed many of the proof techniques used in this paper. Hence, our paper can and should be seen as an extension of the “single-user” \( (r = 1) \) results in [2] to the multi-user setting \( (r > 1) \). However, it will not come as a big surprise to the reader familiar with [2], that there is no simple way to extend the results in [2] to the multi-user setting unless we assume that we have multiple received signals \( y_1, \ldots, y_m \). Indeed, as may be obvious from the length of the proofs in our paper, there are substantial differences in the theoretical derivations between this manuscript and [2]. In particular, the sufficient condition for exact recovery in this paper is more complicated since \( r \) \( (r > 1) \) users are considered and the “incoherence” between users need to be introduced properly. Moreover, the construction of approximate dual certificate is nontrivial as well (See Section 5.4) in the “multi-user” scenario.

The paper [1] considers the following generalization of [2]2. Assume that we are given signals \( y_i = g * z_i, i = 1, \ldots, r, \) the goal is to recover the \( z_i \) and \( g \) from \( y_1, \ldots, y_r \). This setting is somewhat in the spirit of [1], but it is significantly less challenging, since (i) it assumes the same convolution function \( g \) for each signal \( z_i \) and (ii) there are as many output signals \( y_i \) as we have input signals \( z_i \).

Non-blind versions of (1) or (2) can be found for instance in [44, 29, 28, 3]. In the very interesting paper [44], the authors analyze various problems of decomposing a given observation into multiple incoherent components, which can be expressed as

\[
\text{minimize} \quad \sum_i \lambda_i \|Z_i\|_{(i)} \quad \text{subject to} \quad \sum_i Z_i = M. \tag{3}
\]

Here \( \|\cdot\|_{(i)} \) are (decomposable) norms that encourage various types of low-complexity structure. However, as mentioned before, there is no “blind” component in the problems analyzed in [44]. Moreover, while [3] is formally somewhat similar to the semidefinite program that we derive to solve the blind deconvolution-blind demixing problem (see (8)), the dissimilarity of the right-hand sides in [3] and (8) makes all the differences when theoretically analyzing these two problems.

2Since the main result in [1] relies on Lemma 4 of [2], the issues raised in Remark 2.1 apply to [1] as well.
The current manuscript can as well be seen as an extension of our work on self-calibration \cite{26} to the multi-sensor case. In this context, we also refer to related (single-input-single-output) analysis in \cite{20} \cite{21}.

1.2 Organization of this manuscript

In Section \[2\] we describe in detail the setup and the problem we are solving. We also introduce some notations and key concepts used throughout the manuscript. The main results for the noiseless as well as the noisy case are stated in Section \[3\]. Numerical experiments can be found in Section \[4\]. Section \[3\] is devoted to the proofs of these results. We conclude in Section \[6\] and present some auxiliary results in the Appendix.

2 Preliminaries and Basic Setup

2.1 Notation

Before moving to the basic model, we introduce notation which will be used throughout the paper. Matrices and vectors are denoted in boldface such as $\mathbf{Z}$ and $\mathbf{z}$. The individual entries of a matrix or a vector are denoted in normal font such as $Z_{ij}$ or $z_i$. For any matrix $\mathbf{Z}$, $\|\mathbf{Z}\|_*$ denotes nuclear norm, i.e., the sum of its singular values; $\|\mathbf{Z}\|$ denotes operator norm, i.e., its largest singular value, and $\|\mathbf{Z}\|_F$ denotes the Frobenius norm, i.e., $\|\mathbf{Z}\|_F = \sqrt{\sum_{ij} |Z_{ij}|^2}$. For any vector $\mathbf{z}$, $\|\mathbf{z}\|$ denotes its Euclidean norm. For both matrices and vectors, $\mathbf{Z}^T$ and $\mathbf{z}^T$ stand for the transpose of $\mathbf{Z}$ and $\mathbf{z}$ respectively while $\mathbf{Z}^*$ and $\mathbf{z}^*$ denote their complex conjugate transpose. $\bar{\mathbf{z}}$ and $\bar{\mathbf{z}}^*$ denote the complex conjugate of $\mathbf{z}$ and $\mathbf{z}$ respectively. We equip the matrix space $\mathbb{C}^{K \times N}$ with the inner product defined as $\langle \mathbf{U}, \mathbf{V} \rangle := \text{Tr}(\mathbf{U}^\dagger \mathbf{V})$. A special case is the inner product of two vectors, i.e., $\langle \mathbf{u}, \mathbf{v} \rangle = \text{Tr}(\mathbf{u}^\dagger \mathbf{v}^\dagger) = \mathbf{v}^\dagger \mathbf{u} = (\mathbf{u}^\dagger \mathbf{v})^*$. The identity matrix of size $n$ is denoted by $\mathbf{I}_n$. For a given vector $\mathbf{v}$, diag($\mathbf{v}$) represents the diagonal matrix whose diagonal entries are given by the vector $\mathbf{v}$.

Throughout the paper, $C$ stands for a constant and $C_\alpha$ is a constant which depends linearly on $\alpha$ (and on no other numbers). For the two linear subspaces $T_i$ and $T_i^\perp$ defined in \cite{24} and \cite{25}, we denote the projection of $\mathbf{Z}$ on $T_i$ and $T_i^\perp$ as $\mathbf{Z}_{T_i} := \mathcal{P}_{T_i}(\mathbf{Z})$ and $\mathbf{Z}_{T_i^\perp} := \mathcal{P}_{T_i^\perp}(\mathbf{Z})$ respectively. $\mathcal{P}_{T_i}$ and $\mathcal{P}_{T_i^\perp}$ are the corresponding projection operators onto $T_i$ and $T_i^\perp$.

2.2 The basic model

We develop our theory for a more general model than the blind deconvolution/blind demixing model discussed in Section \[1\]. Our framework also covers certain self-calibration scenarios \cite{26} involving multiple sensors. We consider the following setup \cite{3}

$$\mathbf{y} = \sum_{i=1}^{r} \text{diag}(\mathbf{B}_i, \mathbf{h}_i) \mathbf{A}_i \mathbf{x}_i, \quad (4)$$

where $\mathbf{y} \in \mathbb{C}^L$, $\mathbf{B}_i \in \mathbb{C}^{L \times K_i}$, $\mathbf{A}_i \in \mathbb{R}^{L \times N_i}$, $\mathbf{h}_i \in \mathbb{R}^{K_i}$ and $\mathbf{x}_i \in \mathbb{R}^{N_i}$. We assume that all the matrices $\mathbf{B}_i$ and $\mathbf{A}_i$ are given, but none of the $\mathbf{x}_i$ and $\mathbf{h}_i$ are known. Note that all $\mathbf{h}_i$ and $\mathbf{x}_i$ can be of different lengths. We point out that the total number of measurements is given by the length of $\mathbf{y}$, i.e., by $L$. Moreover, we let $K := \max K_i$ and $N := \max N_i$ throughout our presentation.

This model includes the blind deconvolution-blind demixing problem \cite{1} as a special case, as we will explain in Section \[3\]. But it also includes other cases as well. Consider for instance

\[3\] In \cite{3} we assume a common clock among the different sources. For sources whose distance to the receiver differs greatly, his assumption would require additional synchronization. A detailed discussion of this timing aspect is beyond the scope of this paper, as it is application dependent.
a linear system \( y = \sum_{i=1}^{r} A_i(\theta_i)x_i \), where the measurement matrices \( A_i \) are not fully known due to lack of calibration \([10], [4], [26]\) and \( \theta_i \) represents the unknown calibration parameters associated with \( A_i \). An important special situation that arises e.g. in array calibration \([16]\) is the case where we only know the direction of the rows of \( A_i \). In other words, the norms of each of the rows of \( A_i \) are unknown. If in addition each of the \( \theta_i \) belongs to a known subspace represented by \( B_i \), i.e., \( \theta_i = B_ih_i \), then we can write such an \( A_i(\theta_i) \) as \( A_i(\theta_i) = \text{diag}(B_ih_i)A_i \).

Let \( b_{i,l} \) denote the \( l \)-th column of \( B_i^* \) and \( a_{i,l} \) the \( l \)-th column of \( A_i^* \). A simple application of linear algebra gives

\[
y_i = \sum_{i=1}^{r} (B_ih_i)_l x_i^* a_{i,l} = \sum_{i=1}^{r} b_{i,l}^* h_i x_i^* a_{i,l}
\]

where \( y_i \) is the \( l \)-th entry of \( y \). One may find an obvious difficulty of this problem as the nonlinear relation between the measurement vectors \( (b_{i,l}, a_{i,l}) \) and the unknowns \( (h_i, x_i) \). Proceeding with the meanwhile well-established lifting trick \([10]\), we let \( X_i := h_i x_i^* \in \mathbb{R}^{K_i \times N_i} \) and define the linear mapping \( A_i : \mathbb{C}^{K_i \times N_i} \to \mathbb{C}^L \) for \( i = 1, \ldots, r \) by

\[
A_i(Z) := \{b_{i,l}^* Za_{i,l}\}_{l=1}^L.
\]

Note that the adjoint operator of \( A_i \) is

\[
A_i^* : \mathbb{C}^L \to \mathbb{C}^{K_i \times N_i}, \quad A_i^*(z) = \sum_{i=1}^L z_i b_{i,l}^* a_{i,l}
\]

since \( \mathbb{C}^{K_i \times N_i} \) is equipped with the inner product \( \langle U, V \rangle = \text{Tr}(UV^*) \) for any \( U \) and \( V \in \mathbb{C}^{K_i \times N_i} \). \( A_i^*(z) \) can be also written into simple matrix form, i.e., \( A_i^*(z) = B_i^* \text{diag}(z)A_i \), which is easily verified by definition. Thus we have lifted the non-linear vector-valued equations \([1]\) to linear matrix-valued equations given by

\[
y = \sum_{i=1}^{r} A_i(X_i).
\]

Alas, the set of linear equations \([7]\) will be highly underdetermined, unless we make the number of measurements \( L \) very large, which may not be desirable or feasible in practice. Moreover, finding such \( r \) rank-1 matrices satisfying \([7]\) is generally an NP-hard problem \([31], [15]\). Hence, to combat this underdeterminedness, we attempt to recover \( (h_i, x_i)_{i=1}^r \) by solving the following nuclear norm minimization problem,

\[
\min_{Z} \sum_{i=1}^{r} \|Z_i\|_* \quad \text{subject to} \quad \sum_{i=1}^{r} A_i(Z_i) = y.
\]

If the solutions (or the minimizers to \([8]\)) \( \hat{X}_1, \ldots, \hat{X}_r \) are all rank-one, we can easily extract \( h_i \) and \( x_i \) from \( \hat{X}_i \) via a simple matrix factorization. In case of noisy data, the \( \hat{X}_i \) will not be exactly rank-one, in which case we set \( h_i \) and \( x_i \) to be the left and right singular vector respectively, associated with the largest singular value of \( \hat{X}_i \). Naturally, the question arises if and when the solution to \([8]\) coincides with the true solution \( (h_i, x_i)_{i=1}^r \). It is the main purpose of this paper to shed light on this question.

### 2.3 Incoherence conditions on the matrices \( B_i \)

Analogous to matrix completion, where one needs to impose certain incoherence conditions on the singular vectors (see e.g. \([3]\)), we introduce two quantities that describe a notion of incoherence of the matrices \( B_i \). We require \( B_i^*B_i = I_{K_i} \) and define

\[
\mu_{\text{max}}^2 := \max_{1 \leq i \leq L, 1 \leq l \leq r} \frac{L}{K_i} \|b_{i,l}\|^2, \quad \mu_{\text{min}}^2 := \min_{1 \leq i \leq L, 1 \leq l \leq r} \frac{L}{K_i} \|b_{i,l}\|^2.
\]
\(B_i^* B_i = I_{K_i}\) implies that \(1 \leq \mu_{\text{max}}^2 \leq \frac{L}{K_i}\) and \(0 \leq \mu_{\text{min}}^2 \leq 1\). In particular, if each \(B_i\) is a partial DFT matrix then \(\mu_{\text{max}}^2 = \mu_{\text{min}}^2 = 1\). The quantity \(\mu_{\text{min}}^2\) will be useful to establish Theorem 3.3

while the main purpose of introducing \(\mu_{\text{max}}^2\) is to quantify a “joint incoherence pattern” on all \(B_i\). Namely, there is a common partition \(\{\Gamma_p\}_{p=1}^{P}\) of the index set \(\{1, \cdots, L\}\) with \(|\Gamma_p| = Q\) and \(L = PQ\) such that for each pair of \((i,p)\) with \(1 \leq i \leq r\) and \(1 \leq p \leq P\), we have

\[
\max_{1 \leq i \leq r, 1 \leq p \leq P} \left\| T_{i,p} - \frac{Q}{L} I_{K_i} \right\| \leq \frac{Q}{4L}, \quad \text{where} \quad T_{i,p} := \sum_{l \in \Gamma_p} b_{i,l} b_{i,l}^*,
\]

which says that each \(T_{i,p}\) does not deviate too much from \(I_{K_i}\). The key question here is whether such a common partition exists. It is hard to answer it in general. To the best of our knowledge, it is known that for each \(B_i\), there exists a partition \(\{\Gamma_{i,p}\}_{p=1}^{P}\) (where \(\Gamma_{i,p}\) depends on \(i\)) such that

\[
\max_{1 \leq p \leq P} \left\| \sum_{l \in \Gamma_{i,p}} b_{i,l} b_{i,l}^* - \frac{Q}{L} I_{K_i} \right\| \leq \frac{Q}{4L}, \quad \forall 1 \leq i \leq r,
\]

if \(Q \geq C \mu_{\text{max}}^2 K_i \log L\) where this argument is shown to be true in [2] by using Theorem 1.2 in [8]. Based on this observation, at least we have following several special cases which satisfy (10) for a common partition \(\{\Gamma_p\}_{p=1}^{P}\).

1. All \(B_i\) are the same. Then the common partition \(\{\Gamma_p\}_{p=1}^{P}\) can be chosen the same as \(\{\Gamma_{i,p}\}_{p=1}^{P}\) for any particular \(i\).
2. If each \(B_i, i \neq j\) is a submatrix of \(B_j\), then we can simply let \(\Gamma_p = \Gamma_{j,p}\) such that (10) holds.
3. If all \(B_i\) are “low-frequency” DFT matrices, i.e., the first \(K_i\) columns of an \(L \times L\) DFT matrix with \(B_i^* B_i = I_{K_i}\), we can actually create an explicit partition of \(\Gamma_p\) such that

\[
T_{i,p} = \sum_{l \in \Gamma_p} b_{i,l} b_{i,l}^* = \frac{Q}{L} I_{K_i},
\]

(11)

For example, suppose \(L = PQ\) and \(Q \geq K_i\), we can achieve \(T_{i,p} = \frac{Q}{L} I_{K_i}\), and \(|\Gamma_p| = Q\) by letting \(\Gamma_p = \{p, P + p, \cdots, (Q - 1)P + p\}\). A short proof will be provided in Section 7.2.

Some direct implications of (10) are

\[
\|T_{i,p}\| \leq \frac{5Q}{4L}, \quad \|S_{i,p}\| \leq \frac{4L}{3Q}, \quad \forall 1 \leq i \leq r, 1 \leq p \leq P.
\]

(12)

where \(S_{i,p} := T_{i,p}^{-1}\). Now let us introduce the second incoherence quantity, which is also crucial in the proof of Theorem 3.1

\[
\mu_h^2 := \max \left\{ \frac{Q^2}{L} \max_{l \in \Gamma_p, 1 \leq p \leq P, 1 \leq i \leq r} \frac{|\langle S_{i,p} h_i, b_{i,l} \rangle|^2}{\| b_{i,l} \|^2}, L \max_{1 \leq l \leq L, 1 \leq i \leq r} \frac{|\langle h_i, b_{i,l} \rangle|^2}{\| h_i \|^2} \right\}.
\]

(13)

The range of \(\mu_h^2\) is given in Proposition 2.2.

**Remark 2.1.** The attentive reader may have noticed that the definition of \(\mu_h^2\) is a bit more intricate than the one in [2], where \(\mu_h^2\) only depends on \(|\langle h_i, b_{i,l} \rangle|^2\). The reason is that we need to establish a result similar to Lemma 4 in [2], but the proof of Lemma 4 as stated is not entirely accurate, and a fairly simple way to fix this issue is to slightly modify the definition of \(\mu_h^2\).

Another easy way to fix the issue is to consider all \(B_i\) as low-frequency Fourier matrices. If so, \(\mu_h^2\) in (13) reduces to a simpler form of \(\mu_h^2\), i.e., \(\mu_h^2 = L \max_{1 \leq l \leq L} \| b_{i,l} \|^2 / \| h_i \|^2\) in [2] because the explicit partition of low-frequency DFT matrices allows \(T_{i,p} = \frac{Q}{L} I_{K_i}\) and \(S_{i,p} = \frac{1}{Q} I_{K_i}\).
Both $\mu_h^2$ and $\mu_{\max}^2$ measure the incoherence of $B_i$ and the latter one, depending $h_i$, also describes the interplay between $h_i$ and $B_i$. To sum up, for all $1 \leq l \leq L$ and $1 \leq i \leq r$,

$$\|b_{i,l}\|^2 \leq \frac{\mu_{\max}^2 K_i}{L}, \quad |\langle h_{i,l}, b_{i,l} \rangle|^2 \leq \frac{\mu_h^2}{L} \|h_i\|^2, \quad |\langle S_{i,p} h_{i,l}, b_{i,l} \rangle|^2 \leq \frac{L \mu_h^2}{Q^2 \|h_i\|^2}. \quad (14)$$

**Proposition 2.2.** Under the condition of (10) and (12),

$$1 \leq \mu_h^2 \leq \frac{16}{9} \mu_{\max}^2 K_i, \quad \forall 1 \leq i \leq r.$$

**Proof:** We start with (13) and (14) to find the lower bound of $\mu_h^2$ first. Without loss of generality, all $h_i$ are of unit norm. The definition of $\mu_h^2$ and $|\Gamma_p| = Q$ immediately imply that

$$\mu_h^2 \geq \max_{i,p} \left\{ \frac{Q}{L} \sum_{l \in \Gamma_p} |\langle S_{i,p} h_i, b_{i,l} \rangle|^2, \sum_{l = 1}^L |\langle h_i, b_{i,l} \rangle|^2 \right\}$$

$$= \max_{i,p} \left\{ \frac{Q}{L} \sum_{l \in \Gamma_p} h_i^* S_{i,p} b_{i,l} b_{i,l}^* S_{i,p} h_i, \sum_{l = 1}^L h_i^* b_{i,l} b_{i,l}^* h_i \right\}$$

$$= \max_{i,p} \left\{ \frac{Q}{L} h_i^* S_{i,p} h_i, 1 \right\}.$$

Note that

$$1 \leq \max_{i,p} \left\{ \frac{Q}{L} h_i^* S_{i,p} h_i, 1 \right\} \leq \frac{4}{3},$$

which follows from $\|S_{i,p}\| \leq \frac{4L}{3Q}$ and thus we can conclude the lower bound of $\mu_h^2$ is between 1 and $\frac{4}{3}$. We proceed to derive the range of the upper bound for $\mu_h^2$. Using Cauchy-Schwarz inequality gives

$$\mu_h^2 \leq \max \left\{ \frac{Q^2}{L} \max_{l \in \Gamma_p, 1 \leq i \leq L, 1 \leq i \leq r} |\langle S_{i,p} h_i, b_{i,l} \rangle|^2, L \max_{1 \leq i \leq L, 1 \leq i \leq r} |\langle h_i, b_{i,l} \rangle|^2 \right\}$$

$$\leq \max_{p,i,l} \left\{ \frac{Q^2}{L} \|S_{i,p}\|^2 \|b_{i,l}\|^2, L \|b_{i,l}\|^2 \right\}$$

$$\leq \frac{Q^2}{L} \frac{16L^2}{9Q^2} \frac{\mu_{\max}^2 K_i}{L} \leq \frac{16}{9} \mu_{\max}^2 K_i,$$

where $\|S_{i,p}\| \leq \frac{4L}{3Q}$ and $\|b_{i,l}\|^2 \leq \mu_{\max}^2 K_i$. □

### 2.4 Is the incoherence parameter $\mu_h^2$ necessary?

This subsection is devoted to a further discussion of the role of $\mu_h^2$. In order to provide a clearer explanation of the significance of $\mu_h^2$, we first reformulate the recovery of $\{X_i\}_{i=1}^r$ subject to (7) as a rank-$r$ matrix recovery problem. Each entry of $y$ is actually the inner product of two rank-$r$ block-diagonal matrices, i.e.,

$$y_i = \begin{bmatrix} h_{1,l} x_{1,l}^* & 0 & \cdots & 0 \\ 0 & h_{2,l} x_{2,l}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{r,l} x_{r,l}^* \end{bmatrix} \begin{bmatrix} \{b_{1,l} a_{1,l}^* \} & 0 & \cdots & 0 \\ 0 & \{b_{2,l} a_{2,l}^* \} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \{b_{r,l} a_{r,l}^* \} \end{bmatrix}.$$ 

Recall that in matrix completion [35, 30] the left and right singular vectors of the true matrix cannot be too aligned with those of the test matrix. A similar spirit applies to this problem as well, i.e., both

$$\max_{1 \leq i \leq L, 1 \leq i \leq r} L |\langle b_{i,l}, h_i \rangle|^2 / \|h_i\|^2, \quad \max_{1 \leq i \leq L, 1 \leq i \leq r} |\langle a_{i,l}, x_i \rangle|^2 / \|x_i\|^2 \quad (15)$$
are required to be small. We can ensure that the second term in \([15]\) is small since each \(a_{i,l}\) is a Gaussian random vector and randomness contributes a lot to making the quantity small (with high probability). However, the first term is deterministic and could in principle be very large for certain \(h_i\) (more precisely, the worst case could be \(O(K)\)), hence we need to put a constraint on \(\mu^2_h\) in order to control its size. As numerical simulations presented in Section \(\[4\]\) show, the relevance of \(\mu^2_h\) goes beyond “proof-technical reasons”. The required number of measurements for successful recovery does indeed depend on \(\mu^2_h\), see Figure \(\[3\]\) at least when using the suggested approach via semidefinite programing.

### 2.5 Conditions on the matrices \(A_i\)

Throughout the proof of main theorem, we also need to be able to control a certain “mutual incoherence” of the matrices \(A_i\) on the subspaces \(T_i\), cf. \(\[27\]\). This condition involves the quantity

\[
\max_{j \neq k} \| P_{T_j} A_j^* A_k P_{T_k} \|. \tag{16}
\]

This quantity is formulated in terms of the matrices \(A_i\) (and not the \(A_i^*\)), but in order to get a grip on this quantity, it will be convenient and necessary to impose some conditions on the matrices \(A_i\). For instance we may assume that the \(A_i\) are i.i.d. Gaussian random matrices, which we will do henceforth. Thus, we require that the \(l\)-th column of \(A_i^*\), \(a_{i,l} \sim \mathcal{N}(0, I_N)\), i.e., \(a_{i,l}\) is an \(N \times 1\) standard real Gaussian random vector. In that case the expectation of \(A_i^* A_i(Z_i) = \sum_{l=1}^L b_{i,l} b_{i,l}^* Z_i a_{i,l} a_{i,l}^*\) can be computed

\[
\mathbb{E}(A_i^* A_i(Z_i)) = \sum_{l=1}^L b_{i,l} b_{i,l}^* Z_i \mathbb{E}(a_{i,l} a_{i,l}^*) = Z_i, \quad Z_i \in \mathbb{C}^{K_i \times N_i},
\]

which says that the expectation of \(A_i^* A_i\) is the identity. In the proof, we also need to examine \(A_{i,p}^* A_{i,p}\). Considering the common partition \(\{G_p\}_{p=1}^P\) satisfying \(\[10\]\), we define \(A_{i,p} : \mathbb{C}^{K_i \times N_i} \to \mathbb{C}^{Q}\) and \(A_{i,p}^* : \mathbb{C}^{Q} \to \mathbb{C}^{K_i \times N_i}\) correspondingly by

\[
A_{i,p}(Z_i) = \{b_{i,l}^* Z_i a_{i,l}\}_{l \in G_p}, \quad A_{i,p}^*(z) = \sum_{l \in G_p} z_i b_{i,l} a_{i,l}^*.
\]

The definition of \(A_{i,p}\) is the same as that of \(A_i\) except that \(A_{i,p}\) only uses a subset of all measurements. However, the expectation of \(A_{i,p}^* A_{i,p}\) is no longer the identity in general (except the case when all \(B_i\) are low-frequency DFT matrices and satisfy \(\[11\]\)), i.e.,

\[
A_{i,p}^* A_{i,p}(Z_i) = \sum_{l \in G_p} b_{i,l} b_{i,l}^* Z_i a_{i,l} a_{i,l}^*,
\]

and

\[
\mathbb{E}(A_{i,p}^* A_{i,p}(Z_i)) = T_{i,p} Z_i, \quad T_{i,p} := \sum_{l \in G_p} b_{i,l} b_{i,l}^*.
\]

The non-identity expectation of \(A_{i,p}^* A_{i,p}\) will create challenges throughout the proof. However, there is an easy trick to fix this issue. By properly assuming \(Q > K_i\), \(T_{i,p}\) is actually invertible. Consider \(A_{i,p}^* A_{i,p}(S_{i,p} Z_i)\) and its expectation now yields

\[
\mathbb{E}(A_{i,p}^* A_{i,p}(S_{i,p} Z_i)) = T_{i,p} S_{i,p} Z_i = Z_i, \quad S_{i,p} := T_{i,p}^{-1}.
\]

This trick, i.e., making the expectation of \(A_{i,p}^* A_{i,p} S_{i,p}\) equal to the identity, plays an important role in the proof.
3 Main Results

3.1 The noiseless case

Our main finding is that solving \((8)\) enables demixing and blind deconvolution simultaneously. Moreover, our method is also robust to noise.

**Theorem 3.1.** Consider the model in \((4)\) and assume that each \(B_i \in \mathbb{C}^{L \times K_i}\) with \(B_i^*B_i = I_{K_i}\) and each \(A_i\) is a Gaussian random matrix, i.e., each entry in \(A_i\) is \(i.i.d\. \sim \mathcal{N}(0,1)\). Let \(\mu_\text{max}^2\) and \(\mu_h^2\) be as defined in \((9)\) and \((13)\) respectively, and denote \(K := \max_{1 \leq i \leq r} K_i\) and \(N := \max_{1 \leq i \leq r} N_i\).

If

\[
L \geq C_\alpha r^2 \max\{\mu_\text{max}^2 K, \mu_h^2 N\} \log^2 L \log \gamma (r + 1),
\]

where \(\gamma\) is the upper bound of \(\|A_i\|\) and yields \(\gamma \leq \sqrt{N \log(NL/2)} + \alpha \log L\), then solving \((8)\) recovers \(\{X_i = h_i x_i^*, 1 \leq i \leq r\}\) exactly with probability at least \(1 - O(L^{-\alpha+1})\).

Even though the proof of Theorem 3.1 follows a meanwhile well established route, the details of the proof itself are nevertheless quite involved and technical. Hence, for convenience we give a brief overview of the proof architecture. In Section 5.1 we derive a sufficient condition and an approximate dual certificate condition for the minimizer of \((8)\) to be the unique solution to \((4)\). These conditions stipulate that the matrices \(A_i\) need to satisfy two key properties. The first property, proved in Section 5.2, can be considered as a modification of the celebrated Restricted Isometry Property (RIP) \([9]\), as it requires the \(A_i\) to act in a certain sense as “local” approximate isometries \([11,10]\). The second property, proved in Section 5.3, requires the two operators \(A_i\) and \(A_j\) to satisfy a “local” mutual incoherence property. With these two key properties in place, we can now construct an approximate dual certificate that fulfills the conditions derived in Section 5.1. We use the golfing scheme \([19]\) for this purpose, the constructing of which can be found in Section 5.4. With all these tools in place, we assemble the proof of Theorem 3.1 in Section 5.5.

The theorem assumes for convenience that the \(h_i\) and the \(x_i\) are real-valued, but it is easy to modify the proof for complex-valued \(h_i\) and \(x_i\). We leave this modification to the reader. While Theorem 1.1 is the first of its kind, the derived condition on the number of measurements in \((2)\) is not optimal. Numerical experiments suggest (see e.g. Figure 1 in Section 4), that the number of measurements required for a successful solution of the blind deconvolution-blind demixing problem scales with \(r\) and not with \(r^2\). Indeed, the simulations indicate that successful recovery via semidefinite programming is possible with a number of measurements close to the theoretical minimum, i.e., with \(L \gtrsim r(K + N)\), see Section 4. This is a good news from a viewpoint of application and means that there is room for improvement in our theory. Nevertheless, this brings up the question whether we can improve upon our bound. A closer inspection of the proof shows that the \(r^2\)-bottleneck comes from the requirement \(\max_{j \neq k} \|P_{T_j} A_j^* A_k P_{T_k}\| \leq \frac{1}{r}\), see condition \((28)\). In order to achieve this we need that \(L\), the number of measurements, scales essentially like \(r^2 \max\{\mu_\text{max}^2 K, \mu_h^2 N\} \) (up to log-factors), see Section 5.3. Is it possible, perhaps with a different condition that does not rely on mutual incoherence between the \(A_i\), to reduce this requirement on \(L\) to one that scales like \(r \max\{\mu_\text{max}^2 K, \mu_h^2 N\}^2\)?

Now we take a little detour to revisit the blind deconvolution problem described in the introduction and in the informal Theorem 1.1 which is actually contained in our proposed framework as a special case. Recall the model in \((4)\) that \(y\) is actually the sum of Hadamard products of \(B_i h_i\) and \(A_i x_i\). Let \(F\) be the Discrete Fourier Transform matrix of size \(L \times L\) with \(F^*F = I_L\) and let the \(L \times K_i\) matrix \(B_i\) consist of the first \(K_i\) columns of \(F\) (then \(B_i^*B_i = I_{K_i}\)). Now we can express \((4)\) equivalently as the sum of circular convolutions of \(F^{-1}(B_i h_i)\) and \(F^{-1}(A_i x_i)\), i.e.,

\[
\frac{1}{\sqrt{L}} \sum_{i=1}^{r} F^{-1}(B_i h_i) * F^{-1}(A_i x_i) = \sum_{i=1}^{r} (F^{-1}B_i) h_i * (F^{-1}A_i) x_i.
\]
Set
\[ g_i := \begin{bmatrix} h_i \\ 0_{L-K_i} \end{bmatrix}. \]

Then there holds
\[ F^{-1}B_i h_i = F^{-1} \begin{bmatrix} B_i & 0_{L \times (L-K_i)} \end{bmatrix} \begin{bmatrix} h_i \\ 0_{L-K_i} \end{bmatrix} = g_i. \]

Hence with a slight abuse of notation (replacing \( \frac{1}{\sqrt{L}} F^{-1}y \) in (20) by \( y \) and \( F^{-1}A_i \) by \( A_i \), using the fact that the Fourier transform of a Gaussian random matrix is again a Gaussian random matrix), we can express (4) equivalently as
\[ y = \sum_{i=1}^{r} g_i * (A_i x_i), \]
which is exactly (1) up to a normalization factor.

Thus we can easily derive the following corollary from Theorem 3.1 (using the fact that \( \mu_{\text{max}} = 1 \) for the particular choice of \( B_i \) above). This corollary is the precise version of the informal Theorem 1.1.

**Corollary 3.2.** Consider the model in (4), i.e.,
\[ y = \sum_{i=1}^{r} g_i * (A_i x_i), \]
where we assume that \( g_i(k) = 0 \) for \( k > K_i \). Suppose that each \( A_i \) is a Gaussian random matrix, i.e., each entry in \( A_i \) \( \overset{i.i.d.}{\sim} N(0,1) \). Let \( \mu_h^2 \) be as defined in (13). If
\[ L \geq C_\alpha r^2 \max\{K, \mu_h^2 N\} \log^2 L \log \gamma \log(r+1), \]
where \( \max_i \|A_i\| \leq \gamma \leq \sqrt{N \log(NL/2) + \alpha \log L} \), then solving (8) recovers \( \{X_i := h_i x_i^*, 1 \leq i \leq r\} \) exactly with probability at least \( 1 - O(L^{-\alpha+1}) \).

For the special case \( r = 1 \), Corollary 3.2 becomes Theorem 1 in [2] (with the proviso that in principle our \( \mu_h^2 \) is defined slightly differently than in [2], see Remark 2.1. Yet, if we choose the partition of the matrix \( B \) as suggested in the third example in Subsection 2.3, then the difference between the two definitions of \( \mu_h^2 \) vanishes.).

### 3.2 Noisy data

In reality, measurements are always noisy. Hence, suppose \( \hat{y} = y + \epsilon \) where \( \epsilon \) is noise with \( \|\epsilon\| \leq \eta \). In this case we solve the following optimization program to recover \( \{X_i\}_{i=1}^{r} \),
\[ \min \sum_{i=1}^{r} \|Z_i\|_* \text{ subject to } \sum_{i=1}^{r} A_i(Z_i) - \hat{y} \leq \eta. \]

(21)

We should choose \( \eta \) properly in order to make \( X_i \) inside the feasible set and \( \|\hat{y}\| > \eta \). Let \( \{X_i\}_{i=1}^{r} \) be the minimizer to (21). We immediately know
\[ \sum_{i=1}^{r} \|X_i\|_* \leq \sum_{i=1}^{r} \|X_i\|_. \]

(22)

Our goal is to see how \( \sqrt{\sum_{i=1}^{r} \|X_i - X_i\|^2_F} \) varies with respect to the noise level \( \eta \).
We denote \(A\) for them yet. In wireless communication, it is particularly interesting to see the recovery

Let \(B\) are fixed. Each \(\epsilon\) where applying SVD to \(\hat{X}_i\) where each \(\hat{X}_i\) varies from 1 to 7 and \(L\) sources. Here are the parameters and settings used in the simulations: the number of sources \(L\) We investigate empirically the minimal \(L\) such that \(\hat{X}_i\) with probability at least \(1 - \mathcal{O}(L^{-\alpha+1})\). Here, \(\lambda_{\max}\) and \(\lambda_{\min}\) are the largest and the smallest eigenvalue of \(\sum_{i=1}^r A_i A_i^*\), respectively.

Note that with a little modification of Lemma 2 in \([2]\), it can be shown that \(\lambda_{\max} \sim \overline{\mu}\) The proof of Theorem 3.3 will be given in Section 5.6.

With Theorem 3.3 and Wedin’s \(\sin(\theta)\) theorem \([43, 45]\) for singular value/vector perturbation theory, we immediately have the performance guarantees of recovering individual \((h_i, x_i)_{r=1}^r\) by applying SVD to \(\hat{X}_i\).

**Corollary 3.4.** Let \(\hat{h}_i = \sqrt{\sigma_{i1}} \hat{u}_{i1}\) and \(\hat{x}_i = \sqrt{\sigma_{i1}} \hat{v}_{i1}\) where \(\sigma_{i1}\), \(\hat{u}_{i1}\) and \(\hat{v}_{i1}\) are the leading singular value, left and right singular vectors of \(\hat{X}_i\) respectively. Then there exist \(\epsilon_i\) and a constant \(c_0\) such that

\[
\|h_i - c_i \hat{h}_i\| \leq c_0 \min(\epsilon/\|h_i\|, \|h_i\|), \quad \|x_i - c_i^{-1} \hat{x}_i\| \leq c_0 \min(\epsilon/\|x_i\|, \|x_i\|)
\]

where \(\epsilon = \sqrt{\sum_{i=1}^r \|\hat{X}_i - X_i\|^2_F}\).

### 4 Numerical Simulations

In this section we present a range of simulations that illustrate and complement the theoretical results of the previous section.

#### 4.1 Number of measurements \(L\) vs. number of sources \(r, K_i\) and \(N_i\)

We investigate empirically the minimal \(L\) required to simultaneously demix and deconvolve \(r\) sources. Here are the parameters and settings used in the simulations: the number of sources \(r\) varies from 1 to 7 and \(L = 50, 100, \cdots, 750\) and 800. For each \(1 \leq i \leq r\), \(K_i = 30\) and \(N_i = 25\) are fixed. Each \(B_i\) is the first \(K_i\) columns of an \(L \times L\) DFT matrices with \(B_i^* B_i = I_{K_i}\) and each \(A_i\) is an \(L \times N_i\) Gaussian random matrix. \(h_i\) and \(x_i\) yield \(N(0, I_{K_i})\) and \(N(0, I_{N_i})\) respectively. We denote \(X_i = h_i x_i^*\), the “lifted” matrix and solve \([8]\) to recover \(X_i\). For each pair of \((L, r)\), 10 experiments are performed and the recovery is regarded as a success if

\[
\frac{\sqrt{\sum_{i=1}^r \|X_i - \hat{X}_i\|^2_F}}{\sqrt{\sum_{i=1}^r \|X_i\|^2_F}} < 10^{-3}
\]

(23)

where each \(\hat{X}_i\), given by solving \([8]\) via the CVX package \([18]\), serves as an approximation of \(X_i\). Theorem 3.1 implies that the minimal required \(L\) scales with \(r^2\), which is not optimal in terms of number of degrees of freedom. Figure 1 validates the non-optimality of our theory. Figure 1 shows a sharp phase transition boundary between success and failure and furthermore the minimal \(L\) for exact recovery seems to have a strongly linear correlation with number of sources \(r\). Note that if \(L\) is approximately greater than \(80r\), solving \([8]\) gives the exact recovery of \(X_i\) numerically, which is quite close to the theoretical limit \((K_i + N_i)r = 55r\).

Moreover, our method extends to other types of settings although we do not have theories for them yet. In wireless communication, it is particularly interesting to see the recovery
In order to illustrate the robustness of our algorithm with respect to noise as stated in Theorem 3.3, we conduct two simulations to study how the relative error \( \frac{\sqrt{\sum_{i=1}^{r} \| X_i - \hat{X}_i \|_F^2}}{\sqrt{\sum_{i=1}^{r} \| X_i \|_F^2}} \) behaves under different levels of noise. In the first experiment we choose \( r = 3 \), i.e., there are totally 3 sources. They are of different sizes, i.e., \((K_1, N_1) = (20, 20), (K_2, N_2) = (25, 25)\) and \((K_3, N_3) = (20, 20)\). \( L \) is fixed to be 256, the \( B_i \)s are as outlined in Section 4.1 and the \( A_i \)s are Gaussian random matrices. In the simulation, we choose \( \epsilon_i \) to be a normalized Gaussian random vector. Namely, we first sample \( \epsilon_i \) from a multivariate Gaussian distribution and then normalize \( \| \epsilon_i \|_F = \sigma \sqrt{\sum_{i=1}^{r} \| X_i \|_F^2} \) where \( \sigma = 1, 0.5, 0.1, 0.05, 0.01, \ldots \) and 0.0001. For each \( \sigma \), we run 10 experiments and compute the average relative error in the scale of dB, i.e., \( 10 \log_{10}(\text{Avg.RelErr}) \).

We run a similar experiment, this time with \( r = 15 \) sources (all \( N_i \)s are equal to 10, and all \( K_i \)s are equal to 15) and the \( A_i \)s are the “random” Hadamard matrices described above. For both experiments, Figure 4 indicates that the average relative error (dB) is linearly correlated with \( \text{SNR} = 10 \log_{10}(\sum_{i=1}^{r} \| X_i \|_F^2/\| \epsilon_i \|_F^2) \), as one would wish.
5 Proofs

This section is dedicated to proving the theorems presented in Section 3. Since the proofs are rather involved and technical, we have arranged the arguments into individual subsections. We first start out with stating sufficient conditions under which the main theorems will hold, see Subsection 5.1. In Subsection 5.2 we state and analyze a certain form of local restricted isometry property and a specific local incoherence property is established in Subsection 5.3. Both of them are associated with the assumptions of the sufficient conditions in Subsection 5.1. We construct
5.1 Sufficient conditions

Without loss of generality, we assume that the lifted matrix \( X_i = \alpha_i h_i x_i^* \), where \( h_i \) and \( x_i \) are all real and of unit norm and \( \alpha_i \) is a scalar for all \( 1 \leq i \leq r \) throughout Section 5.1–5.6. We also define a linear space which \( h_i x_i^* \) lies in and which will be useful in the further analysis:

\[
T_i = \{ h_i h_i^* Z_i + (I_{K_i} - h_i h_i^*) Z_i x_i x_i^* | Z_i \in \mathbb{C}^{K_i \times N_i} \} \tag{24}
\]

and its corresponding complement is defined as

\[
T_i^\perp = \{ (I_{K_i} - h_i h_i^*) Z_i (I_{N_i} - x_i x_i^*) | Z_i \in \mathbb{C}^{K_i \times N_i} \}. \tag{25}
\]
Figure 4: Performance of (21) under different SNR. Left: $\{A_i\}$ are Gaussian and there are 3 sources and $L = 256$; Right: $A_i = D_i H_i$ where $H_i$ is a partial Hadamard matrix and $D_i$ is a diagonal matrix with random $\pm 1$ entries. Here there are 15 sources in total and $L = 512$.

Now we present the first sufficient condition, under which $\{\alpha_i h_i x_i^*\}_{i=1}^r$ is the unique minimizer. However, Lemma 5.1 is not easy to use in reality and therefore, we derive Lemma 5.2, a more useful condition, from Lemma 5.1.

**Lemma 5.1.** Assume that
\[ \sum_{i=1}^r \langle H_i, h_i x_i^* \rangle + \| H_i, T_i^\perp \|_s > 0. \]  
(26)
for any real $\{H_i\}_{i=1}^r$ satisfying $\sum_{i=1}^r A_i(H_i) = 0$ and at least of $H_i$ is nonzero. Then $\{\alpha_i h_i x_i^*\}_{i=1}^r$ is the unique minimizer to the convex program (8).

**Proof:** For any feasible element of the convex program (8), it must have the form of $\{\alpha_i h_i x_i^* + H_i\}_{i=1}^r$. It suffices to show that the $\sum_{i=1}^r \|\alpha_i h_i x_i^* + H_i\|_s > \sum_{i=1}^r \|\alpha_i h_i x_i^*\|_s$ for any nontrivial set of $\{H_i\}_{i=1}^r$, i.e., at least one of $H_i$ is nonzero. For each $H_i$, there exists a $V_i \in T_i^\perp$ such that
\[ \langle H_i, V_i \rangle = \langle H_i, T_i^\perp, V_i \rangle = \| H_i, T_i^\perp \|_s. \]
where $H_i, T_i^\perp$ is the projection of $H_i$ on $T_i^\perp$ and $\|V_i\| = 1$. Thus $h_i x_i^* + V_i$ belongs to the subdifferential of $\|\cdot\|_s$ at $X_i = \alpha_i h_i x_i^*$.

\[ \sum_{i=1}^r \|\alpha_i h_i x_i^* + H_i\|_s \geq \sum_{i=1}^r \|\alpha_i h_i x_i^*\|_s + \langle h_i x_i^* + V_i, H_i \rangle \]
\[ = \sum_{i=1}^r \|\alpha_i h_i x_i^*\|_s + \langle H_i, T_i^\perp, h_i x_i^* \rangle + \| H_i, T_i^\perp \|_s \]
\[ > \sum_{i=1}^r \|\alpha_i h_i x_i^*\|_s \]
where the first inequality follows from the definition of subgradient and (26). □

Now we consider under what condition on $A_i$, the unique minimizer is $\{\alpha_i h_i x_i^*\}$. Define $\mu$ by
\[ \mu := \max_{j \neq k} \|P_{T_j} A_j^* A_k P_{T_k}\| \]  
(27)
as a measure of incoherence between any pairs of linear operators. \( A_i \mathcal{P}_{T_i} \) is the restriction of \( A_i \) onto \( T_i \).

**Lemma 5.2.** Assume that

\[
\| \mathcal{P}_{T_i} A_i^* A_i \mathcal{P}_{T_i} - \mathcal{P}_{T_i} \| \leq \frac{1}{4}, \quad \mu \leq \frac{1}{4r}, \quad \| A_i \| \leq \gamma
\]  

(28)

and also there exists a \( \lambda \in \mathbb{C}^L \) such that

\[
\| h_i x_i^* - \mathcal{P}_{T_i} (A_i^* (\lambda)) \|_F \leq \alpha, \quad \| \mathcal{P}_{T_i^+} (A_i^* (\lambda)) \| \leq \beta
\]  

(29)

for all \( 1 \leq i \leq r \) and \( (1 - \beta) - 2r \gamma \alpha > 0 \), then \( \{ \alpha_i h_i x_i^* \}_{i=1}^r \) is the unique minimizer to (8). In particular, we can choose \( \alpha = (5r \gamma)^{-1} \) and \( \beta = \frac{1}{2} \). Here \( \| A_i \| := \sup_{Z \neq 0} \| A_i (Z) \|_F / \| Z \|_F \).

**Proof:** It suffices to show that for any nonzero \( \{ H_i \}_{i=1}^r \) with \( \sum_{i=1}^r A_i (H_i) = 0 \), there holds

\[
I_1 := \sum_{i=1}^r (H_i, h_i x_i^* - A_i^* (\lambda)) + \| H_i, T_i^+ \|_* > 0
\]

under (28) and (29). By decomposing the inner product in \( I_1 \) on \( T_i \) and \( T_i^+ \) for each \( i \), we have

\[
I_1 = \sum_{i=1}^r (H_i, T_i, h_i x_i^* - \mathcal{P}_{T_i} (A_i^* (\lambda))) - \left( H_i, T_i^+ , \mathcal{P}_{T_i^+} (A_i^* (\lambda)) \right) + \| H_i, T_i^+ \|_*.
\]

Then, by applying the Cauchy-Schwarz inequality and \( |(U, V)| \leq \|U\|_F \|V\| \) for any matrices \( U, V \) of the same size, we only need to show that a lower bound of \( I_1 \) is positive, i.e., \( I_1 \geq I_2 > 0 \):

\[
I_2 := \sum_{i=1}^r -\| H_i, T_i \|_F \| h_i x_i^* - \mathcal{P}_{T_i} (A_i^* (\lambda)) \|_F + \| H_i, T_i^+ \|_* \left( 1 - \| \mathcal{P}_{T_i^+} (A_i^* (\lambda)) \| \right) > 0.
\]  

(30)

From now on, we aim to show

\[
\frac{1}{2} \left( \sum_{i=1}^r \| H_i, T_i \|_F \right)^2 \leq \gamma \left( \sum_{i=1}^r \| H_i, T_i^+ \|_F \right) \leq \gamma \left( \sum_{i=1}^r \| H_i, T_i^+ \|_* \right)
\]

in order to achieve (30). We start with \( \sum_{i=1}^r A_i (H_i) = 0 \). Note that \( H_i = H_{i, T_i} + H_{i, T_i^+} \), there holds

\[
\| \sum_{i=1}^r A_i (H_{i, T_i}) \|_F = \| \sum_{i=1}^r A_i (H_{i, T_i^+}) \|_F.
\]

It is easy to bound the quantity on the right hand side by \( \| A_i \| \leq \gamma \),

\[
\| \sum_{i=1}^r A_i (H_{i, T_i^+}) \|_F \leq \gamma \left( \sum_{i=1}^r \| H_{i, T_i^+} \|_F \right).
\]  

(31)

The difficulty is to establish the lower bound.

\[
\left\| \sum_{i=1}^r A_i (H_{i, T_i}) \right\|^2_F \geq \sum_{i=1}^r \| A_i (H_{i, T_i}) \|^2 + 2 \sum_{j \neq k} \langle A_j (H_{j, T_j}), A_k (H_{k, T_k}) \rangle
\]

\[
\geq \frac{3}{4} \sum_{i=1}^r \| H_{i, T_i} \|^2_F - 2 \mu \sum_{j \neq k} \| H_{j, T_j} \|_F \| H_{k, T_k} \|_F
\]  

(32)

\[
= \left[ \| H_{1, T_1} \|_F \right]^* \left[ \begin{array}{cccc} \frac{3}{4} & -\mu & \cdots & -\mu \\ -\mu & \frac{3}{4} & \cdots & -\mu \\ \vdots & \vdots & \ddots & \vdots \\ -\mu & -\mu & \cdots & \frac{3}{4} \end{array} \right]_{r \times r} \left[ \begin{array}{c} \| H_{1, T_1} \|_F \\ \| H_{r, T_r} \|_F \end{array} \right].
\]
where (32) follows from (28). It is easy to see that the coefficient matrix inside the quadratic form has its smallest eigenvalue $\frac{3}{4} - (r - 1)\mu \geq \frac{3}{4} - \frac{r-1}{16r} > \frac{1}{2}$ and all the other eigenvalues are \( \frac{3}{4} \). Hence

$$\left\| \sum_{i=1}^{r} A_i(H_{i,T_i}) \right\|_F \geq \frac{1}{2} \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F^2 \geq \frac{1}{2r} \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F.$$

Combining (33) with (51) leads to

$$\frac{1}{2r} \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F \leq \gamma \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F. \tag{34}$$

I₂ in (30) has its lower bound as follows:

$$I_2 \geq \sum_{i=1}^{r} -\left\| H_{i,T_i} \right\|_F \left\| h_i x_i^* - \mathcal{P}_T(A_i^*(\lambda)) \right\|_F + \left\| H_{i,T_i\perp} \right\|_F \left( 1 - \left\| \mathcal{P}_T(\lambda) \right\| \right) \tag{35}$$

$$\geq -\alpha \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F + (1 - \beta) \sum_{i=1}^{r} \left\| H_{i,T_i\perp} \right\|_F \tag{36}$$

$$\geq -2r\gamma\alpha \sum_{i=1}^{r} \left\| H_{i,T_i} \right\|_F + (1 - \beta) \sum_{i=1}^{r} \left\| H_{i,T_i\perp} \right\|_F \tag{37}$$

$$\geq (-2r\gamma\alpha + (1 - \beta)) \sum_{i=1}^{r} \left\| H_{i,T_i\perp} \right\|_F \geq 0,$$

where (35) uses $\left\| \cdot \right\| \geq \left\| \cdot \right\|_F$, (36) follows from the assumption (29), and (37) follows from (34). Under the condition $-2r\gamma\alpha + (1 - \beta) > 0$, (30) holds if at least one of the terms $\left\| H_{i,T_i\perp} \right\|_F$ is nonzero. If $H_{i,T_i\perp} = 0$ for all $1 \leq i \leq r$, then $H_i = 0$ via (34). ■

### 5.2 Local Isometry Property

Our goal in this subsection is to prove that the first assumption in (28) of Lemma 5.2 holds with high probability if $L$ is large enough. Instead of studying $\left\| \mathcal{P}_T A_i^* A_i \mathcal{P}_T - \mathcal{P}_T \right\|_F$ directly, we will focus on the more general expression $\left\| \mathcal{P}_T A_i^* A_i S_i S_i \mathcal{P}_T - \mathcal{P}_T \right\|_F$, where $A_i S_i$ and $S_i S_i$ are defined in (17) and (19) respectively.

#### 5.2.1 An explicit formula for $\mathcal{P}_T A_i^* A_i S_i S_i \mathcal{P}_T$

For each fixed pair of $(i, p)$ where $1 \leq i \leq r$ and $1 \leq p \leq P$, the proof of $\left\| \mathcal{P}_T A_i^* A_i S_i S_i \mathcal{P}_T - \mathcal{P}_T \right\|_F \leq \frac{1}{2}$ is actually the same. Therefore, for simplicity of notation, we omit the subscript $i$ and denote $\mathcal{P}_T A_i^* A_i S_i S_i \mathcal{P}_T$ by $\mathcal{P}_T A_i^* A_i S_i S_i \mathcal{P}_T$ throughout the proof of Proposition 5.3. By definition, $A_i S_i \mathcal{P}_T(Z) = \{b_i^* S_i \mathcal{P}_T(Z) a_l \}_{l \in \Gamma_p}$ for any $Z \in \mathbb{C}^{K \times N}$. Using (24) gives us an explicit expression of $b_i^* S_i \mathcal{P}_T(Z) a_l$, i.e.,

$$b_i^* S_i \mathcal{P}_T(Z) a_l = b_i^* S_i [h h^* Z + (I_K - h h^*) Z x x^*] a_l = \langle S_i h, b_i \rangle h^* Z a_l + \langle a_l, x \rangle b_i^* S_i (I_k - h h^*) Z x = h^* Z v_l + u_i^* Z x, \quad l \in \Gamma_p,$$

where $\mathcal{P}_T(Z) = h h^* Z + (I_K - h h^*) Z x x^*$ and both $h$ and $x$ are assumed to be real and of unit norm. Similarly,

$$b_i^* \mathcal{P}_T(Z) a_l = h^* Z v_l + u_i^* Z x, \quad l \in \Gamma_p,$$
where
\[ v_t := \langle h, b_i \rangle a_t, \]
\[ u_t := \langle a_t, x \rangle (I_K - hh^*) b_t, \]
\[ \tilde{v}_t := \langle S_p h, b_i \rangle a_t, \]
\[ \tilde{u}_t := \langle a_t, x \rangle (I_K - hh^*) S_p b_t. \]  (38)

Now we have
\[ A \mathcal{P}_T(S_p Z) = \{ \langle Z, hh^* \tilde{v}_t + \tilde{u}_t x^* \rangle \}_{t \in \Gamma_p}, \]
\[ \mathcal{P}_T A^*(z) = \sum_{t \in \Gamma_p} z_t \langle hv_t^* + u_t x^* \rangle. \]

By combining the terms we arrive at
\[ \mathcal{P}_T A^*_p A_p S_p \mathcal{P}_T (Z) = \sum_{t \in \Gamma_p} [hh^* Z \tilde{v}_t v_t^* + h \tilde{u}_t^* Z x v_t^* + u_t h^* Z \tilde{v}_t x^* + u_t \tilde{u}_t^* Z xx^*]. \]  (39)

The explicit form of each component in this summation is
\[ hh^* Z \tilde{v}_t v_t^* = \langle h, b_i \rangle \langle S_p h, b_t \rangle hh^* Z a_t a_t^*, \]
\[ h \tilde{u}_t^* Z x v_t^* = \langle h, b_i \rangle \langle h, b_t \rangle S_p (I_K - hh^*) Z xx^* a_t a_t^*, \]
\[ u_t h^* Z \tilde{v}_t x^* = \langle S_p h, b_i \rangle (I_K - hh^*) b_t h^* Z a_t a_t^* x x^*, \]
\[ u_t \tilde{u}_t^* Z xx^* = |\langle a_t, x \rangle|^2 (I_K - hh^*) b_t b_t^* S_p (I_K - hh^*) Z xx^*. \]

It is easy to compute the expectation of those random matrices by using \( \mathbb{E}(a_t a_t^*) = I_N \) and \( \mathbb{E}(|a_t, x|^2) = \|x\|^2 = 1 \). Our goal here is to estimate the operator norm of \( \mathcal{P}_T A^*_p A_p S_p \mathcal{P}_T - \mathcal{P}_T \) which is the sum of four components, i.e.,
\[ \mathcal{P}_T A^*_p A_p S_p \mathcal{P}_T - \mathcal{P}_T = \sum_{s=1}^4 \mathcal{M}_s \]

where each \( \mathcal{M}_i \) is a random linear operator with zero mean. More precisely, each of \( \mathcal{M}_s \) is given by
\[ \mathcal{M}_1(Z) := \sum_{t \in \Gamma_p} \langle h, b_i \rangle \langle S_p h, b_t \rangle hh^* Z (a_t a_t^* - I_N), \]
\[ \mathcal{M}_2(Z) := \sum_{t \in \Gamma_p} \langle h, b_i \rangle \langle h, b_t \rangle S_p (I_K - hh^*) Z xx^* (a_t a_t^* - I_N), \]
\[ \mathcal{M}_3(Z) := \sum_{t \in \Gamma_p} \langle S_p h, b_i \rangle (I_K - hh^*) b_i h^* Z (a_t a_t^* - I_N) x x^*, \]  (40)
\[ \mathcal{M}_4(Z) := \sum_{t \in \Gamma_p} (|\langle a_t, x \rangle|^2 - 1) (I_K - hh^*) b_i b_i^* S_p (I_K - hh^*) Z xx^*. \]

Each \( \mathcal{M}_s \) can be treated as a \( KN \times KN \) matrix because it is a linear operator from \( \mathbb{C}^{KN} \) to \( \mathbb{C}^{KN} \).

**Proposition 5.3.** Under the assumption of \[ (14) \] and \[ (10) \] and that \( \{a_{i,t}\} \) are standard Gaussian random vectors of length \( N_i \),
\[ \| \mathcal{P}_T A^*_p A_t S_p \mathcal{P}_T - \mathcal{P}_T \| \leq \frac{1}{4}, \quad 1 \leq i \leq r, 1 \leq p \leq P \]  (41)
holds simultaneously with probability at least \( 1 - L^{-\alpha + 1} \) if \( Q \geq C_\alpha \max \{ \mu_{max}^2 K, \mu_{max}^2 N \} \log^2 L \log (r + 1) \) where \( K := \max K_i \) and \( N := \max N_i \).
The following corollary, which is a special case of Proposition 5.3 (simply set $Q = L$ and $S_{i,p} = I_{K_i}$), indicates the first condition in (28) holds with high probability.

**Corollary 5.4.** Under the assumption of (14) and (10) and that $\{a_{i,l}\}$ are standard Gaussian random vectors of length $N_i$,

\[
\|P_{T_i}A_i^*A_iP_{T_i} - P_{T_i}\| \leq \frac{1}{4}, \quad 1 \leq i \leq r
\]

holds with probability at least $1 - L^{-\alpha + 1}$ if $L \geq C_\alpha \max\{\mu_{\max}^2, \mu_h^2\}N^2 \log^2 L \log(r + 1)$ where $K := \max K_i$ and $N := \max N_i$.

**Remark 5.5.** Although Proposition 5.3 and Corollary 5.4 are quite similar to Lemma 3 in [12] at the first glance, we include $S_{i,p}$ and the new definition of $\mu_h^2$ in our result. The purpose is to resolve the issue mentioned in Remark 2.1 by making $E(P_{T_i}A_i^*A_iS_{i,p}P_{T_i}) = P_{T_i}$. Therefore we would prefer to rewrite the proof for the sake of completeness in our presentation, although the main tools are quite alike.

The proof of Proposition 5.3 is given as follows.

**Proof:** To prove Proposition 5.3, it suffices to show that $\|M_s\| \leq \frac{1}{16}$ for $1 \leq s \leq 4$ and then take the union bound over all $1 \leq p \leq P$ and $1 \leq i \leq r$. For each fixed pair of $(p, i)$, it is shown in Lemma 5.6 that

\[
\|P_{T_i}A_{i,p}S_{i,p}P_{T_i} - P_{T_i}\| \leq \frac{1}{4}
\]

with probability at least $1 - 4L^{-\alpha}$ if $Q \geq C_\alpha \max\{\mu_{\max}^2, \mu_h^2\}N^2 \log^2 L$. Now we simply take the union bound over all $1 \leq p \leq P$ and $1 \leq i \leq r$ and obtain

\[
P\left(\|P_{T_i}A_{i,p}S_{i,p}P_{T_i} - P_{T_i}\| \leq \frac{1}{4}, \quad \forall 1 \leq i \leq r, 1 \leq p \leq P\right) \geq 1 - 4PrL^{-\alpha} \geq 1 - 4rL^{-\alpha + 1}
\]

if $Q \geq C_\alpha \max\{\mu_{\max}^2, \mu_h^2\}N^2 \log^2 L$ where there are $Pr$ events and $L = PQ$. In order to compensate for the loss of probability due to the union bound and to make the probability of success at least $1 - L^{-\alpha + 1}$, we can just choose $\alpha' = \alpha + \log r$, or equivalently, $Q \geq C_\alpha \max\{\mu_{\max}^2, \mu_h^2\}N^2 \log^2 L \log(r + 1)$.

\[\blacksquare\]

### 5.2.2 Estimation of $\|M_s\|$  

**Lemma 5.6.** Under the assumptions of (14), (10) and (12) and that $a_l \sim \mathcal{N}(0, I_N)$ independently, the estimate

\[
\|M_s\| \leq \frac{1}{16} \quad \text{for } s = 1, 2, 3, 4,
\]

holds with probability at least $1 - L^{-\alpha}$ if $Q \geq C_\alpha \max\{\mu_{\max}^2, \mu_h^2\}N^2 \log^2(L)$.

**Proof:** We will only prove the bound for $M_2$. The proof for $M_3$ is very similar and the proofs for $M_1$ and $M_4$ are easier and are left to the reader.

By definition of $M_2$ in (40),

\[
M_2(Z) := \sum_{l \in \Gamma_p} Z_l(Z), \quad Z_l(Z) = (\overline{h, b_l}h|b_l^*S_{p}(I - hh^*)Zxx^*(a_l a_l^* - I_N)).
\]
Immediately, we have \( \|Z_t\| \leq \| (h, b_t) \| h b_t \| \cdot \| (a_t a_t^* - I_N) \| \) and \( Z_t \) is actually a \( KN \times KN \) matrix. Then we estimate \( \|Z_t\|_{\psi_1} \) as follows:
\[
\|Z_t\|_{\psi_1} \leq \| (h, b_t) \| hh b_t \| \cdot \| (a_t a_t^* - I_N) \|_{\psi_1} \\
\leq \frac{\mu_h}{\sqrt{L}} \cdot \frac{4L \mu_{\max} \sqrt{K}}{3Q} \cdot \| (a_t a_t^* - I_N) \|_{\psi_1} \\
\leq C \frac{\mu_{\max} \mu_h \sqrt{KN}}{Q} \\
\leq C \frac{\max \{ \mu_{\max}^2 K, \mu_h^2 N \}}{Q},
\]
where \( \| (a_t a_t^* - I_N) \|_{\psi_1} \leq C \sqrt{N} \) follows from (42). Therefore we have \( R := \max_{l \in \Gamma_p} \|Z_l\|_{\psi_1} \leq C \frac{\max \{ \mu_{\max}^2 K, \mu_h^2 N \}}{Q} \). Now we proceed to estimate \( \sigma^2 \). By definition, the adjoint of \( Z_t \) in the form of
\[
Z_t^*(Z) = \langle h, b_t \rangle (I_K - hh^*) S_p b_t h^* Z (a_t a_t^* - I_N) x x^*.
\]
Then \( Z^* Z \) and \( ZZ^* \) are easily obtained by definition,
\[
Z_t^* Z_t(Z) = \langle h, b_t \rangle^2 (I_K - hh^*) S_p b_t b_t^* S_p (I_K - hh^*) Z x x^* (a_t a_t^* - I_N)^2 x x^* \\
\text{and} \\
Z_t Z_t^*(Z) = \langle h, b_t \rangle^2 h b_t S_p (I_K - hh^*) S_p b_t h^* Z (a_t a_t^* - I_N) x x^* (a_t a_t^* - I_N).
\]
The expectation of \( Z_t^* Z_t \) and \( Z_t Z_t^* \) are computed via
\[
\mathbb{E}(Z_t^* Z_t(Z)) = \mathbb{E} \langle h, b_t \rangle^2 (I_K - hh^*) S_p b_t b_t^* S_p (I_K - hh^*) Z x x^* (a_t a_t^* - I_N)^2 x x^* \\
= (N + 1) \langle h, b_t \rangle^2 (I_K - hh^*) S_p b_t b_t^* S_p (I_K - hh^*) Z x x^* \\
= (N + 1) I_N \text{ follows from (88). Similarly}, \\
\mathbb{E}(Z_t Z_t^*(Z)) = \mathbb{E} \langle h, b_t \rangle^2 h b_t S_p (I_K - hh^*) S_p b_t h^* Z (a_t a_t^* - I_N) x x^* (a_t a_t^* - I_N)) \\
= \langle h, b_t \rangle^2 h b_t S_p (I_K - hh^*) S_p b_t h^* Z (I_N + x x^*) \\
\text{where } \mathbb{E} \langle a_t a_t^* - I_N \rangle x x^* (a_t a_t^* - I_N) = \|x\|^2 I_N + x x^* \text{ from (43) and the fact that } x \text{ is real. Taking the sum of } \mathbb{E}(Z_t^* Z_t) \text{ and } \mathbb{E}(Z_t Z_t^*) \text{ over } l \in \Gamma_p \text{ gives}
\]
\[
\left\| \sum_{l \in \Gamma_p} \mathbb{E}(Z_t^* Z_t) \right\| = (N + 1) \left\| \sum_{l \in \Gamma_p} \langle h, b_t \rangle^2 (I_K - hh^*) S_p b_t b_t^* S_p (I_K - hh^*) \right\| \\
\leq 2 \frac{\mu_h^2 N}{L} \langle (I_K - hh^*) S_p (I_K - hh^*) \rangle \\
\leq 2 \frac{\mu_h^2 N}{L} \cdot \frac{4L}{3Q} = \frac{8 \mu_h^2 N}{3Q}
\]
and
\[
\left\| \sum_{l \in \Gamma_p} \mathbb{E}(Z_t Z_t^*) \right\| = \left\| \sum_{l \in \Gamma_p} \langle h, b_t \rangle^2 b_t^* S_p (I_K - hh^*) S_p b_t \cdot \|I_N + x x^*\| \right\| \\
\leq 2 \max_{l \in \Gamma_p} \{ b_t^* S_p (I_K - hh^*) S_p b_t \} \sum_{l \in \Gamma_p} \langle h, b_t \rangle^2 \\
\leq 2 \left[ \|S_p\|^2 \max_{l \in \Gamma_p} \|b_t\|^2 \right] \cdot \|T_p\| \\
\leq \frac{32L^2}{9Q^2} \cdot \frac{\mu_{\max}^2 K}{L} \cdot \frac{5Q}{4L} \\
\leq \frac{40 \mu_{\max}^2 K}{9Q}.
\]
Thus the variance $\sigma^2$ is bounded above by

$$\sigma^2 \leq C \max\{\mu_{\text{max}}^2 K, \mu_h^2 N\}$$

and $\log \left(\frac{\sqrt{QR}}{\sigma}\right) \leq C_1 \log L$ for some constant $C_1$. Then we just use (34) to estimate the deviation of $\mathcal{M}_2$ from 0 by choosing $t = \alpha \log L$. Setting $Q \geq C_\alpha \max\{\mu_{\text{max}}^2 K, \mu_h^2 N\} \log L/\delta^2$ gives us

$$\mathcal{M}_2 \leq C \max\left\{\sqrt{\max\{\mu_{\text{max}}^2 K, \mu_h^2 N\} (\alpha \log L + \log(2KN))}, \max\{\mu_{\text{max}}^2 K, \mu_h^2 N\} (\alpha \log L + \log(2KN)) \log L\right\} \leq \delta$$

where $K$ and $N$ are properly assumed to be smaller than $L$. In particular, we take $\delta = \frac{1}{16}$ and have

$$\|\mathcal{M}_2\| \leq \frac{1}{16}$$

with the probability at least $1 - L^{-\alpha}$.

5.3 Proof of $\mu \leq \frac{1}{4r}$

In this section, we aim to show that $\mu \leq \frac{1}{4r}$, where $\mu$ is defined in (27), i.e., the second condition in (28) holds with high probability. The main idea here is first to show that a more general and stronger version of incoherent property,

$$\|P_T^j A^*_{j,p} A^*_k, p S_k, p P_T^k\| \leq \frac{1}{4r}$$

holds with high probability for any $1 \leq p \leq P$ and $j \neq k$. Since the derivation is essentially the same for all different pairs of $(j, k)$ with $j \neq k$, without loss of generality, we take $j = 1$ and $k = 2$ as an example throughout this section. We finish the proof by taking the union bound over all possible sets of $(j, k, p)$.

5.3.1 An explicit expression for $P_T^1 A^*_{1,p} A^*_1, p S_1, p P_T^1$

Following the same procedures as the previous section, we have explicit expressions for $A_{1,p} \mathcal{P}_T^1$ and $P_T^2 A^*_2, p$:

$$P_T^1 A^*_{1,p} A^*_1, p S_1, p P_T^1 = \{\langle Z, h_1 \tilde{v}_1, l + \tilde{u}_1, l x_1^* \rangle\}_{l \in \Gamma_p} \quad P_T^2 A^*_2(z) = \sum_{l \in \Gamma_p} z_l (h_2 v_{2, l}^* + u_{2, l} x_2^*)$$

where $\tilde{u}_1, l, \tilde{v}_1, l$ and $u_{2, l}$ and $v_{2, l}$ are defined in (38) except the notation, where we omit subscript $i$ in the previous section. By combining $P_T^2 A^*_2, p$ and $A_{1,p} S_1, p P_T^1$, we arrive at

$$P_T^2 A^*_2, p A_{1,p} S_1, p P_T^1 (Z) = \sum_{l \in \Gamma_p} \left[ h_2 h_1^* Z \tilde{v}_1, l v_{2, l}^* + h_2 \tilde{u}_1, l Z x_1 v_{2, l}^* + u_2, l h_1^* Z \tilde{v}_1, l x_2^* + u_2, l \tilde{u}_1, l Z x_1 x_2^* \right].$$

(43)
Note that the expectations of all terms are equal to 0 because \( \{ u_{1,t}, v_{1,t} \} \) is independent of \( \{ u_{2,t}, v_{2,t} \} \) and both \( u_{1,t} \) and \( v_{1,t} \) have zero mean. Define \( \mathcal{M}_{s,\text{mix}} \) as

\[
\mathcal{M}_{1,\text{mix}}(Z) := \sum_{t \in \Gamma_p} h_2^* h_1^* Z v_{1,t} v_{2,t}^* = \sum_{t \in \Gamma_p} \langle S_{1,p} h_1, b_{1,t} \rangle \langle h_2, b_{2,t} \rangle h_2^* Z a_{1,t} a_{2,t}^*,
\]

\[
\mathcal{M}_{2,\text{mix}}(Z) := \sum_{t \in \Gamma_p} h_2^* \tilde{u}_{1,t} S_{1,t} v_{1,t} v_{2,t}^* = \sum_{t \in \Gamma_p} \langle a_{1,t}, x_1 \rangle \langle h_2, b_{2,t} \rangle h_2 b_{1,t}^\ast S_{1,p} (I_{K_1} - h_1^* h_1) Z x_1 a_{2,t}^*,
\]

\[
\mathcal{M}_{3,\text{mix}}(Z) := \sum_{t \in \Gamma_p} u_{2,t} h_1^* Z v_{1,t} x_{2,t}^* = \sum_{t \in \Gamma_p} \langle a_{2,t}, x_2 \rangle \langle I_{K_2} - h_2^* h_2^\ast \rangle h_2 b_{1,t}^\ast S_{1,p} (I_{K_1} - h_1^* h_1) Z x_1 x_{2,t}^*,
\]

and there holds

\[
\mathcal{P}_{T_2} A_{2,p}^* A_{1,p} S_{1,p} \mathcal{P}_{T_1} = \sum_{s=1}^4 \mathcal{M}_{s,\text{mix}}.
\]

Each \( \mathcal{M}_{s,\text{mix}} \) can be treated as a \( K_2 N_2 \times K_1 N_1 \) matrix because it is a linear operator from \( \mathbb{C}^{K_1 \times N_1} \) to \( \mathbb{C}^{K_2 \times N_2} \).

**Proposition 5.7.** Under the assumption of (14) and (10) and that \( \{ a_{i,t} \} \) are standard Gaussian random vectors of length \( N_i \),

\[
\| \mathcal{P}_{T_j} A_{j,p}^* A_{k,p} S_{k,p} \mathcal{P}_{T_k} \| \leq \frac{1}{4^r}, \quad 1 \leq j \neq k \leq r, 1 \leq p \leq P
\]

holds with probability at least \( 1 - L^{-\alpha+1} \) if \( Q \geq C_{\alpha} r^2 \max\{ \mu_{\max}^2 K, \mu_{h}^2 N \} \log^2 L \log(r+1) \) where \( K := \max K_i \) and \( N := \max N_i \).

By setting \( Q = L \), we immediately have \( \mu \leq \frac{1}{4^r} \), which is written into the following corollary.

**Corollary 5.8.** Under the assumption of (14) and (10) and that \( \{ a_{i,t} \} \) are standard Gaussian random vectors of length \( N_i \),

\[
\| \mathcal{P}_{T_j} A_{j,p}^* A_{k,p} \mathcal{P}_{T_k} \| \leq \frac{1}{4^r}, \quad 1 \leq j \neq k \leq r, 1 \leq p \leq P
\]

holds with probability at least \( 1 - L^{-\alpha+1} \) if \( Q \geq C_{\alpha} r^2 \max\{ \mu_{\max}^2 K, \mu_{h}^2 N \} \log^2 L \log(r+1) \) where \( K := \max K_i \) and \( N := \max N_i \). In other words, \( \mu \leq \frac{1}{4^r} \).

The proof of Proposition 5.7 follows two steps. First we will show each \( \| \mathcal{M}_{s,\text{mix}} \| \leq \frac{1}{16^r} \) holds with high probability, followed by taking the union bound over all \( j \neq k \) and \( 1 \leq p \leq P \).

**Proof:** For any fixed set of \( (j, k, p) \) with \( j \neq k \), it has been shown, in Lemma 5.9, that

\[
\| \mathcal{P}_{T_j} A_{j,p}^* A_{k,p} S_{k,p} \mathcal{P}_{T_k} \| \leq \frac{1}{4^r}
\]

with probability at least \( 1 - 4L^{-\alpha} \) if \( Q \geq C_{\alpha} r^2 \max\{ \mu_{\max}^2 K, \mu_{h}^2 N \} \log^2 L \). Then we simply take the union bound over all \( 1 \leq p \leq P \) and \( 1 \leq j \neq k \leq r \) and it leads to

\[
P \left( \| \mathcal{P}_{T_j} A_{j,p}^* A_{k,p} S_{k,p} \mathcal{P}_{T_k} \| \leq \frac{1}{4^r}, \quad \forall j \neq k, 1 \leq p \leq P \right) \geq 1 - 4L^{-\alpha} Pr^2 / 2 \geq 1 - 2L^{-\alpha+1} r^2
\]

if \( Q \geq C_{\alpha} r^2 \max\{ \mu_{\max}^2 K, \mu_{h}^2 N \} \log^2 L \) where there are at most \( Pr^2 / 2 \) events and \( L = PQ \). In order to make the probability of success at least \( 1 - L^{-\alpha+1} \), we can just choose \( \alpha' = \alpha + 2 \log r \), or equivalently, \( Q \geq C_{\alpha} r^2 \max\{ \mu_{\max}^2 K, \mu_{h}^2 N \} \log^2 L \log(r+1) \).
5.3.2 Estimation of $\|\mathcal{M}_{s,\text{mix}}\|

Lemma 5.9. Under the assumptions of (14), (10) and (12) and that $a_{i,l} \sim \mathcal{N}(0, I_{N_i})$ independently for $i = 1, 2$ and $l \in \Gamma_p$, there holds

$$\|\mathcal{M}_{s,\text{mix}}\| \leq \frac{1}{16r}$$

for $s = 1, 2, 3, 4$,

with probability $1 - L^{-\alpha}$ if $Q \geq C\alpha r^2 \max\{\mu_2^2 K_1, \mu_2^2 N\} \log^2 L$.

Proof: We only prove the bound for $\mathcal{M}_{2,\text{mix}}$, the proofs of the bounds for $\mathcal{M}_{1,\text{mix}}, \mathcal{M}_{3,\text{mix}},$ and $\mathcal{M}_{4,\text{mix}}$ use similar arguments and are left to the reader.

Following from the definition in (14),

$$\mathcal{M}_{2,\text{mix}} := \sum_{l \in \Gamma_p} Z_l(Z), \quad Z_l(Z) = \langle a_{1,l}, x_1 \rangle \langle h_2, b_{2,l} \rangle h_2 b_{1,l}^* S_{1,p}(I_{K_1} - h_1 h_1^*) Z x_1 a_{2,l}^*$$

and $\|Z_l\| \leq \|\langle a_{1,l}, x_1 \rangle \langle h_2, b_{2,l} \rangle\| \|h_2 b_{1,l}^* S_{1,p}\| \|x_1 a_{2,l}^*\|$. By using Lemma 7.2 and 7.5

$$\|Z_l\|_{\psi_1} \leq \|\langle h_2, b_{2,l} \rangle\| \|S_{1,p} b_{1,l} \| \cdot \|\langle a_{1,l}, x_1 \rangle\| \cdot \|a_{2,l}^*\|_{\psi_1} \leq C \frac{\mu_{\max} \mu h \sqrt{K_1}}{Q} \|\langle a_{1,l}, x_1 \rangle\| \cdot \|a_{2,l}^*\|_{\psi_1} \leq C \frac{\mu_{\max} \mu h \sqrt{K_1 N_2}}{Q} \|\langle a_{1,l}, x_1 \rangle\| \cdot \|a_{2,l}^*\|_{\psi_1} \leq C \frac{\max\{\mu_2^2 K_1, \mu_2^2 N_2\}}{Q}$$

where $\|\langle a_{1,l}, x_1 \rangle\| \cdot \|a_{2,l}^*\|_{\psi_1} \leq C \sqrt{N_2}$ follows from Lemma 7.5. We proceed to estimate $\sigma^2$ by first finding $Z_l^* (Z)$,

$$Z_l^* (Z) = \langle a_{1,l}, x_1 \rangle \langle h_2, b_{2,l} \rangle (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l} h_2^* Z a_{2,l}^* x_1^*$$

$Z_l^* Z_l(Z)$ and $Z_l Z_l^* (Z)$ have the following forms:

$$Z_l^* Z_l(Z) = \|\langle a_{1,l}, x_1 \rangle \langle h_2, b_{2,l} \rangle\|^2 \|a_{2,l}^*\|^2 \|S_{1,p} b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) S_{1,p} (I_{K_1} - h_1 h_1^*) Z x_1^*\|$$

and

$$Z_l Z_l^* (Z) = \|\langle a_{1,l}, x_1 \rangle \langle h_2, b_{2,l} \rangle\|^2 \|b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) Z x_1^*\|$$

The expectations of $Z_l^* Z_l$ and $Z_l Z_l^*$ are

$$\mathbb{E}(Z_l^* Z_l(Z)) = N_2 \|\langle h_2, b_{2,l} \rangle\|^2 \|S_{1,p} b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l}^*\| \cdot h_2 h_2^* Z$$

and

$$\mathbb{E}(Z_l Z_l^* (Z)) = \|\langle h_2, b_{2,l} \rangle\|^2 \|b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l} b_{1,l}^* Z x_1^*\|$$

Taking the sum over $l \in \Gamma_p$ leads to

$$\left\| \sum_{l \in \Gamma_p} \mathbb{E}(Z_l^* Z_l) \right\| = N_2 \left\| \sum_{l \in \Gamma_p} \|\langle h_2, b_{2,l} \rangle\|^2 \|S_{1,p} b_{1,l}^* S_{1,p} (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l}^*\| \right\| \leq \frac{\mu_2^2 N_2}{L} \left\| \sum_{l \in \Gamma_p} S_{1,p} b_{1,l} b_{1,l}^* S_{1,p} \right\| \leq \frac{\mu_2^2 N_2}{L} \cdot \|S_{1,p}\| \leq \frac{\mu_2^2 N_2}{3Q} \leq \frac{4\mu_2^2 N_2}{3Q}$$
and
\[
\left\| \sum_{l \in \Gamma} E(Z_l Z_l^*) \right\| = \sum_{l \in \Gamma_p} |\langle h_2, b_{2,l} \rangle|^2 b_{1,l}^* S_{1,p}(I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l} \\
= \sum_{l \in \Gamma_p} \| (I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,l} \|^2 |\langle h_2, b_{2,l} \rangle|^2 \\
\leq \max_{l \in \Gamma_p} \{ \| S_{1,p} \|^2 \| b_{1,l} \|^2 \} \| T_{2,p} \| \\
\leq \frac{16L^2}{9Q^2} \frac{\mu_{\text{max}}^2 K_1}{L} \cdot \frac{5Q}{4L} \leq \frac{20\mu_{\text{max}}^2 K_1}{9Q}.
\]

Thus the variance \( \sigma^2 \) is bounded by
\[
\sigma^2 \leq C \max \{ \mu_{\text{max}}^2 K_1, \mu_{\text{max}}^2 N \} \leq C \max \{ \mu_{\text{max}}^2 K_1, \mu_{\text{max}}^2 N \}.
\]

Then we just apply (84) to estimate the deviation of \( M_{2,\text{mix}} \) from 0 by choosing \( t = \alpha \log L \).

Letting \( Q \geq C_\alpha \max \{ \mu_{\text{max}}^2 K_1, \mu_{\text{max}}^2 N \} \log^2 L/\delta^2 \) gives us
\[
M_{2,\text{mix}} \leq C \max \left\{ \sqrt{\frac{\max \{ \mu_{\text{max}}^2 K_1, \mu_{\text{max}}^2 N \}}{Q} (\alpha \log L + \log(2KN))}, \frac{\mu_{\text{max}}^2 K_1}{Q} (\alpha \log L + \log(2KN)) \log L \right\} \leq \delta,
\]
with probability at least \( 1 - L^{-\alpha} \) where \( K \) and \( N \) are properly assumed to be smaller than \( L \).

Let \( \delta = \frac{1}{16r} \) and \( Q \geq C_\alpha r^2 \max \{ \mu_{\text{max}}^2 K_1, \mu_{\text{max}}^2 N \} \log^2 L \),
\[
\| M_{2,\text{mix}} \| \leq \frac{1}{16r}
\]
with the probability at least \( 1 - L^{-\alpha} \).

5.4 Constructing a dual certificate
In this section, we will construct a \( \lambda \) such that
\[
\| h_i x_i^* - \mathcal{P}_{T_i} (A_i^*(\lambda)) \|_F \leq (5r\gamma)^{-1}, \quad \| \mathcal{P}_{T_i^\perp} (A_i^*(\lambda)) \| \leq \frac{1}{2}
\]
holds simultaneously for all \( 1 \leq i \leq r \). If such a \( \lambda \) exists, then solving (8) yields exact recovery according to Lemma 5.2. The difficulty of this mission is obvious since we require all \( A_i^*(\lambda) \) to be close to \( h_i x_i^* \) and “small” on \( T_i^\perp \). However, it becomes possible with help of the incoherence between \( A_i \) and \( A_j \). The method to achieve this goal is to apply a well-known and widely used technique called golfing scheme, developed by Gross in [19].

5.4.1 Golfing scheme
The approximate dual certificate \( \{ Y_i := A_i^*(\lambda) \}_{i=1}^r \) satisfying Lemma 5.2 is constructed via a sequence of random matrices, by following the philosophy of golfing scheme. The constructed sequence \( \{ Y_{i,p} \}_{p=1}^P \) would approach \( h_i x_i^* \) on \( T_i \) exponentially fast while keeping \( Y_{i,p} \) “small” on \( T_i^\perp \) at the same time.
Construct an approximate dual certificate via golfing scheme

1. Initialize $Y_{i,0} = 0_{K \times N_i}$ for all $1 \leq i \leq r$ and

$$\lambda_0 := \sum_{j=1}^{r} A_{j,1}(S_{j,1} h_j x_j^* ) \in \mathbb{C}^L.$$  

2. For $p$ from 1 to $P$ (where $P$ will be specified later in Lemma 5.10), we define the following recursive formula:

$$\begin{align*}
\lambda_{p-1} & := \sum_{j=1}^{r} A_{j,p}(S_{j,p}(h_j x_j^* - P_{T_j}(Y_{j,p-1}))) , \\
Y_{i,p} & := Y_{i,p-1} + A_{i,p}^*(\lambda_{p-1}) , \quad 1 \leq i \leq r .
\end{align*}$$  

3. $Y_{i,p}$ denotes the result after $p$-th iteration and let $Y_i := Y_{i,p}$, i.e., the final outcome for each $i$.

Denote $W_{i,p}$ as the difference between $Y_{i,p}$ and $h_i x_i^*$ on $T_i$, i.e.,

$$W_{i,p} = h_i x_i^* - P_{T_i}(Y_{i,p}) \in T_i , \quad W_{i,0} = h_i x_i^*$$  

and (48) can be rewritten into the following form:

$$\lambda_{p-1} = \sum_{j=1}^{r} A_{j,p}(S_{j,p}W_{j,p-1}).$$

Moreover, $W_{i,p}$ yields the following relation:

$$W_{i,p} = W_{i,p-1} - \sum_{j=1}^{r} P_{T_i} A_{i,p}^* A_{j,p}(S_{j,p}W_{j,p-1})$$  

from (49) and (50). An important observation here is that each $A_{i,p}^*(\lambda_{p-1})$ is an unbiased estimator of $W_{i,p-1}$, i.e.,

$$\mathbb{E}(A_{i,p}^*(\lambda_{p-1})) = \sum_{j=1}^{r} \mathbb{E}(A_{i,p}^* A_{j,p}(S_{j,p}W_{j,p-1})) = W_{i,p-1} .$$  

where $\mathbb{E}(A_{i,p}^* A_{j,p}(S_{j,p}W_{j,p-1})) = 0$ for all $j \neq i$ due to the independence between $A_{j,p}$ and $A_{i,p}$ and $\mathbb{E}(A_{i,p}^* A_{j,p}(S_{j,p}W_{j,p-1})) = W_{i,p-1}$. Remember that $\{W_{j,p-1}\}_{j=1}^{r}$ are independent of $\{A_{i,p}\}_{i=1}^{r}$, which is based on the construction of sequences in (48) and (49). More precisely, the expectation above should be treated as the conditional expectation of $A_{i,p}^*(\lambda_{p-1})$ given $\{W_{j,p-1}\}_{j=1}^{r}$ are known.

5.4.2 $\|P_{T_1}(Y_i) - h_i x_i^*\|_F$ decays exponentially fast

**Lemma 5.10.** Conditioned on (41) and (46), the golfing scheme (48) and (49) generate a sequence of $\{Y_{i,p}\}_{p=1}^{P}$ such that

$$\|W_{i,p}\|_F = \|P_{T_1}(Y_{i,p}) - h_i x_i^*\|_F \leq 2^{-p}$$

hold simultaneously for all $1 \leq i \leq r$. In particular, if $P \geq \log_2(5r\gamma)$,

$$\|P_{T_1}(Y_i) - h_i x_i^*\| \leq 2^{-\log_2(5r\gamma)} \leq \frac{1}{5r\gamma}$$

where $Y_i := Y_{i,p}$. In other words, the first condition in (47) holds.
**Proof:** Directly following from (51) leads to

\[ W_{i,p} = W_{i,p-1} - \mathcal{P}_i \mathcal{A}^*_i \mathcal{A}_{i,p}(S_{i,p}W_{i,p-1}) - \sum_{j \neq i} \mathcal{P}_j \mathcal{A}^*_i \mathcal{A}_{j,p}(S_{j,p}W_{j,p-1}). \]  

(53)

Note that \( W_{j,p-1} \in T_j \) and thus \( W_{j,p-1} = \mathcal{P}_j(W_{j,p-1}) \). By applying (41) and (46),

\[ \| W_{i,p} \|_F \leq \frac{1}{4} \| W_{i,p-1} \|_F + \frac{1}{4r} \sum_{j \neq i} \| W_{j,p-1} \|_F, \quad 1 \leq i \leq r. \]

It is easy to see that

\[ \max_{1 \leq i \leq r} \| W_{i,p} \|_F \leq \frac{1}{4} \max_{1 \leq i \leq r} \| W_{i,p-1} \|_F. \]

Recall that \( \| W_{i,0} \|_F = \| h_i x_i^* \|_F = 1 \) for all \( 1 \leq i \leq r \) and by the induction above, we prove that

\[ \| W_{i,p} \|_F \leq 2^{-p}, \quad 1 \leq p \leq P, \quad 1 \leq i \leq r. \]

\[ \text{5.4.3 Proof of } \| \mathcal{P}_i^\perp(Y_i) \| \leq \frac{1}{2} \]

In the previous section, we have already shown that \( \mathcal{P}_i^\perp(Y_{i,p}) \) approaches \( h_i x_i^* \) exponentially fast with respect to \( p \). The only missing piece of the proof is to show that \( \| \mathcal{P}_i^\perp(Y_{i,p}) \| \) is bounded by \( \frac{1}{2} \) for all \( 1 \leq i \leq r \), i.e., the second condition in (29) holds. Without loss of generality, we set \( i = 1 \). Following directly from (48) and (49),

\[ Y_1 := Y_{1,p} = \sum_{p=1}^{P} \mathcal{A}^*_1 \mathcal{A}_{1,p}(\lambda_{p-1}). \]

There holds

\[ \| \mathcal{P}_1^\perp(Y_1) \| = \left\| \mathcal{P}_1^\perp \left( \sum_{p=1}^{P} \mathcal{A}^*_1 \mathcal{A}_{1,p}(\lambda_{p-1}) - W_{1,p-1} \right) \right\| \]

\[ \leq \sum_{p=1}^{P} \mathcal{A}^*_1 \mathcal{A}_{1,p}(\lambda_{p-1}) - W_{1,p-1}, \]

where \( \mathcal{P}_1^\perp(W_{1,p-1}) = 0 \). It suffices to demonstrate that \( \| \mathcal{A}^*_1 \mathcal{A}_{1,p}(\lambda_{p-1}) - W_{1,p-1} \| \leq 2^{-p-1} \) for \( 1 \leq p \leq P \) in order to justify \( \| \mathcal{P}_1^\perp(Y_1) \| \leq \frac{1}{2} \) since

\[ \| \mathcal{P}_1^\perp(Y_1) \| \leq \sum_{p=1}^{P} 2^{-p-1} < \frac{1}{2}. \]

Before moving to the proof, we first define the quantity \( \mu_p \) which will be useful in the proof,

\[ \mu_p := \frac{Q}{\sqrt{L}} \max_{1 \leq i \leq r, l \in \Gamma_{p+1}} \| W_{i,p} S_{i,p+1} b_{i,l} \|. \]  

(54)

In particular, \( \mu_0 \leq \mu_h \) because of

\[ \mu_0 = \frac{Q}{\sqrt{L}} \max_{i,l \in \Gamma_1} \| x_i h_i^* S_{i,1} b_{i,l} \| = \frac{Q}{\sqrt{L}} \max_{i,l \in \Gamma_1} \| h_i^* S_{i,1} b_{i,l} \| \leq \mu_h. \]

26
and the definition of $\mu_h$ in (13). Also we define $w_{i,l}$ as

$$w_{i,l} := W_{i,p-1}^*s_{i,l}b_{i,l} \in \mathbb{C}^{N_i}, \ l \in \Gamma_p, \ 1 \leq i \leq r$$

and there holds

$$\max_{1 \leq i \leq r, l \in \Gamma_p} \|w_{i,l}\| \leq \frac{\sqrt{L}}{Q} \mu_{p-1}. \tag{56}$$

**Remark 5.11.** The definition of $\mu_p$ is a little complicated but the idea is simple. Since we have already shown in Lemma 5.10 that $W_{i,p} \in T_i$ is very close to $h_{i,x_i^*}$ for large $p$, $\mu_p$ can be viewed as a measure of the incoherence between $W_{i,p}$ (an approximation of $h_{i,x_i}$) and $\{b_{i,l}\}_{l \in \Gamma_{p+1}}$. We would like to have “small” $\mu_p$, i.e., $\mu_p \leq \|W_{i,p}\| \mu_h \leq 2^{-p} \mu_h$ which guarantees that $A_{i,p}(\lambda_{p-1})$ concentrates well around $W_{i,p-1}$ for all $i$ and $p$. This insight leads us to the following lemma.

**Lemma 5.12.** Let $\mu_p$ be defined in (54) and $W_{i,p}$ satisfy

$$\mu_p \leq 2^{-p} \mu_h, \ \|W_{i,p}\|_F \leq 2^{-p}, \ 1 \leq p, 1 \leq i \leq r.$$  

If $Q \geq C_{\alpha}r \max\{\mu_{\text{max}}^2, \mu_h^2 N\} \log^2 L \log(r + 1)$, then

$$\|A_{i,p}(\lambda_{p-1}) - W_{i,p-1}\| \leq 2^{-p-1},$$  

simultaneously for $(p,i)$ with probability at least $1 - L^{-\alpha+1}$. Thus, the second condition in (47),

$$\|P_{T_i}(Y_i)\| \leq \frac{1}{2}$$  

holds simultaneously for all $1 \leq i \leq r$.

**Remark 5.13.** The assumption $\mu_p \leq 2^{-p} \mu_h$ is justified in Lemma 5.14.

**Proof:** Without loss of generality, we start with $i = 1$. It is shown in (52) that

$$\mathbb{E} (A_{1,p}^*(\lambda_{p-1}) - W_{1,p-1}) = 0.$$

First we rewrite $A_{1,p}^*(\lambda_{p-1}) - W_{1,p-1}$ into the sum of rank-1 matrices with mean $0$ by (48) and (17),

$$A_{1,p}^*(\lambda_{p-1}) - W_{1,p-1} = \sum_{i \in \Gamma_p} \left[ b_{1,l}b_{1,l}^*s_{1,p}W_{1,p-1}(a_{1,l}a_{1,l}^* - I_{N_1}) + \sum_{j \neq 1} b_{1,l}b_{j,l}^*s_{j,p}W_{j,p-1}a_{j,l}a_{1,l}^* \right]. \tag{57}$$

Denote $Z_l$ by

$$Z_l := b_{1,l}w_{1,l}^*(a_{1,l}a_{1,l}^* - I_{N_1}) + \sum_{j \neq 1} b_{1,l}w_{j,l}^*a_{j,l}a_{1,l}^* \in \mathbb{C}^{K_i \times N_i}, \tag{58}$$

where $w_{j,l}$ is defined in (55). The goal is to bound the operator norm of (57), i.e., $\|\sum_{l \in \Gamma_p} Z_l\|$, by $2^{-p-1}$. An important fact here is that $\mu_{p-1}$ is independent of all $a_{i,l}$ with $l \in \Gamma_p$ because $\mu_{p-1}$ is a function of $\{a_{i,k}\}_{k \in \Gamma_r, s < p}$. Following from (54) and the assumption $\mu_p \leq 2^{-p} \mu_h$, we have

$$\|w_{i,l}\| \leq \frac{\sqrt{L}}{Q} \mu_{p-1} \leq \frac{\sqrt{L}}{Q} 2^{-p+1} \mu_h, \ \forall l \in \Gamma_p. \tag{59}$$

The proof is more or less a routine: estimate $\|Z_l\|_{\psi_1}$, $\sigma^2$ and apply (84). For any fixed $l \in \Gamma_p$,

$$\|Z_l\| \leq \|b_{1,l}w_{1,l}^*(a_{1,l}a_{1,l}^* - I_{N_1})\| + \sum_{j \neq 1} \|b_{1,l}w_{j,l}^*a_{j,l}a_{1,l}^*\|$$

$$\leq \frac{\mu_{\text{max}} \sqrt{K_i}}{\sqrt{L}} \left[ \|w_{1,l}^*(a_{1,l}a_{1,l}^* - I_{N_1})\| + \sum_{j \neq 1} \|w_{j,l}^*a_{j,l}a_{1,l}^*\| \right]. \tag{27}$$
Note that for $j \neq 1$, $\mathbf{w}_j^* \mathbf{a}_{j,l} \sim \mathcal{N}(0, \| \mathbf{w}_{j,l} \|^2)$ and $\| \mathbf{a}_{1,l} \|^2 \sim \chi^2_{N_1}$. There holds
\[
\| (\mathbf{w}_j^* \mathbf{a}_{j,l} \cdot \mathbf{a}_{1,l}) \|_{\psi_1} \leq C \sqrt{N_1} \| \mathbf{w}_{j,l} \| \leq C \frac{2^{p+1} \mu \sqrt{LN_1}}{Q}
\]
\[
\| \mathbf{w}_j^*(\mathbf{a}_{1,l} \mathbf{a}_{j,l}^* - \mathbf{I}_{N_1}) \|_{\psi_1} \leq C \sqrt{N_1} \| \mathbf{w}_{1,l} \| \leq C \frac{2^{p+1} \mu \sqrt{LN_1}}{Q}
\]
follow from (92), (59) and Lemma 7.5. Taking the sum over $j$, from 1 to $r$, gives
\[
\| \mathbf{Z}_l \|_{\psi_1} \leq C \frac{2^{p+1} r \mu_{\text{max}} \mu \sqrt{K_1 N_1}}{Q} \leq C \frac{2^{p+1} r \mu_{\text{max}}^2 K_1^2 N_1}{Q}, \quad l \in \Gamma_p.
\]
Thus we have $R := \max_{l \in \Gamma_p} \| \mathbf{Z}_l \|_{\psi_1} \leq C \frac{2^{p+1} r \mu_{\text{max}}^2 K_1^2 N_1}{Q}$. Now let’s move on to the estimation of $\sigma^2$. From (58), we have
\[
\mathbf{Z}_l = (\mathbf{a}_{1,l} \mathbf{a}_{j,l}^* - \mathbf{I}_{N_1}) \mathbf{w}_{1,l} \mathbf{b}_{j,l}^* + \sum_{j \neq 1} \mathbf{a}_{1,l} \mathbf{a}_{j,l}^* \mathbf{w}_{j,l} \mathbf{b}_{j,l}^*.
\]
The corresponding $\mathbf{Z}_l^* \mathbf{Z}_l$ and $\mathbf{Z}_l \mathbf{Z}_l^*$ have quite complicated expressions. However, all the cross terms have zero expectation, which simplifies $\mathbb{E}(\mathbf{Z}_l^* \mathbf{Z}_l)$ and $\mathbb{E}(\mathbf{Z}_l \mathbf{Z}_l^*)$ a lot.
\[
\mathbb{E}(\mathbf{Z}_l^* \mathbf{Z}_l) = \mathbb{E} \left( \| \mathbf{b}_{1,l} \|^2 (\mathbf{a}_{1,l} \mathbf{a}_{1,l}^* - \mathbf{I}_{N_1}) \mathbf{w}_{1,l} \mathbf{w}_{j,l}^* (\mathbf{a}_{1,l} \mathbf{a}_{j,l}^* - \mathbf{I}_{N_1}) + \| \mathbf{b}_{1,l} \|^2 \sum_{j \neq 1} |\mathbf{w}_{j,l} \mathbf{a}_{j,l}^*|^2 \mathbf{a}_{1,l} \mathbf{a}_{1,l}^* \right)
\]
\[
= \| \mathbf{b}_{1,l} \|^2 \left( \sum_{j=1}^r \| \mathbf{w}_{j,l} \|^2 \right) \mathbf{I}_{N_1} + \| \mathbf{b}_{1,l} \|^2 \mathbf{w}_{1,l} \mathbf{w}_{1,l}^*,
\]
which follows from (93).
\[
\mathbb{E}(\mathbf{Z}_l \mathbf{Z}_l^*) = \mathbb{E} \left( \| (\mathbf{a}_{1,l} \mathbf{a}_{j,l}^* - \mathbf{I}_{N_1}) \mathbf{w}_{1,l} \|^2 \mathbf{b}_{j,l} \mathbf{b}_{j,l}^* + \sum_{j \neq 1} \| \mathbf{a}_{1,l} \|^2 |\langle \mathbf{w}_{j,l}, \mathbf{a}_{j,l} \rangle|^2 \mathbf{b}_{j,l} \mathbf{b}_{j,l}^* \right)
\]
\[
= N_1 \sum_{j=1}^r \| \mathbf{w}_{j,l} \|^2 \mathbf{b}_{j,l} \mathbf{b}_{j,l}^* + \| \mathbf{w}_{1,l} \|^2 \mathbf{b}_{1,l} \mathbf{b}_{1,l}^*,
\]
which follows from (88) and $\mathbb{E} \| \mathbf{a}_{1,l} \|^2 = N_1$.
where the last inequality follows from \( \| W_{i,p-1} \|_F \leq 2^{-p+1} \) and \( \| \cdot \|_s \) is the dual norm of \( \| \cdot \| \).

\[
\left\| \sum_{t \in \Gamma_p} E(\mathcal{Z}_t Z_t^*) \right\| = \left\| \sum_{t \in \Gamma_p} \left[ N_1 \sum_{j=1}^{r} \| w_{j,t} \| \right] \right\|
\leq \max_{j,t} \left\| w_{j,t} \right\|^2 \left\| \sum_{t \in \Gamma_p} \left[ rN_1 b_{j,t} b_{j,t}^* + b_{1,t} b_{1,t}^* \right] \right\|
\leq \frac{\mu_{p-1} L}{Q^2} \cdot 2rN_1 \left\| T_{1,p} \right\| = \frac{5r \mu_{p-1}^2 N_1}{2Q}
\leq C \frac{4^{-p+1} r \mu_{p}^2 N_1}{Q}
\]

where \( \| w_{i,t} \| \leq \frac{\sqrt{\mu_{p-1}}}{Q} \leq \frac{2^{-p+1} \sqrt{\mu_{p-1}}}{Q} \). Finally we have an upper bound of \( \sigma^2 \):

\[
\sigma^2 \leq C \frac{4^{-p+1} r \max \{ \mu_{max}^2 K, \mu_h^2 N \}}{Q} \leq C \frac{4^{-p+1} r \max \{ \mu_{max}^2 K, \mu_h^2 N \}}{Q}
\]

By using Bernstein inequality \[84\] with \( t = \alpha \log L \) and \( \log \left( \frac{\sqrt{QR}}{\sigma} \right) \leq C_1 \log L \), we have

\[
\left\| \sum_{t \in \Gamma_p} Z_t \right\| \leq C_0 \frac{2^{-p+1} \max \left\{ \sqrt{\alpha \log \{ \mu_{max}^2 K, \mu_h^2 N \} \log L}, \alpha r \log \{ \mu_{max}^2 K, \mu_h^2 N \} \log^2 L \right\}}{L}
\]

In order to let \( \left\| \sum_{t \in \Gamma_p} Z_t \right\| \leq 2^{-p+1} \) hold with probability at least \( 1 - L^{-\alpha} \), it suffices to let \( Q \geq C_0 r \log \{ \mu_{max}^2 K, \mu_h^2 N \} \log^2 L \). This finishes the proof for case when \( i = 1 \). Then we take the union bound over all \( 1 \leq p \leq P \) and \( 1 \leq i \leq r \), i.e., totally \( rP \) events and then

\[
\left\| A_{i,p}(\mathbf{X}_{p-1}) - W_{i,p-1} \right\| \leq 2^{-p+1}
\]

holds simultaneously for all \( (p,i) \) with probability at least \( 1 - rPL^{-\alpha} \geq 1 - rL^{-\alpha+1} \). To compensate the loss of probability from the union bound, we can choose \( \alpha' = \alpha + \log r \), which gives \( Q \geq C_0 r \log \{ \mu_{max}^2 K, \mu_h^2 N \} \log^2 L \log(r + 1) \).

\[\blacksquare\]

5.4.4 Proof of \( \mu_p \leq \frac{1}{2} \mu_{p-1} \)

Recall that \( \mu_p \) is defined in \[54\] as \( \mu_p = \frac{Q}{L} \max_{1 \leq i \leq r, t \in \Gamma_{p+1}} \left( \| b_{i,t}^* S_{i,p+1} W_{i,p} \| \right) \). The goal is to show that \( \mu_p \leq \frac{1}{2} \mu_{p-1} \) and thus \( \mu_p \leq 2^{-p} \mu_h \) hold with high probability.

**Lemma 5.14.** Under the assumption of \[14\], \[10\] and \[12\] and that \( \mathbf{a}_{i,t} \sim \mathcal{N}(0, I_N) \) independently for \( 1 \leq i \leq r \) then

\[
\mu_p \leq \frac{1}{2} \mu_{p-1},
\]

with probability at least \( 1 - L^{-\alpha+1} \) if \( Q \geq C_0 r^2 \max \{ \mu_{max}^2 K, \mu_h^2 N \} \log^2 L \log(r + 1) \).

**Proof:** In order to show that \( \mu_p \leq \frac{1}{2} \mu_{p-1} \), it is equivalent to prove

\[
\frac{Q}{L} \| b_{i,t}^* S_{i,p+1} W_{i,p} \| \leq \frac{1}{2} \mu_{p-1}
\] (60)
for all $l \in \Gamma_{p+1}$ and $1 \leq i \leq r$. From now on, we set $i = 1$ and fix $l \in \Gamma_{p+1}$ and show that
\[
\frac{Q}{\sqrt{L}} \|b_{i,l}^{*}S_{i,p+1}W_{i,p}\| \leq \frac{1}{2} \mu_{p-1}
\]
holds with high probability. Then taking the union bound over $(i,l)$ completes the proof. Following from (53) and (55), there holds
\[
-W_{1,p} = \mathcal{P}_{T_1} \left( \sum_{k \in \Gamma_p} b_{1,k} w_{1,k} (a_{1,k} a_{1,k}^* - I_{N_1}) \right) + \sum_{j \neq 1} \mathcal{P}_{T_1} \left( \sum_{k \in \Gamma_p} b_{1,k} w_{1,k} a_{j,k} a_{1,k}^* \right)
=: \Pi_1 + \Pi_2.
\]
Obviously, (60) follows directly from the following two inequalities,
\[
\|b_{i,l}^{*}S_{i,p+1+\Pi_1}\| \leq \frac{\sqrt{T}}{4Q} \mu_{p-1}, \quad \|b_{i,l}^{*}S_{i,p+1+\Pi_2}\| \leq \frac{\sqrt{T}}{4Q} \mu_{p-1}.
\]

**Step 1: Proof of $\|b_{i,l}^{*}S_{1,p+1+\Pi_1}\| \leq \frac{\sqrt{T} \mu_{p-1}}{4Q}$**

For a fixed $l \in \Gamma_{p+1}$,
\[
b_{i,l}^{*}S_{1,p+1+\Pi_1} = \sum_{k \in \Gamma_p} b_{1,l}^{*}S_{1,p+1} \left[ h_{1,l} h_{1,l} w_{1,k} (a_{1,k} a_{1,k}^* - I_{N_1}) \right] + (I_{N_1} - h_{1,l} h_{1,l}^{*}) b_{1,l} w_{1,k} (a_{1,k} a_{1,k}^* - I_{N_1}) x_1 x_1^*.
\]

where $\mathcal{P}_{T_1}$ has an explicit form in (24). Define
\[
z_k := (a_{1,k} a_{1,k}^* - I_{N_1}) w_{1,k}, b_{1,l}^{*} h_{1,l} S_{1,p+1} b_{1,l} \in \mathbb{C}^{N_1}
\]
and
\[
z_k := b_{1,l}^{*} S_{1,p+1} (I_{N_1} - h_{1,l} h_{1,l}^{*}) b_{1,l} w_{1,k} (a_{1,k} a_{1,k}^* - I_{N_1}) x_1.
\]

There holds
\[
\|b_{i,l}^{*}S_{1,p+1+\Pi_1}\| \leq \left\| \sum_{k \in \Gamma_p} z_k \right\| + \sum_{k \in \Gamma_p} \left\| z_k \right\|.
\]

Our goal now is to bound both $\| \sum_{k \in \Gamma_p} z_k \|$ and $\| \sum_{k \in \Gamma_p} z_k \|$ by $\frac{\sqrt{T} \mu_{p-1}}{8Q}$. First we take a look at $\sum_{k \in \Gamma_p} z_k$. For each $k$,
\[
\|z_k\|_{\psi_1} = |b_{1,l}^{*} S_{1,p+1} h_{1,l}| \cdot |(a_{1,k} a_{1,k}^* - I_{N_1}) w_{1,k}| \cdot \|z_k\|_{\psi_1}
\leq C \frac{\sqrt{T} \mu_{h}}{Q} \frac{\sqrt{T} \sqrt{N_1}}{\mu_{h}} \|w_{1,k}\| = C \frac{\mu_{h}^2 \sqrt{N_1}}{Q} \|w_{1,k}\|,
\]
which follows from (14) and $\| (a_{1,k} a_{1,k}^* - I_{N_1}) w_{1,k} \|_{\psi_1} \leq C \sqrt{N_1} \|w_{1,k}\|$ in (92). The expectation of $E(z_k^* z_k)$ and $E(z_k z_k^*)$ can be easily computed,
\[
E(z_k^* z_k) = |b_{1,l}^{*} S_{1,p+1} h_{1,l}|^2 |h_{1,l} b_{1,k}|^2 E[w_{1,k}^* (a_{1,k} a_{1,k}^* - I_{N_1})^2 w_{1,k}]
= |N_1 + 1| |b_{1,l}^{*} S_{1,p+1} h_{1,l}|^2 |h_{1,l} b_{1,k}|^2 \|w_{1,k}\|^2,
\]
\[
E(z_k z_k^*) = |b_{1,l}^{*} S_{1,p+1} h_{1,l}|^2 |h_{1,l} b_{1,k}|^2 E[(a_{1,k} a_{1,k}^* - I_{N_1}) w_{1,k} w_{1,k}^* (a_{1,k} a_{1,k}^* - I_{N_1})]
= |b_{1,l}^{*} S_{1,p+1} h_{1,l}|^2 |h_{1,l} b_{1,k}|^2 \|w_{1,k}\|^2 I_{N_1} + w_{1,k} w_{1,k}^*.
\]

which follows from (88) and (93).
The estimation of $\|\sum_{k \in \Gamma_p} \mathbb{E}(z_k z_k^*)\|$ is quite similar to that of $\|\sum_{k \in \Gamma_p} \mathbb{E}(z_k^* z_k)\|$ and thus we give the result directly without going to the details,

$$\left\| \sum_{k \in \Gamma_p} \mathbb{E}(z_k z_k^*) \right\| \leq \frac{5 \mu_h^2 \max_{k \in \Gamma_p} \|w_{1,k}\|^2}{2Q}.$$ 

Therefore,

$$R := \max_{k \in \Gamma_p} \|z_k\|_{\psi_1} \leq C \frac{\mu_h^2 \sqrt{N}}{Q} \max_{k \in \Gamma_p} \|w_{1,k}\|$$

and similarly, we have

$$\sigma^2 \leq C \frac{\mu_h^2 N \max_{k \in \Gamma_p} \|w_{1,k}\|^2}{Q}.$$ 

Then we just apply (84) with $t = \alpha \log L$ and $\log(\sqrt{QR}/\sigma) \leq C_1 \log L$ to estimate $\|\sum_{k \in \Gamma_p} z_k\|$

$$\left\| \sum_{k \in \Gamma_p} z_k \right\| \leq C \max_{k \in \Gamma_p} \|w_{1,k}\|^2 \max \left\{ \sqrt{\frac{\alpha \mu_h^2 N}{Q} \log L}, \frac{\alpha \mu_h^2 \sqrt{N}}{Q} \log^2 L \right\}. \quad (65)$$

Note that $\max_{k \in \Gamma_p} \|w_{1,k}\| \leq \frac{\sqrt{T_{\mu_k} - 1}}{Q}$ in (56) and thus it suffices to let $Q \geq C_1 \alpha \mu_h^2 N \log^2 L$ to ensure that $\|\sum_{k \in \Gamma_p} z_k\| \leq \frac{\sqrt{T_{\mu_k} - 1}}{8Q}$ holds with probability at least $1 - L^{-\alpha}$.

Concerning $z_k$ in (63), we first estimate $\|z_k\|_{\psi_1}$:

$$\|z_k\|_{\psi_1} = \|b_1^* S_{1,p+1}(I_{K_1} - h_1 h_1^*) b_{1,k}\| \|w_{1,k}^*(a_{1,k} a_{1,k}^* - I_{N_1}) x_1\|_{\psi_1} = \|b_1^* \| S_{1,p+1} \| b_{1,k}\| \|w_{1,k}^*(a_{1,k} a_{1,k}^* - I_{N_1}) x_1\|_{\psi_1} \leq C \frac{\mu_h^2 \max_{k \in \Gamma_p} K_1}{L} \frac{4L}{3Q} \|w_{1,k}\| \leq C \frac{\mu_h^2 \max_{k \in \Gamma_p} K_1}{Q} \|w_{1,k}\|$$

where $\|w_{1,k}^*(a_{1,k} a_{1,k}^* - I_{N_1}) x_1\|_{\psi_1} \leq C \|w_{1,k}\|$ in (94). Thus $R := \max\{\|z_k\|_{\psi_1}\} \leq C \frac{\mu_h^2 \max_{k \in \Gamma_p} K_1}{Q} \|w_{1,k}\|$.

Furthermore,

$$\mathbb{E}|z_k|^2 = \|b_1^* S_{1,p+1}(I_{K_1} - h_1 h_1^*) b_{1,k}\|^2 \mathbb{E}\left[ \|w_{1,k}^*(a_{1,k} a_{1,k}^* - I_{N_1}) x_1 x_1^* a_{1,k} a_{1,k}^* - I_{N_1}) w_{1,k}\| \right] = \|b_1^* S_{1,p+1}(I_{K_1} - h_1 h_1^*) b_{1,k}\|^2 \|w_{1,k}^*(I_{N_1} + x_1 x_1^*) w_{1,k}\|$$

which follows from (63). The variance $\sum_{k \in \Gamma_p} z_k$ is bounded by

$$\sigma^2 \leq \|b_1^* S_{1,p+1}(I_{K_1} - h_1 h_1^*) T_{1,p}(I_{K_1} - h_1 h_1^*) S_{1,p} b_{1,k} \| \max_{k \in \Gamma_p} \|w_{1,k}^*(I_{N_1} + x_1 x_1^*) w_{1,k}\| \leq \|b_1^* S_{1,p+1}(I_{K_1} - h_1 h_1^*) T_{1,p} \| \max_{k \in \Gamma_p} \|w_{1,k}^*(I_{N_1} + x_1 x_1^*) w_{1,k}\| \leq 2 \mu_h^2 \max_{k \in \Gamma_p} K_1 16L^2 5Q \max_{k \in \Gamma_p} \|w_{1,k}\|^2 \leq \frac{40 \mu_h^2 \max_{k \in \Gamma_p} K_1}{9Q} \max_{k \in \Gamma_p} \|w_{1,k}\|^2.$$ 

Similar to what we have done in (65),

$$\left\| \sum_{k \in \Gamma_p} z_k \right\| \leq C \max_{k \in \Gamma_p} \|w_{1,k}\|^2 \max \left\{ \sqrt{\frac{\alpha \mu_h^2 \max_{k \in \Gamma_p} K_1}{Q} \log L}, \frac{\alpha \mu_h^2 \max_{k \in \Gamma_p} K_1}{Q} \log^2 L \right\}. \quad (66)$$
Note that $\max_{k \in \Gamma_p} \|w_{1,k}\| \leq \frac{\sqrt{T} \mu p - 1}{Q}$ and thus $Q \geq C_\alpha \mu_{\max}^2 K \log^2 L$ guarantees that $|\sum_{k \in \Gamma_p} z_k| \leq \frac{\sqrt{T} \mu p - 1}{8Q}$ holds with probability at least $1 - L^{-\alpha}$. Combining (65) and (66) gives

$$\mathbb{P}\left(\|b_{1,l}^* S_{1,p+1}\| \geq \frac{\sqrt{T} \mu p - 1}{4Q}\right) \leq \mathbb{P}\left(\|\sum_{k \in \Gamma_p} z_k\| \geq \frac{\sqrt{T} \mu p - 1}{8Q}\right) + \mathbb{P}\left(\|\sum_{k \in \Gamma_p} z_k\| \geq \frac{\sqrt{T} \mu p - 1}{8Q}\right) \leq 2L^{-\alpha},$$

if $Q \geq C_\alpha \max\{\mu_{\max}^2 K, \mu_h^2 N\} \log^2 L$.

**Step 2: Proof of** $\|b_{1,l}^* S_{1,p+1}\| \leq \frac{\sqrt{T} \mu p - 1}{4Q}$: For any fixed $l \in \Gamma_{p+1}$,

$$b_{1,l}^* S_{1,p+1} = b_{1,l}^* S_{1,p+1} \sum_{j \neq 1} \mathcal{P}_{\Pi_1} \left(\sum_{k \in \Gamma_p} b_{1,k} w_{j,k}^* a_{j,k} a_{1,k}^*\right).$$

Now we rewrite $b_{1,l}^* S_{1,p+1}$ into

$$b_{1,l}^* S_{1,p+1} = \sum_{j \neq 1} \left(\sum_{k \in \Gamma_p} z_{j,k}^* + z_{j,k} x_{1,k}^*\right)$$

where

$$z_{j,k} := b_{1,l}^* S_{1,p+1} h_1 h_{1,k}^* b_{1,k} w_{j,k}^* a_{j,k} a_{1,k},$$

$$z_{j,k} := b_{1,l}^* S_{1,p+1} (I - h_1 h_{1,k}^*) b_{1,k} w_{j,k}^* a_{j,k} a_{1,k} x_1.$$

By the triangle inequality,

$$\|b_{1,l}^* S_{1,p+1}\| \leq \sum_{j \neq 1, j \leq r} \left[\|\sum_{k \in \Gamma_p} z_{j,k}\| + |\sum_{k \in \Gamma_p} z_{j,k}|\right].$$

In order to bound $\|b_{1,l}^* S_{1,p+1}\|$ by $\frac{\sqrt{T} \mu p - 1}{4Q}$, it suffices to prove that for all $1 \leq j \leq r$,

$$\left|\sum_{k \in \Gamma_p} z_{j,k}\right| \leq \frac{\sqrt{T} \mu p - 1}{8rQ},$$

(71)

For $\sum_{k \in \Gamma_p} z_{j,k}$,

$$\|z_{j,k}\| \leq \|b_{1,l}^* S_{1,p+1} h_1 h_{1,k}^* b_{1,k} w_{j,k}^* a_{j,k} a_{1,k}\| \leq C \sqrt{\frac{T \mu_h}{Q}} \sqrt{\frac{\mu_h}{Q}} \|w_{j,k}\| \leq C \sqrt{\frac{\mu_h^2}{Q}} \|w_{j,k}\|,$$

where $(\|w_{j,k}^* a_{j,k}\| \cdot |a_{1,k}^*|) \psi_1 \leq C \sqrt{\mathcal{N}_1} \|w_{j,k}\|$ follows from Lemma 7.5. Now we move on to the estimation of $\sigma^2$. 

32
Thus and similarly,
\[
\left\| \sum_{k \in \Gamma_p} \mathbb{E} z_{j,k}^* z_{j,k} \right\| = \sum_{k \in \Gamma_p} \left| b_{1,t}^* S_{1,p+1} h_1 h_1^* b_{1,k} \right|^2 \mathbb{E} \left[ \left| w_{j,k}^* a_{j,k} \right|^2 \right] \\
= N_1 \sum_{k \in \Gamma_p} \left| b_{1,t}^* S_{1,p+1} h_1 h_1^* b_{1,k} \right|^2 \left\| w_{j,k} \right\|^2 \\
\leq N_1 \frac{L \mu_2^2}{Q^2} \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \left\| T_{1,p} \right\| \\
\leq \frac{5 \mu_2^2 N_1 \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2}{4Q}
\]
and similarly,
\[
\left\| \sum_{k \in \Gamma_p} \mathbb{E} z_{j,k} z_{j,k}^* \right\| \leq \frac{5 \mu_2^2 \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2}{4Q}.
\]
Thus \( \sigma^2 \leq C \frac{\mu_2^2 N_1 \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2}{Q} \). By applying Bernstein inequality \( \text{[84]} \), we have
\[
\left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \leq C \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \max \left\{ \sqrt{\frac{\alpha \mu_2^2 N}{Q} \log L}, \frac{\alpha \mu_2^2 N}{Q} \log^2 L \right\}
\]
where \( \max_{k \in \Gamma_p} \left\| w_{j,k} \right\| \leq \frac{\sqrt{L \mu_{p-1}}}{Q} \). Choosing \( Q \geq C \alpha^2 \mu_2^2 N \log^2 L \) leads to
\[
\left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \leq \frac{\sqrt{L \mu_{p-1}}}{8rQ} \tag{72}
\]
with probability at least \( 1 - L^{-\alpha} \) for a fixed \( j : 1 \leq j \leq r \).

For \( \sum_{k \in \Gamma_p} z_{j,k} \) defined in \( \text{[69]} \) and fixed \( j \),
\[
R : = \max \left| z_{j,k} \right| \leq \max_{k \in \Gamma_p} \left\| b_{1,t}^* S_{1,p+1}(I_K - h_1 h_1^*) b_{1,k} \right\| \cdot \max_{k \in \Gamma_p} \left\| w_{j,k}^* a_{j,k} a_{j,k}^* x_1 \right\|_{\psi_1}
\leq C \mu_2^2 K_1 \frac{4L}{3Q} \max_{k \in \Gamma_p} \left\| w_{j,k} \right\| = C \mu_2^2 K_1 \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|\]
where \( \left\| w_{j,k}^* a_{j,k} a_{j,k}^* x_1 \right\|_{\psi_1} \leq C \left\| w_{j,k} \right\| \) follows from Lemma \( \text{[75]} \). Now we proceed to compute the variance by
\[
\sigma^2 := \sum_{k \in \Gamma_p} \mathbb{E} \left| z_{j,k} \right|^2 = \sum_{k \in \Gamma_p} \left| b_{1,t}^* S_{1,p+1}(I_K - h_1 h_1^*) b_{1,k} \right|^2 \left\| w_{j,k} \right\|^2 \\
\leq \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \left| b_{1,t}^* S_{1,p+1}(I_K - h_1 h_1^*) T_{1,p}(I_K - h_1 h_1^*) S_{1,p+1} b_{1,t} \right|^2 \\
\leq \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \left\| S_{1,p+1} \right\|^2 \left\| T_{1,p} \right\|^2 \left\| b_{1,t} \right\|^2 \\
\leq \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \frac{16L^2 5Q}{9Q^2} \frac{\mu_2^2 K_1}{4L} \leq C \frac{\max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \mu_2^2 K_1}{Q}
\]
Then we apply Bernstein inequality to get an upper bound of \( \left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \) for fixed \( j \),
\[
\left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \leq C \max_{k \in \Gamma_p} \left\| w_{j,k} \right\|^2 \max \left\{ \sqrt{\frac{\alpha \mu_2^2 K}{Q} \log L}, \frac{\alpha \mu_2^2 K}{Q} \log^2 L \right\} \leq \frac{\sqrt{L \mu_{p-1}}}{8rQ}
\]

33
with probability $1 - L^{-\alpha}$ if $Q \geq C_\alpha r^2 \mu_{\max}^2 K \log^2 L$. Thus combined with (72), we have proven that for fixed $j$,

$$\left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| + \left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \leq \frac{\sqrt{L} \mu_{p-1}}{4rQ}$$

holds with probability at least $1 - 2L^{-\alpha}$. By taking union bound over $1 \leq j \leq r$ and using (70), we can conclude that

$$\left\| b_{1,i}^* S_{1,p+1} \Pi_2 \right\| \leq \frac{\sqrt{L} \mu_{p-1}}{4Q}$$

with probability $1 - rL^{-\alpha}$ if $Q \geq C_\alpha r^2 \mu_{\max}^2 K \log^2 L$.

**Final step: Proof of (60)** To sum up, we have already shown that for fixed $l \in \Gamma_p$ and $i = 1$,

$$\frac{Q}{\sqrt{L}} \left\| b_{1,i}^* S_{1,p+1} W_{1,p} \right\| \leq \left\| \sum_{k \in \Gamma_p} z_k \right\| + \left\| \sum_{k \in \Gamma_p} z_k \right\| + \left\| \sum_{j \neq 1} \left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| + \left\| \sum_{k \in \Gamma_p} z_{j,k} \right\| \right\| \leq \frac{1}{2} \mu_{p-1}$$

with probability at least $1 - (r + 2)L^{-\alpha}$ if $Q \geq C_\alpha r^2 \max\{\mu_{\max}^2 K, \mu_{\max}^2 N\} \log^2 L$. Then we take union bound over all $1 \leq i \leq r$ and $l \in \Gamma_p$ and $1 \leq p \leq P$ and obtain

$$\mathbb{P} \left( \frac{Q}{\sqrt{L}} \max_{i,l,p} \left\| b_{1,i}^* S_{1,p+1} W_{1,p} \right\| \geq \frac{1}{2} \mu_{p-1} \right) \geq 1 - r(r + 2)PQL^{-\alpha} = 1 - r(r + 2)L^{-\alpha + 1}.$$ 

If we choose a slightly larger $\alpha$ as $\tilde{\alpha} = \alpha + 2 \log r$, i.e., $Q \geq C_\alpha r^2 \max\{\mu_{\max}^2 K, \mu_{\max}^2 N\} \log^2 L \log(r + 1)$, then $\mu_p \leq \frac{1}{2} \mu_{p-1}$ holds for all $p$ with probability at least $1 - L^{-\alpha + 1}$.

5.5 Proof of Theorem 3.1

We now assemble the various intermediate and auxiliary results to establish Theorem 3.1. We recall that Theorem 3.1 follows immediately from Lemma 5.2 which in turn hinges on the validity of the conditions (28) and (29). Let us focus on condition (28) first, i.e., we need to show that

$$\max_{1 \leq i \leq r} \left\| \mathcal{P}_{T_i} A_i^* A_i \mathcal{P}_{T_i} - \mathcal{P}_{T_i} \right\| \leq \frac{1}{4}, \quad (73)$$

$$\max_{1 \leq i \neq k \leq r} \left\| \mathcal{P}_{T_i} A_i^* A_k \mathcal{P}_{T_k} \right\| \leq \frac{1}{4r}, \quad (74)$$

$$\max_{1 \leq i \leq r} \left\| A_i \right\| \leq \gamma. \quad (75)$$

Under the assumptions of Theorem 3.1, Proposition 5.3 ensures that condition (73) holds with probability at least $1 - L^{-\alpha + 1}$ if $Q \geq C_\alpha \max\{\mu_{\max}^2 K, \mu_{\max}^2 N\} \log^2 L \log(r + 1)$ where $K := \max K_i$ and $N := \max N_i$. Moving on to the incoherence condition (74), Proposition 5.7 implies that this condition holds with probability at least $1 - L^{-\alpha + 1}$ if $Q \geq C_\alpha r^2 \max\{\mu_{\max}^2 K, \mu_{\max}^2 N\} \log^2 L \log(r + 1)$. Furthermore, $\gamma$ in condition (75) is bounded by $\sqrt{N(\log NL/2 + \alpha \log L)}$ with probability $1 - rL^{-\alpha}$ according to Lemma 1 in [2]. We now turn our attention to condition (29). Under the assumption that properties (41) and (46) hold, Lemma 5.10 implies the first part of condition (29). The two properties (41) and (46) have been established in Propositions 5.3 and 5.7 respectively. The second part of the approximate dual certificate condition in (29) is established in Lemma 5.12 with the aid of Lemma 5.14 with probability at least $1 - 2L^{-\alpha + 1}$ if $Q \geq C_\alpha r^2 \max\{\mu_{\max}^2 K, \mu_{\max}^2 N\} \log^2 L \log(r + 1)$. 

34
By “summing up” all the probabilities of failure in each substep,

\[ \Pr(\hat{X}_i = X_i, \forall l \leq i \leq r) \geq 1 - 5L^{-\alpha + 1} \]

if \( Q \geq C_a r^2 \max\{\mu_{max}^2 K, \mu_h^2 N\} \log^2 L \log(r + 1) \). Since \( L = PQ \) and \( P \) is chosen to be greater than \( \log_2(5r\gamma) \), it suffices to let \( L \) yield:

\[ L \geq C_a r^2 \max\{\mu_{max}^2 K, \mu_h^2 N\} \log^2 L \log \gamma \log(r + 1) \]

with \( \gamma \leq \sqrt{N\log(NL/2) + \alpha \log L} \). Thus, the sufficient conditions stated in Lemma 5.2 are fulfilled with probability at least \( 1 - O(L^{-\alpha + 1}) \), hence Theorem 3.1 follows now directly from Lemma 5.2.

5.6 Stability theory – Proof of Theorem 3.3

Since we do not assume \( \{X_i\}_{i=1}^r \) are of the same size, notation will be an issue during the discussion. We introduce a few notations in order to make the derivations easier. Recall \( \sum_{i=1}^r A_i(Z_i) \) is actually a linear mapping from \( \mathbb{C}^{K_1 \times N_1} \oplus \cdots \oplus \mathbb{C}^{K_r \times N_r} \) to \( \mathbb{C}^L \). This linear operator can be easily written into matrix form: define \( \Phi := [\Phi_1 \cdots \Phi_r] \) with \( \Phi_i \in \mathbb{C}^{L \times K_i N_i} \) and \( \Phi \in \mathbb{C}^{L \times (\sum_{i=1}^r K_i N_i)} \) as

\[
\Phi_i \text{vec}(Z_i) := \text{vec}(A_i(Z_i)), \quad \Phi \begin{bmatrix} \text{vec}(Z_1) \\ \vdots \\ \text{vec}(Z_r) \end{bmatrix} := \text{vec}(\sum_{i=1}^r A_i(Z_i))
\]

where \( Z_i \in \mathbb{C}^{K_i \times N_i} \). The operation “\( \text{vec} \)” vectorizes a matrix into a column vector. \( \Phi \) and \( \Phi_i \) are well-defined here. It could be be shown by slightly modifying the proof of Lemma 2 in [2] that

\[
\Phi \Phi^* = \sum_{i=1}^r \Phi_i \Phi_i^* \in \mathbb{C}^{L \times L}
\]

is well conditioned, which means the largest and smallest eigenvalues of \( \Phi \Phi^* \), denoted by \( \lambda_{max}^2 \) and \( \lambda_{min}^2 \) respectively, are of the same scale. More precisely,

\[
0.48 \mu_{min}^2 \frac{\sum_{i=1}^r K_i N_i}{L} \leq \lambda_{min}^2 \leq \lambda_{max}^2 \leq 4.5 \mu_{max}^2 \frac{\sum_{i=1}^r K_i N_i}{L}
\]

with probability at least \( 1 - O(L^{-\alpha + 1}) \) if \( \sum_{i=1}^r K_i N_i \geq \frac{C}{\mu_{min}^2} L \log^2 L \) with \( \mu_{min}^2 \) defined in [9]. Note that \( \sum_{i=1}^r K_i N_i \) is usually much larger than \( L \) in applications.

Let \( E_i = \hat{X}_i - X_i \in \mathbb{C}^{K_i \times N_i}, 1 \leq i \leq r \) be the difference between \( \hat{X}_i \) and \( X_i \). Define

\[
e_i := \text{vec}(E_i), \quad e := \begin{bmatrix} e_1 \\ \vdots \\ e_r \end{bmatrix} \in \mathbb{C}^{(\sum_{i=1}^r K_i N_i) \times 1},
\]

where \( e \) is a long vector consisting of all \( e_i, 1 \leq i \leq r \). We also consider \( e \) being projected on \( \text{Ran}(\Phi^*) \), denoted by \( e_{\Phi} \),

\[
e_{\Phi} := \Phi^* (\Phi \Phi^*)^{-1} \Phi e
\]

where \( \Phi e = \sum_{i=1}^r \Phi_i e_i = \sum_{i=1}^r A_i(E_i) \). From [21], we know that

\[
\|\Phi e\|_F = \left\| \sum_{i=1}^r A_i(E_i) \right\|_F \leq \left\| \sum_{i=1}^r A_i(\hat{X}_i) - \hat{y} \right\|_F + \left\| \sum_{i=1}^r A_i(X_i) - \hat{y} \right\|_F \leq 2\eta
\]
since both \( \{X_i\}_{i=1}^{r} \) and \( \{X_i\}_{i=1}^{r} \) are inside the feasible set. Similarly, define \( e_{\Phi^+} := e - e_{\Phi} \in \text{Null}(\Phi) \) and denote \( H_i \in \mathbb{C}^{K_i \times N_i} \) and \( J_i \in \mathbb{C}^{K_i \times N_i} \), \( 1 \leq i \leq r \), as matrices satisfying

\[
e_{\Phi^+} := \begin{bmatrix} \text{vec}(H_1) \\ \vdots \\ \text{vec}(H_r) \end{bmatrix}, \quad e_{\Phi} := \begin{bmatrix} \text{vec}(J_1) \\ \vdots \\ \text{vec}(J_r) \end{bmatrix}
\] (78)

where \( \sum_{i=1}^{r} A_i(H_i) = \Phi e_{\Phi^+} = 0 \) and \( H_i + J_i = E_i \) follows from the definition of \( H_i \) and \( J_i \).

Define \( P_{T_i} \) as the projection matrix from \( \text{vec}(Z) \) to \( \text{vec}(P_{T_i}(Z)) \), as

\[
P_{T_i} \text{vec}(Z) = \text{vec}(P_{T_i}(Z)), \quad P_{T_i} \in \mathbb{C}^{(K_iN_i) \times (K_iN_i)}
\]

and

\[
P_T := \begin{bmatrix} P_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_{T_r} \end{bmatrix}, \quad P_T e_{\Phi^+} := \begin{bmatrix} \text{vec}(H_{1,T_1}) \\ \vdots \\ \text{vec}(H_{r,T_r}) \end{bmatrix}, \quad P_T e_{\Phi^+} := \begin{bmatrix} \text{vec}(H_{1,T_1}) \\ \vdots \\ \text{vec}(H_{r,T_r}) \end{bmatrix}.
\] (79)

Actually the definitions above immediately give the following equations:

\[
P_T e = \begin{bmatrix} P_{T_1} e_1 \\ \vdots \\ P_{T_r} e_r \end{bmatrix}, \quad P_T e_{\Phi^+} = \begin{bmatrix} \text{vec}(H_{1,T_1}) \\ \vdots \\ \text{vec}(H_{r,T_r}) \end{bmatrix}, \quad P_T e_{\Phi^+} = \begin{bmatrix} \text{vec}(H_{1,T_1}) \\ \vdots \\ \text{vec}(H_{r,T_r}) \end{bmatrix}.
\]

We will prove that if the observation \( \hat{y} \) is contaminated by noise, the minimizer \( \hat{X}_i \) to the convex program (21) yields,

\[
\|e\| \leq C r \lambda_{\text{max}} \sqrt{\max\{K, N\}} \lambda_{\text{min}}^{-1}(1 - \beta - 2r \gamma) \eta.
\]

**Proof:** The proof basically follows similar arguments as [2, 7]. First we decompose \( e \) into several linear subspaces. By using orthogonality and Pythagorean Theorem,

\[
\|e\|^2_F = \|e_{\Phi}\|^2 + \|P_T e_{\Phi^+}\|^2_F + \|P_{T^\perp} e_{\Phi^+}\|^2_F.
\] (80)

Following from (79), (33) and (31) gives an estimate of the second term in (80),

\[
\|P_T e_{\Phi^+}\|^2_F = \sum_{i=1}^{r} \|H_{i,T_i}\|^2_F \leq 2 \left( \sum_{i=1}^{r} A_i(H_{i,T_i}) \right)^2_F = 2 \left( \sum_{i=1}^{r} A_i(H_{i,T_i}) \right)^2_F \leq 2 \gamma^2 \sum_{i=1}^{r} \|H_{i,T_i}\|^2_F \leq 2r \gamma^2 \|P_{T^\perp} e_{\Phi^+}\|^2_F \leq 2r \lambda_{\text{max}}^2 \|P_{T^\perp} e_{\Phi^+}\|^2_F
\]

where \( \max \|A_i\| \leq \gamma, \lambda_{\text{max}}^2 \) is largest eigenvalue of \( \Phi \Phi^* \) and obviously \( \gamma \leq \lambda_{\text{max}} \). The second equality holds since \( \sum_{i=1}^{r} A_i(H_i) = 0 \). For the third term in (80), by reversing the arguments
in the proof of Lemma 5.2, we have
\[ \| P_{T^orth} e_{\Phi^*} \|_F = \sqrt{\sum_{i=1}^{r} \| H_{i,T^orth} \|_F^2} \leq \sum_{i=1}^{r} \| H_{i,T} \|_F \]
\[ \leq \frac{1}{1 - \beta - 2r\gamma\alpha} \sum_{i=1}^{r} \langle H_i, h_i x_i^* \rangle + \| H_{i,T} \|_s \]
\[ \leq \frac{1}{1 - \beta - 2r\gamma\alpha} \sum_{i=1}^{r} \| X_i + H_i \|_s - \| X_i \|_s \]
\[ \leq \frac{1}{1 - \beta - 2r\gamma\alpha} \sum_{i=1}^{r} \| X_i + H_i \|_s - \| \hat{X}_i \|_s \]
where the first equality comes from (79), the third inequality is due to Lemma 5.1 and the last inequality follows from \( \sum_{i=1}^{r} \| X_i \|_s \leq \sum_{i=1}^{r} \| X_i \|_s \) in (22). From the definition of \( H_i \) and \( J_i \) in (78), \( \hat{X}_i = X_i + E_i + H_i + J_i \) and triangle inequality gives,
\[ \| P_{T^orth} e_{\Phi^*} \|_F \leq \frac{1}{1 - \beta - 2r\gamma\alpha} \sum_{i=1}^{r} \| J_i \|_s \leq \frac{\sqrt{\max\{K,N\}}}{1 - \beta - 2r\gamma\alpha} \sum_{i=1}^{r} \| J_i \|_F. \]
In other words,
\[ \| P_{T^orth} e_{\Phi^*} \|^2_F \leq \frac{r \max\{K,N\}}{(1 - \beta - 2r\gamma\alpha)^2} \sum_{i=1}^{r} \| J_i \|^2_F \leq \frac{r \max\{K,N\}}{(1 - \beta - 2r\gamma\alpha)^2} \| e_{\Phi} \|^2_F \]
(81)
where \( \| e_{\Phi} \|^2_F = \sum_{i=1}^{r} \| J_i \|^2_F \) follows from (78). By combining all those estimations together, i.e., \( \| P_{T^orth} e_{\Phi^*} \|^2_F \leq 4r \lambda_{max}^2 \| P_{T^orth} e_{\Phi^*} \|^2_F \), (81) and (80), we arrive at
\[ \| e \|^2_F \leq \| e_{\Phi} \|^2_F + (2r \lambda_{max}^2 + 1) \| P_{T^orth} e_{\Phi^*} \|^2_F \]
\[ \leq C r^2 \lambda_{max}^2 \| \hat{X} \|_s \| e_{\Phi} \|^2_F \]
where \( \lambda_{min}^2 \) is the smallest eigenvalue of \( \Phi \Phi^* \). By applying \( \| \Phi e \| \leq 2\eta \) in (77), we have
\[ \| e \|_F \leq C r^{\lambda_{max}} \sqrt{\max\{K,N\}} \lambda_{min} \| \Phi e \|_F \leq C r \lambda_{max} \sqrt{\max\{K,N\}} \lambda_{min} (1 - \beta - 2r\gamma\alpha)^{\eta} \eta. \]
In particular, if we choose \( \alpha = (5r\gamma)^{-1} \) and \( \beta = \frac{1}{2} \) according to Lemma 5.2 then \( \frac{1}{1 - \beta - 2r\gamma\alpha} = 10. \)
This completes the proof of Theorem 3.3.

6 Conclusion

We have developed a theoretical and numerical framework for simultaneously blindly deconvolve and demix multiple transmitted signals from just one received signal. The reconstruction of the transmitted signals and the impulse responses can be accomplished by solving a semidefinite program. Our findings are of interest for a variety of applications, in particular for the field of multiuser wireless communications. Our theory provides a bound for the number of
measurements needed to guarantee successful recovery. While this bound scales quadratically in the number of unknown signals, it seems that our theory is somewhat pessimistic. Indeed, numerical experiments indicate, surprisingly, that the proposed algorithm succeeds already even if the number of measurements is fairly close to the theoretical limit with respect to the number of degrees of freedom. It would be very desirable to develop a theory that can explain this remarkable phenomenon.

Hence, this paper does not only provide answers, but it also triggers numerous follow-up questions. Some key questions are: (i) Can we derive a theoretical bound that scales linearly in $r$, rather than quadratic in $r$ as our current theory? (ii) Is it possible to develop satisfactory theoretical bounds for deterministic matrices $A_i$? (iii) Do there exist faster numerical algorithms that do not need to resort to solving a semidefinite program (say in the style of the phase retrieval Wirtinger-Flow algorithm [6]) with provable performance guarantees? (iv) Can we develop a theoretical framework where the signals $x_i$ belong to some non-linear subspace, e.g. for sparse $x_i$? (v) How do the relevant parameters change when we have multiple (but less than $r$) receive signals? Answers to these questions could be particularly relevant in connection with the future Internet-of-Things.

Acknowledgement

The authors want to thank Holger Boche and Benjamin Friedlander for insightful discussions on the topic of the paper.

7 Appendix

7.1 Some useful auxiliary results

The key concentration inequality we use throughout our paper comes from Proposition 2 in [21, 22].

**Theorem 7.1.** Consider a finite sequence of $Z_l$ of independent centered random matrices with dimension $M_1 \times M_2$. Assume that $\|Z_l\|_{\psi_1} \leq R$ where the norm $\|\cdot\|_{\psi_1}$ of a matrix is defined as

$$\|Z\|_{\psi_1} := \inf_{u \geq 0} \{E[\exp(\|Z\|/u)] \leq 2\}. \quad (82)$$

and introduce the random matrix

$$S = \sum_{l=1}^Q Z_l.$$

Compute the variance parameter

$$\sigma^2 = \max \left\{ \| \sum_{l=1}^Q E(Z_lZ_l^*) \|, \| \sum_{l=1}^Q E(Z_l^*Z_l) \| \right\}, \quad (83)$$

then for all $t \geq 0$, we have the tail bound on the operator norm of $S$,

$$\|S\| \leq C_0 \max \{ \sigma \sqrt{t + \log(M_1 + M_2)}, R \log \left( \frac{\sqrt{QR}}{\sigma} \right) (t + \log(M_1 + M_2)) \} \quad (84)$$

with probability at least $1 - e^t$ where $C_0$ is an absolute constant.

---

4After completion of this manuscript we have developed such an algorithm for the case $r = 1$, see [23], but the case $r > 1$ is still open at this point in time.
For convenience we collect some results used throughout the proofs. Before we proceed, we note that there is a quantity equivalent to $\| \cdot \|_{\psi_1}$ defined in (82), i.e.,

$$c_1 \sup_{q \geq 1} q^{-1} (E |z|^q)^{1/q} \leq \|z\|_{\psi_1} \leq c_2 \sup_{q \geq 1} q^{-1} (E |z|^q)^{1/q},$$

(85)

where $c_1$ and $c_2$ are two universal positive constants, see Section 5.2.4 in [41]. Therefore, $\sup_{q \geq 1} q^{-1} (E |z|^q)^{1/q}$ will be used to quantify $\|z\|_{\psi_1}$ in this section since it is easier to use in explicit calculations.

**Lemma 7.2.** Let $z$ be a random variable which obeys $\mathbb{P}\{|z| > u\} \leq ae^{-bu}$, then

$$\|z\|_{\psi_1} \leq (1 + a)/b$$

which is proven in Lemma 2.2.1 in [39]. Moreover, it is easy to verify that for a scalar $\lambda \in \mathbb{C}$

$$\|\lambda z\|_{\psi_1} = |\lambda| \|z\|_{\psi_1}.$$

For another independent random variable $w$ with an exponential tail

$$\|z + w\|_{\psi_1} \leq C (\|z\|_{\psi_1} + \|w\|_{\psi_1})$$

(86)

**Proof:** We only prove (86) by using the equivalent quantity introduced in (85). Suppose that both $z$ and $w$ yield (85), there holds

$$\|z + w\|_{\psi_1} \leq c_2 \sup_{q \geq 1} q^{-1} (E |z + w|^q)^{1/q}$$

$$\leq c_2 \sup_{q \geq 1} q^{-1} \left[ (E |z|^q)^{1/q} + (E |w|^q)^{1/q} \right]$$

$$\leq c_1 c_2 (\|z\|_{\psi_1} + \|w\|_{\psi_1}),$$

where the second inequality follows from triangle inequality on $L^p$ spaces.

**Lemma 7.3.** Let $\mathbf{u} \in \mathbb{R}^n \sim \mathcal{N}(0, I_n)$, then $\|\mathbf{u}\|^2 \sim \chi_n^2$ and

$$\|\|\mathbf{u}\|^2\|_{\psi_1} = \|\langle \mathbf{u}, \mathbf{u} \rangle\|_{\psi_1} \leq 2n.$$

(87)

Furthermore,

$$E \left[ (\mathbf{u}^* - I_n) \mathbf{u}^2 \right] = (n + 1) I_n.$$

(88)

**Lemma 7.4** (Lemma 10-13 in [2]). Let $\mathbf{u} \in \mathbb{R}^n \sim \mathcal{N}(0, I_n)$ and $\mathbf{q} \in \mathbb{C}^n$ be any deterministic vector, then the following properties hold

$$|\langle \mathbf{u}, \mathbf{q} \rangle|^2 \sim \|\mathbf{q}\|^2 \chi_1^2,$$

(89)

$$\|\|\langle \mathbf{u}, \mathbf{q} \rangle\|^2\|_{\psi_1} \leq C \|\mathbf{q}\|^2,$$

(90)

$$\|\|\langle \mathbf{u}, \mathbf{q} \rangle\|^2 - \|\mathbf{q}\|^2\|_{\psi_1} \leq C \|\mathbf{q}\|^2,$$

(91)

$$\|\langle \mathbf{u}^* - I_n \rangle \mathbf{q}\|_{\psi_1} \leq C \sqrt{n} \|\mathbf{q}\|,$$

(92)

$$E \left[ (\mathbf{u}^* - I_n) \mathbf{q} \mathbf{q}^* (\mathbf{u}^* - I_n) \right] = \|\mathbf{q}\|^2 I_n + \overline{\mathbf{q}} \mathbf{q}^*.$$

(93)

Let $\mathbf{p} \in \mathbb{C}^n$ be another deterministic vector, then

$$\|\langle \mathbf{u}, \mathbf{q} \rangle \langle \mathbf{p}, \mathbf{u} \rangle - \langle \mathbf{q}, \mathbf{p} \rangle\|_{\psi_1} \leq \|\mathbf{q}\| \|\mathbf{p}\|.$$

(94)
Proof: \((89)\) to \((92)\) and \((94)\) directly follow from Lemma 10-13 in [2], except for small differences in the constants. We only prove \((93)\) here.

\[
E[(uu^* - I_n)qq^*(uu^* - I_n)] = E[(|u, q|^2uu^*) - qq^*].
\]

For each \((i, j)\)-th entry of \(R_{ij} = \langle |u, q|^2u_iu_j = q^*u_iu_ju_j^* \rangle_q \).

\[
E[u_iu_juu^*] = \begin{cases} 
E_{ij} + E_{ji} & i \neq j \\
I_n + E_{ii} & i = j
\end{cases}
\]

where \(E_{ij}\) is an \(n \times n\) matrix with the \((i, j)\)-th entry equal to 1 and the others being 0. The expectation of \(R_{ij}\)

\[
E R_{ij} = \begin{cases} 
q_i^*q_j + q_*^*q_i & i \neq j \\
|q|^2 + |q_i|^2 & i = j
\end{cases}
\]

and

\[
E[|u, q|^2uu^*] - qq^* = ||q||^2I_n + qq^* + \bar{q}\bar{q}^* - qq^* = ||q||^2I_n + \bar{q}\bar{q}^*
\]

where \(\bar{q}\) is the complex conjugate of \(q\).

Lemma 7.5. Assume \(u \sim \mathcal{N}(0, I_n)\) and \(v \sim \mathcal{N}(0, I_m)\) are two independent Gaussian random vectors, then

\[
\|\|u\|^2 + |v|^2\|_{\psi_1} \leq n + m
\]

and

\[
\|u\| \cdot \|v\|_{\psi_1} \leq C\sqrt{mn}.
\]

Proof: Let us start with the first one.

\[
\|\|u\|^2 + |v|^2\|_{\psi_1} \leq \|\|u\|^2\|_{\psi_1} + \|\|u\|^2\|_{\psi_1} \leq n + m,
\]

which directly follows from \((86)\) and \((87)\). Following from independence,

\[
\|\|u\| \cdot \|v\|\|_{\psi_1} \leq c_2 \sup_{q \geq 1} q^{-1}(E\|u\|^q\|v\|^q)^{1/q} \leq c_2 \sup_{q} q^{-1}(E\|u\|^q)^{1/q}(E\|v\|^q)^{1/q}.
\]

Let \(t = q/2,\)

\[
\|\|u\| \cdot \|v\|\|_{\psi_1} \leq c_2 \sup_{t \geq 1} \frac{1}{2t}(E\|u\|^{2t})^{1/2t}(E\|v\|^{2t})^{1/2t} \leq \frac{c_2}{2} \left( \sup_{t \geq 1} \frac{1}{t} (E\|u\|^{2t})^{1/2t} \right)^{1/2} \left( \sup_{t \geq 1} \frac{1}{t} (E\|v\|^{2t})^{1/2t} \right)^{1/2} \leq \frac{c_1c_2}{2} \sqrt{\|u\|_{\psi_1} \cdot \|v\|_{\psi_1} \leq C\sqrt{mn},
\]

where \(\|u\|^2 \sim \chi_n^2\) and \(\|v\|^2 \sim \chi_m^2\) and \(\|u\|_{\psi_1}\) and \(\|v\|_{\psi_1}\) are given by \((87)\).

7.2 A useful fact about the “low-frequency” DFT matrix

Suppose that \(B\) is a “low-frequency” Fourier matrix, i.e.,

\[
B = \frac{1}{\sqrt{L}}(e^{-2\pi i l/k})_{l,k} \in \mathbb{C}^{L \times K},
\]

where \(1 \leq k \leq K\) and \(1 \leq l \leq L\) with \(K \leq L\). Assume there exists a \(Q\) such that \(L = QP\) with \(Q \geq K\). We choose \(\Gamma_p = \{p, P + p, \ldots, (Q - 1)P + p\}\) with \(1 \leq p < P\) such that \(|\Gamma_p| = Q,\)
\[ \bigcup_{1 \leq p \leq P} \Gamma_p = \{1, \cdots, L\} \] and they are mutually disjoint. Let \( B_p \) be the \( Q \times K \) matrix by choosing its rows from those of \( B \) with indices in \( \Gamma_p \). Then we can rewrite \( B_p \) as

\[
B_p = \frac{1}{\sqrt{L}} (e^{-2\pi i (tP - P + p)k/(PQ)})_{1 \leq t \leq Q, 1 \leq k \leq K} \in \mathbb{C}^{Q \times K},
\]

and it actually equals

\[
B_p = \frac{1}{\sqrt{L}} (e^{-2\pi i k/Q} e^{2\pi i (P - p)k/(PQ)})_{1 \leq t \leq Q, 1 \leq k \leq K} \in \mathbb{C}^{Q \times K}.
\]

Therefore

\[
B_p = \sqrt{\frac{Q}{L}} F_Q \text{diag}(e^{2\pi i (P - p)/Q}, \ldots, e^{2\pi i K(P - p)/Q})
\]

where \( F_Q \) is the first \( K \) columns of a \( Q \times Q \) DFT matrix with \( F_Q^* F_Q = I_K \). There holds

\[
\sum_{l \in \Gamma_p} b_l b_l^* = B_p^* B_p = \frac{Q}{L} I_K
\]

where \( b_l \) is the \( l \)-th column of \( B^* \).

References


