Blind Deconvolution Meets Blind Demixing: Algorithms and Performance Bounds *

Shuyang Ling and Thomas Strohmer
Department of Mathematics
University of California at Davis
Davis CA 95616
{syling,strohmer}@math.ucdavis.edu

December 23, 2015

Abstract

Consider r sensors, each one intends to send a function $\boldsymbol{x}_i$ (e.g. a signal or image) to a receiver common to all r sensors. Before transmission, each $\boldsymbol{x}_i$ is multiplied by an "encoding matrix" $\boldsymbol{A}_i$. During transmission each $\boldsymbol{A}_i\boldsymbol{x}_i$ gets convolved with a function $\boldsymbol{h}_i$. The receiver records the function $\boldsymbol{y}$, given by the sum of all these convolved signals. Assume that the receiver knowns all the $\boldsymbol{A}_i$, but does neither know the $\boldsymbol{x}_i$ nor the $\boldsymbol{h}_i$. When and under which conditions is it possible to recover the individual signals $\boldsymbol{x}_i$ and the channels $\boldsymbol{h}_i$ from just one received signal $\boldsymbol{y}$? 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 $\boldsymbol{x}_i$ and the impulse responses $\boldsymbol{h}_i$ in a robust, reliable, and efficient manner from just one single received function $\boldsymbol{y}$ 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 $\boldsymbol{z}_i$ (e.g. a signal or image) to a receiver common to all r sensors. During transmission each $\boldsymbol{z}_i$ gets convolved with a function $\boldsymbol{g}_i$ (the $\boldsymbol{g}_i$ may all differ from each other). The receiver records the function $\boldsymbol{y}$, given by the sum of all these convolved signals. More precisely,

$$\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{g}_i * \boldsymbol{z}_i + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{w}$ is additive noise. Assume that the receiver does neither know the $\boldsymbol{z}_i$ nor the $\boldsymbol{g}_i$. When and under which conditions is it possible to recover all the individual signals $\boldsymbol{z}_i$ and $\boldsymbol{g}_i$ from just one received signal $\boldsymbol{y}$?

*This research was partially supported by the NSF via grant DTRA-DMS 1322393.

Blind deconvolution 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 [26], image processing [33, 31], neuroscience [36], spectroscopy [37], astronomy [12]. It also arises in wireless communications[1] [41] and is expected to play a central role in connection with the future Internet-of-Things [44]. 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 $\boldsymbol{z}_i \in \mathbb{R}^N$ lies in a known subspace of dimension N, i.e., there exists matrices $\boldsymbol{A}_i$ of size $L \times N$ such that $\boldsymbol{z}_i = \boldsymbol{A}_i \boldsymbol{x}_i$. In addition the matrices $\boldsymbol{A}_i$ need to satisfy a certain "local" mutual incoherence condition described in detail in (25). This condition can be satisfied if the $\boldsymbol{A}_i$ 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 noisefree scenario, that is, $\boldsymbol{w} = 0$. Below and throughout the paper $*$ denotes circular convolution.

Theorem 1.1 (Informal version). *Let $\boldsymbol{x}_i \in \mathbb{R}^N$ and let the $\boldsymbol{A}_i$ be $L \times N$ i.i.d. Gaussian random matrices. Furthermore, assume that the impulse responses $\boldsymbol{g}_i \in \mathbb{C}^N$ have* maximum delay spread *K, i.e., for each $\boldsymbol{g}_i$ there holds $\boldsymbol{g}_i(k) = 0$ if $k \geq K$. Let μ_h^2 be a certain "incoherence parameter" related to the measurement matrices, defined in* (13). *Suppose we are given*

$$\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{g}_i * (\boldsymbol{A}_i \boldsymbol{x}_i). \tag{2}$$

Then, as long as the number of measurements L satisfies

$$L \gtrsim C r^2 \max\{K, \mu_h^2 N\} \log^3 L \log(r+1),$$

(where C is a numerical constant), all $\boldsymbol{x}_i$ (and thus $\boldsymbol{z}_i = \boldsymbol{A}_i \boldsymbol{x}_i$) as well as all $\boldsymbol{g}_i$ can be recovered from $\boldsymbol{y}$ with high probability by solving a semidefinite program.

Recovering $\{\boldsymbol{x}_i\}_{i=1}^r$ and $\{\boldsymbol{g}_i\}_{i=1}^r$ is only possible up to a constant, since we can always multiply each $\boldsymbol{x}_i$ with $\alpha_i \neq 0$ and each $\boldsymbol{g}_i$ with $1/\alpha_i$ and still get the same result. Hence, here and throughout the paper, recovery of the vectors $\boldsymbol{x}_i$ and $\boldsymbol{g}_i$ always means recovery modulo constants α_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, 41, 26, 33, 32, 23, 31, 36, 37, 12, 44]. 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 $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$. 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 [35]. However, in this paper we are interested in the case where we have only one single received signal $\boldsymbol{y}$ – a single snapshot, in the jargon of array processing. Hence, there is little overlap between these methods heavily relying on multiple snapshots (manu 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 [39]. There, the matrix $\boldsymbol{A}_i$ represents the spreading matrix assigned to the i-th user and $\boldsymbol{g}_i$ models the associated multipath channel. There are numerous papers on blind channel estimation in connection with CDMA, including the previously cited articles [17, 41, 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 $\boldsymbol{g}_i$ 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 $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$. 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 7) in the "multi-user" scenario.

The paper [1] considers the following generalization of [2][2]. Assume that we are given signals $\boldsymbol{y}_i = \boldsymbol{g} * \boldsymbol{x}_i, i = 1, \dots, r$, the goal is to recover the $\boldsymbol{x}_i$ and $\boldsymbol{g}$ from $\boldsymbol{y}_1, \dots, \boldsymbol{y}_r$. This setting is somewhat in the spirit of (1), but it is significantly less challenging, since (i) it assumes the same convolution function $\boldsymbol{g}$ for each signal $\boldsymbol{x}_i$ and (ii) there are as many output signals $\boldsymbol{y}_i$ as we have input signals $\boldsymbol{x}_i$.

Non-blind versions of (1) or (2) can be found for instance in [43, 28, 27, 3]. In the very interesting paper [43], the authors analyze various problems of decomposing a given observation into multiple incoherent components, which can be expressed as

$$\text{minimize} \ \sum_i \lambda_i \|\boldsymbol{X}_i\|_{(i)} \qquad \text{subject to} \quad \sum_i \boldsymbol{X}_i = \boldsymbol{M}. \tag{3}$$

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 [43]. 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.

[2]Since 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 [25] to the multi-sensor case. In this context, we also refer to related (single-input-single-output) analysis in [24, 14].

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. Sections 4–9 are devoted to the proofs of these results. Numerical experiments can be found in Section 10. We conclude in Section 11 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 $\boldsymbol{Z}$ and $\boldsymbol{z}$. The individual entries of a matrix or a vector are denoted in normal font such as Z_{ij} or z_i. For any matrix $\boldsymbol{Z}$, $\|\boldsymbol{Z}\|_*$ denotes nuclear norm, i.e., the sum of its singular values; $\|\boldsymbol{Z}\|$ denotes operator norm, i.e., its largest singular value, and $\|\boldsymbol{Z}\|_F$ denotes the Frobenius norm, i.e., $\|\boldsymbol{Z}\|_F = \sqrt{\sum_{ij}|Z_{ij}|^2}$. For any vector $\boldsymbol{z}$, $\|\boldsymbol{z}\|$ denotes its Euclidean norm. For both matrices and vectors, $\boldsymbol{Z}^T$ and $\boldsymbol{z}^T$ stand for the transpose of $\boldsymbol{Z}$ and $\boldsymbol{z}$ respectively while $\boldsymbol{Z}^*$ and $\boldsymbol{z}^*$ denote their complex conjugate transpose. $\bar{z}$ and $\bar{\boldsymbol{z}}$ denote the complex conjugate of z and $\boldsymbol{z}$ respectively. We equip the matrix space $\mathbb{C}^{K\times N}$ with the inner product defined as $\langle \boldsymbol{U}, \boldsymbol{V}\rangle := \operatorname{Tr}(\boldsymbol{U}\boldsymbol{V}^*)$. A special case is the inner product of two vectors, i.e., $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \operatorname{Tr}(\boldsymbol{u}\boldsymbol{v}^*) = \boldsymbol{v}^*\boldsymbol{u} = (\boldsymbol{u}^*\boldsymbol{v})^*$. The identity matrix of size n is denoted by $\boldsymbol{I}_n$. For a given vector $\boldsymbol{v}$, $\operatorname{diag}(\boldsymbol{v})$ represents the diagonal matrix whose diagonal entries are given by the vector $\boldsymbol{v}$.
Throughout the paper, C stands for a constant and C_α is a constant which depends linearly on α (and on no other numbers). For the two linear subspaces T_i and $T_i^\perp$ defined in (23) and (24), we denote the projection of $\boldsymbol{Z}$ on T_i and $T_i^\perp$ as $\boldsymbol{Z}_{T_i} := \mathcal{P}_{T_i}(\boldsymbol{Z})$ and $\boldsymbol{Z}_{T_i^\perp} := \mathcal{P}_{T_i^\perp}(\boldsymbol{Z})$ 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 [25] involving multiple sensors. We consider the following setup[3]

$$\boldsymbol{y} = \sum_{i=1}^{r} \operatorname{diag}(\boldsymbol{B}_i\boldsymbol{h}_i)\boldsymbol{A}_i\boldsymbol{x}_i, \tag{4}$$

where $\boldsymbol{y}\in\mathbb{C}^L$, $\boldsymbol{B}_i\in\mathbb{C}^{L\times K_i}$, $\boldsymbol{A}_i\in\mathbb{R}^{L\times N_i}$, $\boldsymbol{h}_i\in\mathbb{R}^{K_i}$ and $\boldsymbol{x}_i\in\mathbb{R}^{N_i}$. We assume that all the matrices $\boldsymbol{B}_i$ and $\boldsymbol{A}_i$ are given, but none of the $\boldsymbol{x}_i$ and $\boldsymbol{h}_i$ are known. Note that all $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ can be of different lengths. We point out that the total number of measurements is given by the length of $\boldsymbol{y}$, 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 (1) as a special case, as we will explain in Section 3. But it also includes other cases as well. Consider for instance a linear

[3] In (4) 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.

system $\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{A}_i(\theta_i)\boldsymbol{x}_i$, where the measurement matrices $\boldsymbol{A}_i$ are not fully known due to lack of calibration [16, 4, 25] and θ_i represents the unknown calibration parameters associated with $\boldsymbol{A}_i$. 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 $\boldsymbol{A}_i$. In other words, the norms of each of the rows of $\boldsymbol{A}_i$ are unknown. If in addition each of the θ_i belongs to a known subspace represented by $\boldsymbol{B}_i$, i.e., $\theta_i = \boldsymbol{B}_i\boldsymbol{h}_i$, then we can write such an $\boldsymbol{A}_i(\theta_i)$ as $\boldsymbol{A}_i(\theta_i) = \operatorname{diag}(\boldsymbol{B}_i\boldsymbol{h}_i)\boldsymbol{A}_i$.

Let $\boldsymbol{b}_{i,l}$ denote the l-th column of $\boldsymbol{B}_i^*$ and $\boldsymbol{a}_{i,l}$ the l-th column of $\boldsymbol{A}_{i,l}^T$. A simple application of linear algebra gives

$$y_l = \sum_{i=1}^{r} (\boldsymbol{B}_i\boldsymbol{h}_i)_l \boldsymbol{x}_i^T \boldsymbol{a}_{i,l} = \sum_{i=1}^{r} \boldsymbol{b}_{i,l}^* \boldsymbol{h}_i \boldsymbol{x}_i^T \boldsymbol{a}_{i,l}. \tag{5}$$

where y_l is the l-th entry of $\boldsymbol{y}$. One may find an obvious difficulty of this problem as the nonlinear relation between the measurement vectors $(\boldsymbol{b}_{i,l}, \boldsymbol{a}_{i,l})$ and the unknowns $(\boldsymbol{h}_i, \boldsymbol{x}_i)$. Proceeding with the meanwhile well-established lifting trick [10], we let $\boldsymbol{X}_i := \boldsymbol{h}_i\boldsymbol{x}_i^T \in \mathbb{R}^{K_i \times N_i}$ and define the *linear* mapping $\mathcal{A}_i : \mathbb{C}^{K_i \times N_i} \to \mathbb{C}^L$ for $i = 1, \dots, r$ by

$$\mathcal{A}_i(\boldsymbol{Z}) := \{\boldsymbol{b}_{i,l}^* \boldsymbol{Z} \boldsymbol{a}_{i,l}\}_{l=1}^{L}.$$

Note that the adjoint of $\mathcal{A}_i$ is

$$\mathcal{A}_i^* : \mathbb{C}^L \to \mathbb{C}^{K_i \times N_i}, \qquad \mathcal{A}_i^*(\boldsymbol{z}) = \sum_{l=1}^{L} z_l \boldsymbol{b}_{i,l} \boldsymbol{a}_{i,l}^*. \tag{6}$$

since $\mathbb{C}^{K_i \times N_i}$ is equipped with the inner product $\langle \boldsymbol{U}, \boldsymbol{V} \rangle = \operatorname{Tr}(\boldsymbol{U}\boldsymbol{V}^*)$ for any $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{C}^{K_i \times N_i}$. Thus we have lifted the *non-linear vector-valued* equations (4) to *linear matrix-valued* equations given by

$$\boldsymbol{y} = \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{X}_i). \tag{7}$$

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 [30, 15]. Hence, to combat this underdeterminedness, we attempt to recover $(\boldsymbol{h}_i, \boldsymbol{x}_i)_{i=1}^r$ by solving the following nuclear norm minimization problem,

$$\min \sum_{i=1}^{r} \|\boldsymbol{Z}_i\|_* \quad \text{subject to} \quad \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{Z}_i) = \boldsymbol{y}. \tag{8}$$

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

2.3 Incoherence conditions on the matrices $\boldsymbol{B}_i$

Analogous to matrix completion, where one needs to impose certain incoherence conditions on the singular vectors (see e.g. [5]), we introduce two quantities that describe a notion of incoherence of the matrices $\boldsymbol{B}_i$. We require $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$ and define

$$\mu_{\max}^2 := \max_{1 \le l \le L, 1 \le i \le r} \frac{L}{K_i} \|\boldsymbol{b}_{i,l}\|^2, \quad \mu_{\min}^2 := \min_{1 \le l \le L, 1 \le i \le r} \frac{L}{K_i} \|\boldsymbol{b}_{i,l}\|^2. \tag{9}$$

With a little knowledge of linear algebra, it is easy to show, using only $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$, that $1 \leq \mu^2_{\max} \leq \frac{L}{K_i}$ and $0 \leq \mu^2_{\min} \leq 1$. In particular, if each $\boldsymbol{B}_i$ is a partial DFT matrix then $\mu^2_{\max} = \mu^2_{\min} = 1$. The quantity $\mu^2_{\min}$ will be useful to establish Theorem 3.3, while the main purpose of introducing $\mu^2_{\max}$ is to quantify a "joint incoherence pattern" on all $\boldsymbol{B}_i$. 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

$$\max_{1\leq i\leq r, 1\leq p\leq P} \|\boldsymbol{T}_{i,p} - \frac{Q}{L}\boldsymbol{I}_{K_i}\| \leq \frac{Q}{4L}, \quad \text{where} \quad \boldsymbol{T}_{i,p} := \sum_{l\in\Gamma_p} \boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^*, \tag{10}$$

which says that each $\boldsymbol{T}_{i,p}$ does not deviate too much from $\boldsymbol{I}_{K_i}$. 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 $\boldsymbol{B}_i$, there exists a partition $\{\Gamma_{i,p}\}_{p=1}^P$ (where $\Gamma_{i,p}$ depends on i) such that

$$\max_{1\leq p\leq P} \|\sum_{l\in\Gamma_{i,p}} \boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^* - \frac{Q}{L}\boldsymbol{I}_{K_i}\| \leq \frac{Q}{4L}, \quad \forall 1 \leq i \leq r,$$

if $Q \geq C\mu^2_{\max}K_i \log L$ 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 $\boldsymbol{B}_i$ 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 $\boldsymbol{B}_i, i \neq j$ is a submatrix of $\boldsymbol{B}_j$, then we can simply let $\Gamma_p = \Gamma_{j,p}$ such that (10) holds.

3. If all $\boldsymbol{B}_i$ are "low-frequency" DFT matrices, i.e., the first K_i columns of an $L \times L$ DFT matrix with $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$, we can actually create an *explicit* partition of Γ_p such that

$$\boldsymbol{T}_{i,p} = \sum_{l\in\Gamma_p} \boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^* = \frac{Q}{L}\boldsymbol{I}_{K_i}. \tag{11}$$

   For example, suppose $L = PQ$ and $Q \geq K_i$, we can achieve $\boldsymbol{T}_{i,p} = \frac{Q}{L}\boldsymbol{I}_{K_i}$ and $|\Gamma_p| = Q$ by letting $\Gamma_p = \{p, P+p, \cdots, (Q-1)P+p\}$. A short proof will be provided in Section 12.2.

Some direct implications of (10) are

$$\|\boldsymbol{T}_{i,p}\| \leq \frac{5Q}{4L}, \quad \|\boldsymbol{S}_{i,p}\| \leq \frac{4L}{3Q}, \quad \forall 1 \leq i \leq r, 1 \leq p \leq P. \tag{12}$$

where $\boldsymbol{S}_{i,p} := \boldsymbol{T}_{i,p}^{-1}$. Now let us introduce the second incoherence quantity, which is also crucial in the proof of Theorem 3.1,

$$\mu^2_h := \max\left\{\frac{Q^2}{L} \max_{l\in\Gamma_p, 1\leq p\leq P, 1\leq i\leq r} \frac{|\langle \boldsymbol{S}_{i,p}\boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2}{\|\boldsymbol{h}_i\|^2}, \quad L \max_{1\leq l\leq L, 1\leq i\leq r} \frac{|\langle \boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2}{\|\boldsymbol{h}_i\|^2}\right\}. \tag{13}$$

The range of μ^2_h is given in Proposition 2.2.

Remark 2.1. *The attentive reader may have noticed that the definition of μ^2_h is a bit more intricate than the one in [2], where μ^2_h only depends on $|\langle \boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2$. 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 μ^2_h. Another easy way to fix the issue is to consider all $\boldsymbol{B}_i$ as low-frequency Fourier matrices. If so, μ^2_h in* (13) *reduces to a simpler form of μ^2_h, i.e., $\mu^2_h = L\max\{|\langle \boldsymbol{b}_l, \boldsymbol{h}\rangle|^2/\|\boldsymbol{h}\|^2\}$ in [2] because the explicit partition of low-frequency DFT matrices allows $\boldsymbol{T}_{i,p} = \frac{Q}{L}\boldsymbol{I}_{K_i}$ and $\boldsymbol{S}_{i,p} = \frac{L}{Q}\boldsymbol{I}_{K_i}$.*

Both $\mu_{\max}^2$ and μ_h^2 measure the incoherence of $\boldsymbol{B}_i$ and the latter one, depending $\boldsymbol{h}_i$, also describes the interplay between $\boldsymbol{h}_i$ and $\boldsymbol{B}_i$. To sum up, for all $1 \le l \le L$ and $1 \le i \le r$,

$$\|\boldsymbol{b}_{i,l}\|^2 \le \frac{\mu_{\max}^2 K_i}{L}, \quad |\langle \boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2 \le \frac{\mu_h^2}{L}\|\boldsymbol{h}_i\|^2, \quad |\langle \boldsymbol{S}_{i,p}\boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2 \le \frac{L\mu_h^2}{Q^2}\|\boldsymbol{h}_i\|^2. \tag{14}$$

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

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

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

$$\begin{aligned}
\mu_h^2 &\ge \max_{i,p}\left\{\frac{Q}{L}\sum_{l\in\Gamma_p}|\langle \boldsymbol{S}_{i,p}\boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2, \sum_{l=1}^{L}|\langle \boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2\right\} \\
&= \max_{i,p}\left\{\frac{Q}{L}\sum_{l\in\Gamma_p}\boldsymbol{h}_i^*\boldsymbol{S}_{i,p}\boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p}\boldsymbol{h}_i, \sum_{l=1}^{L}\boldsymbol{h}_i^*\boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^*\boldsymbol{h}_i\right\} \\
&= \max_{i,p}\left\{\frac{Q}{L}\boldsymbol{h}_i^*\boldsymbol{S}_{i,p}\boldsymbol{h}_i, 1\right\}.
\end{aligned}$$

Note that

$$1 \le \max_{i,p}\left\{\frac{Q}{L}\boldsymbol{h}_i^*\boldsymbol{S}_{i,p}\boldsymbol{h}_i, 1\right\} \le \frac{4}{3},$$

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

$$\begin{aligned}
\mu_h^2 &\le \max\left\{\frac{Q^2}{L}\max_{l\in\Gamma_p, 1\le p\le P, 1\le i\le r}|\langle \boldsymbol{S}_{i,p}\boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2, L\max_{1\le i\le L, 1\le i\le r}|\langle \boldsymbol{h}_i, \boldsymbol{b}_{i,l}\rangle|^2\right\} \\
&\le \max_{p,i,l}\left\{\frac{Q^2}{L}\|\boldsymbol{S}_{i,p}\|^2\|\boldsymbol{b}_{i,l}\|^2, L\|\boldsymbol{b}_{i,l}\|^2\right\} \\
&\le \frac{Q^2}{L}\frac{16L^2}{9Q^2}\cdot\frac{\mu_{\max}^2 K_i}{L} \le \frac{16}{9}\mu_{\max}^2 K_i.
\end{aligned}$$

where $\|\boldsymbol{S}_{i,p}\| \le \frac{4L}{3Q}$ and $\|\boldsymbol{b}_{i,l}\|^2 \le \frac{\mu_{\max}^2 K_i}{L}$. ■

2.4 Is the incoherence parameter μ_h^2 necessary?

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

$$y_l = \left\langle \begin{bmatrix} \boldsymbol{h}_1\boldsymbol{x}_1^T & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_2\boldsymbol{x}_2^T & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{h}_r\boldsymbol{x}_r^T \end{bmatrix}, \begin{bmatrix} \boldsymbol{b}_{1,l}\boldsymbol{a}_{1,l}^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{b}_{2,l}\boldsymbol{a}_{2,l}^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{b}_{r,l}\boldsymbol{a}_{r,l}^* \end{bmatrix} \right\rangle.$$

Recall that in matrix completion [5, 29] 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

$$\max_{1\leq l\leq L, 1\leq i\leq r} L|\langle \boldsymbol{b}_{i,l}, \boldsymbol{h}_i\rangle|^2/\|\boldsymbol{h}_i\|^2, \qquad \max_{1\leq l\leq L, 1\leq i\leq r} |\langle \boldsymbol{a}_{i,l}, \boldsymbol{x}_i\rangle|^2/\|\boldsymbol{x}_i\|^2 \tag{15}$$

are required to be small. We can ensure that the second term in (15) is small since each $\boldsymbol{a}_{i,l}$ 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 $\boldsymbol{h}_i$ (more precisely, the worst case could be $\mathcal{O}(K)$), hence we need to put a constraint on μ_h^2 in order to control its size. As numerical simulations presented in Section 10 show, the relevance of μ_h^2 goes beyond "proof-technical reasons". The required number of measurements for successful recovery does indeed depend on μ_h^2, see Figure 3, at least when using the suggested approach via semidefinite programing.

2.5 Conditions on the matrices $\boldsymbol{A}_i$

Throughout the proof of main theorem, we also need to be able to control a certain "mutual incoherence" of the matrices $\mathcal{A}_i$ on the subspaces T_i, cf. (25). This condition involves the quantity

$$\max_{j\neq k} \|\mathcal{P}_{T_j}\mathcal{A}_j^*\mathcal{A}_k\mathcal{P}_{T_k}\|. \tag{16}$$

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

$$\mathbb{E}(\mathcal{A}_i^*\mathcal{A}_i(\boldsymbol{Z}_i)) = \sum_{l=1}^L \boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^*\boldsymbol{Z}_i\,\mathbb{E}(\boldsymbol{a}_{i,l}\boldsymbol{a}_{i,l}^*) = \boldsymbol{Z}_i, \quad \boldsymbol{Z}_i\in\mathbb{C}^{K_i\times N_i},$$

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

$$\mathcal{A}_{i,p}(\boldsymbol{Z}_i) = \{\boldsymbol{b}_{i,l}^*\boldsymbol{Z}_i\boldsymbol{a}_{i,l}\}_{l\in\Gamma_p}, \quad \mathcal{A}_{i,p}^*(\boldsymbol{z}) = \sum_{l\in\Gamma_p} z_l\boldsymbol{b}_{i,l}\boldsymbol{a}_{i,l}^*. \tag{17}$$

The definition of $\mathcal{A}_{i,p}$ is the same as that of $\mathcal{A}_i$ except that $\mathcal{A}_{i,p}$ only uses a subset of all measurements. However, the expectation of $\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}$ is no longer the identity in general (except the case when all $\boldsymbol{B}_i$ are low-frequency DFT matrices and satisfy (11)), i.e.,

$$\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{Z}_i) = \sum_{l\in\Gamma_p}\boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^*\boldsymbol{Z}_i\boldsymbol{a}_{i,l}\boldsymbol{a}_{i,l}^*,$$

and

$$\mathbb{E}(\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{Z}_i)) = \boldsymbol{T}_{i,p}\boldsymbol{Z}_i, \quad \boldsymbol{T}_{i,p} := \sum_{l\in\Gamma_p}\boldsymbol{b}_{i,l}\boldsymbol{b}_{i,l}^* \tag{18}$$

The non-identity expectation of $\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}$ will create challenges throughout the proof. However, there is an easy trick to fix this issue. By properly assuming $Q > K_i$, $\boldsymbol{T}_{i,p}$ is actually invertible. Consider $\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{S}_{i,p}\boldsymbol{Z}_i)$ and its expectation now yields

$$\mathbb{E}(\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{S}_{i,p}\boldsymbol{Z}_i)) = \boldsymbol{T}_{i,p}\boldsymbol{S}_{i,p}\boldsymbol{Z}_i = \boldsymbol{Z}_i, \quad \boldsymbol{S}_{i,p} := \boldsymbol{T}_{i,p}^{-1}. \tag{19}$$

This trick, i.e., making the expectation of $\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}$ 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* $\boldsymbol{B}_i \in \mathbb{C}^{L\times K_i}$ *with* $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$ *and each* $\boldsymbol{A}_i$ *is a Gaussian random matrix, i.e., each entry in* $\boldsymbol{A}_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$*. Let* $\mu^2_{\max}$ *and* μ_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*

$$L \geq C_\alpha r^2 \max\{\mu^2_{\max} K, \mu_h^2 N\} \log^2 L \log\gamma \log(r+1),$$

where $\gamma \leq \sqrt{N\log(NL/2) + \alpha \log L}$*, then the solution of* (8) *satisfies*

$$\hat{\boldsymbol{X}}_i = \boldsymbol{X}_i, \quad \textit{for all } i = 1,\dots,r,$$

with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$.

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 4 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 $\mathcal{A}_i$ need to satisfy two key properties. The first property, proved in Section 5, can be considered as a modification of the celebrated Restricted Isometry Property (RIP) [9], as it requires the $\mathcal{A}_i$ to act in a certain sense as "local" approximate isometries [11, 10]. The second property, proved in Section 6, requires the two operators $\mathcal{A}_i$ and $\mathcal{A}_j$ 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 4. We use the golfing scheme [19] for this purpose, the constructing of which can be found in Section 7. With all these tools in place, we assemble the proof of Theorem 3.1 in Section 8.

The theorem assumes for convenience that the $\boldsymbol{h}_i$ and the $\boldsymbol{x}_i$ are real-valued, but it is easy to modify the proof for complex-valued $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$. 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 10), 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 10. 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 $\max_{j\neq k} \|\mathcal{P}_{T_j}\mathcal{A}_j^*\mathcal{A}_k\mathcal{P}_{T_k}\| \leq \frac{1}{4r}$, see conditon (26). In order to achieve this we need that L, the number of measurements, scales essentially like $r^2 \max\{\mu^2_{\max}K, \mu_h^2 N\}$ (up to log-factors), see Section 6. Is it possible, perhaps with a different condition that does not rely on mutual incoherence between the $\mathcal{A}_j$, to reduce this requirement on L to one that scales like $r\max\{\mu^2_{\max}K, \mu^2_h N\}$?

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 $\boldsymbol{y}$ is actually the sum of Hadamard products of $\boldsymbol{B}_i\boldsymbol{h}_i$ and $\boldsymbol{A}_i\boldsymbol{x}_i$. Let $\boldsymbol{F}$ be the Discrete Fourier Transform matrix of size $L\times L$ with $\boldsymbol{F}^*\boldsymbol{F} = \boldsymbol{I}_L$ and let the $L\times K_i$ matrix $\boldsymbol{B}_i$ consist of the first K_i columns of $\boldsymbol{F}$ (then $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$). Now we can express (4) equivalently as the sum of circular convolutions of $\boldsymbol{F}^{-1}(\boldsymbol{B}_i\boldsymbol{h}_i)$ and $\boldsymbol{F}^{-1}(\boldsymbol{A}_i\boldsymbol{x}_i)$, i.e.,

$$\boldsymbol{F}^{-1}\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{F}^{-1}(\boldsymbol{B}_i\boldsymbol{h}_i) * \sqrt{L}\boldsymbol{F}^{-1}(\boldsymbol{A}_i\boldsymbol{x}_i) = \sum_{i=1}^{r}(\boldsymbol{F}^{-1}\boldsymbol{B}_i)\boldsymbol{h}_i * (\sqrt{L}\boldsymbol{F}^{-1}\boldsymbol{A}_i)\boldsymbol{x}_i. \tag{20}$$

Set

$$\boldsymbol{g}_i := \begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{0}_{L-K_i} \end{bmatrix}.$$

Then there holds

$$\boldsymbol{F}^{-1}\boldsymbol{B}_i\boldsymbol{h}_i = \boldsymbol{F}^{-1}\begin{bmatrix}\boldsymbol{B}_i & \boldsymbol{0}_{L,L-K_i}\end{bmatrix}\begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{0}_{L-K_i} \end{bmatrix} = \boldsymbol{g}_i.$$

Hence with a slight abuse of notation (replacing $\boldsymbol{F}^{-1}\boldsymbol{y}$ in (20) by $\boldsymbol{y}$ and $\sqrt{L}\boldsymbol{F}^{-1}\boldsymbol{A}_i$ by $\boldsymbol{A}_i$, using the fact that the Fourier transform of a Gaussian random matrix is again a Gaussian random matrix), we can express (4) equivalently as

$$\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{g}_i * (\boldsymbol{A}_i\boldsymbol{x}_i),$$

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 $\mu_{\max} = 1$ for the particular choice of $\boldsymbol{B}_i$ above). This corollary is the precise version of the informal Theorem 1.1.

Corollary 3.2. *Consider the model in* (4)*, i.e.,*

$$\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{g}_i * (\boldsymbol{A}_i\boldsymbol{x}_i),$$

where we assume that $\boldsymbol{g}_i(k) = 0$ for $k \geq K_i$. Suppose that each $\boldsymbol{A}_i$ is a Gaussian random matrix, i.e., each entry in $\boldsymbol{A}_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. Let μ_h^2 be as defined in (13)*. If*

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

where $\gamma \leq \sqrt{N\log(NL/2) + \alpha \log L}$ then solving (8) *recovers $\boldsymbol{X}_i := \boldsymbol{h}_i\boldsymbol{x}_i^T$ exactly with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$.*

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

3.2 Noisy data

In reality measurements are noisy. Hence, suppose $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is noise with $\|\boldsymbol{\epsilon}\| \leq \eta$. In this case we solve the following optimization program to recover $\{\boldsymbol{X}_i\}_{i=1}^r$,

$$\min \sum_{i=1}^{r} \|\boldsymbol{Z}_i\|_* \quad \text{subject to} \quad \|\sum_{i=1}^{r}\mathcal{A}_i(\boldsymbol{Z}_i) - \hat{\boldsymbol{y}}\| \leq \eta. \tag{21}$$

We should choose η properly in order to make $\boldsymbol{X}_i$ inside the feasible set and $\|\hat{\boldsymbol{y}}\| > \eta$. Let $\{\hat{\boldsymbol{X}}_i\}_{i=1}^r$ be the minimizer to (21). We immediately know

$$\sum_{i=1}^{r} \|\hat{\boldsymbol{X}}_i\|_* \leq \sum_{i=1}^{r} \|\boldsymbol{X}_i\|_*. \tag{22}$$

Our goal is to see how $\sqrt{\sum_{i=1}^r \|\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\|_F^2}$ varies with respect to the noise level η.

Theorem 3.3. *Assume we observe* $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\epsilon} = \sum_{i=1}^r \mathcal{A}_i(\boldsymbol{X}_i) + \boldsymbol{\epsilon}$ *with* $\|\boldsymbol{\epsilon}\| \leq \eta$. *Then, under the same conditions as in Theorem 3.1, the minimizer* $\{\hat{\boldsymbol{X}}_i\}_{i=1}^r$ *to* (21) *satisfies*

$$\sqrt{\sum_{i=1}^r \|\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\|_F^2} \leq C\frac{\lambda_{\max}}{\lambda_{\min}} r\sqrt{\max\{K,N\}}\eta.$$

with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$. *Here,* $\lambda_{\max}^2$ *and* $\lambda_{\min}^2$ *are the largest and the smallest eigenvalue of* $\sum_{i=1}^r \mathcal{A}_i\mathcal{A}_i^*$, *respectively.*

Note that with a little modification of Lemma 2 in [2], it can be shown that $\frac{\lambda_{\max}}{\lambda_{\min}} \sim \frac{\mu_{\max}}{\mu_{\min}}$. The proof of Theorem 3.3 will be given in Section 9.

With Theorem 3.3 and Wedin's $\sin(\theta)$ theorem [42, 34] for singular value/vector perturbation theory, we immediately have the performance guarantees of recovering individual $(\boldsymbol{h}_i, \boldsymbol{x}_i)_{i=1}^r$ by applying SVD to $\hat{\boldsymbol{X}}_i$.

Corollary 3.4. *Let* $\hat{\boldsymbol{h}}_i = \sqrt{\hat{\sigma}_{i1}}\hat{\boldsymbol{u}}_{i1}$ *and* $\hat{\boldsymbol{x}}_i = \sqrt{\hat{\sigma}_{i1}}\hat{\boldsymbol{v}}_{i1}$ *where* σ_{i1}, $\hat{\boldsymbol{u}}_{i1}$ *and* $\hat{\boldsymbol{v}}_{i1}$ *are the leading singular value, left and right singular vectors of* $\hat{\boldsymbol{X}}_i$ *respectively. Then there exist* $\{c_i\}_{i=1}^r$ *and a constant* c_0 *such that*

$$\|\boldsymbol{h}_i - c_i\hat{\boldsymbol{h}}_i\| \leq c_0 \min(\epsilon/\|\boldsymbol{h}_i\|, \|\boldsymbol{h}_i\|), \quad \|\boldsymbol{x}_i - c_i^{-1}\hat{\boldsymbol{x}}_i\| \leq c_0 \min(\epsilon/\|\boldsymbol{x}_i\|, \|\boldsymbol{x}_i\|)$$

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

4 Sufficient conditions

Without loss of generality, we assume that the lifted matrix $\boldsymbol{X}_i = \alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T$, where $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ are all real and of unit norm and α_i is a scalar for all $1 \leq i \leq r$ throughout Section 4–9. We also define a linear space which $\boldsymbol{h}_i\boldsymbol{x}_i^T$ lies in and which will be useful in the further analysis:

$$T_i = \{\boldsymbol{h}_i\boldsymbol{h}_i^T\boldsymbol{Z}_i + (\boldsymbol{I} - \boldsymbol{h}_i\boldsymbol{h}_i^T)\boldsymbol{Z}_i\boldsymbol{x}_i\boldsymbol{x}_i^T | \boldsymbol{Z}_i \in \mathbb{C}^{K_i\times N_i}\} \tag{23}$$

and similarly

$$T_i^\perp = \{(\boldsymbol{I} - \boldsymbol{h}_i\boldsymbol{h}_i^T)\boldsymbol{Z}_i(\boldsymbol{I} - \boldsymbol{x}_i\boldsymbol{x}_i^T) | \boldsymbol{Z}_i \in \mathbb{C}^{K_i\times N_i}\}. \tag{24}$$

Lemma 4.1. *Assume that*

$$\sum_{i=1}^r \langle \boldsymbol{H}_i, \boldsymbol{h}_i\boldsymbol{x}_i^T\rangle + \|\boldsymbol{H}_{i,T_i^\perp}\|_* > 0.$$

for any real $\{\boldsymbol{H}_i\}_{i=1}^r$ *satisfying* $\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_i) = 0$ *and at least of* $\boldsymbol{H}_i$ *is nonzero. Then* $\{\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T\}_{i=1}^r$ *is the unique minimizer to the convex program* (8).

Proof: For any feasbile element of the convex program (8), it must have the form of $\{\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T + \boldsymbol{H}_i\}_{i=1}^r$. It suffices to show that the $\sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T + \boldsymbol{H}_i\|_* > \sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T\|_*$ for any nontrivial set of $\{\boldsymbol{H}_i\}_{i=1}^r$, i.e., at least one of $\boldsymbol{H}_i$ is nonzero. For each $\boldsymbol{H}_i$, there exists a $\boldsymbol{V}_i \in T_i^\perp$ such that

$$\langle \boldsymbol{H}_i, \boldsymbol{V}_i\rangle = \langle \boldsymbol{H}_{i,T_i^\perp}, \boldsymbol{V}_i\rangle = \|\boldsymbol{H}_{i,T_i^\perp}\|_*.$$

where $\boldsymbol{H}_{i,T_i^\perp}$ is the projection of $\boldsymbol{H}_i$ on $T_i^\perp$ and $\|\boldsymbol{V}_i\| = 1$. Thus $\boldsymbol{h}_i\boldsymbol{x}_i^T + \boldsymbol{V}_i$ belongs to the subdifferential of $\|\cdot\|_*$ at $\boldsymbol{X}_i = \alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T$.

$$\begin{aligned}
\sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T + \boldsymbol{H}_i\|_* &\geq \sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T\|_* + \langle \boldsymbol{h}_i\boldsymbol{x}_i^T + \boldsymbol{V}_i, \boldsymbol{H}_i\rangle \\
&= \sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T\|_* + \langle \boldsymbol{H}_{i,T_i}, \boldsymbol{h}_i\boldsymbol{x}_i^T\rangle + \|\boldsymbol{H}_{i,T_i^\perp}\|_* \\
&> \sum_{i=1}^r \|\alpha_i\boldsymbol{h}_i\boldsymbol{x}_i^T\|_*.
\end{aligned}$$

where the first inequality follows from the definition of subgradient and the last one is given by the assumption. ■

Now we consider under what condition on $\mathcal{A}_i$, the unique minimizer is $\{\alpha_i \boldsymbol{h}_i \boldsymbol{x}_i^T\}$. Define μ by

$$\mu := \max_{j \neq k} \|\mathcal{P}_{T_j} \mathcal{A}_j^* \mathcal{A}_k \mathcal{P}_{T_k}\| \tag{25}$$

as a measure of incoherence between any pairs of linear operators. $\mathcal{A}_{i,T_i} = \mathcal{A}_i \mathcal{P}_{T_i}$ is the restriction of $\mathcal{A}_i$ onto T_i.

Lemma 4.2. *Assume that*

$$\|\mathcal{P}_{T_i} \mathcal{A}_i^* \mathcal{A}_i \mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \leq \frac{1}{4}, \quad \mu \leq \frac{1}{4r}, \quad \|\mathcal{A}_i\| \leq \gamma \tag{26}$$

and also there exists a $\boldsymbol{\lambda} \in \mathbb{C}^L$ *such that*

$$\|\boldsymbol{h}_i \boldsymbol{x}_i^T - (\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i}\|_F \leq \alpha, \quad \|(\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i^\perp}\| \leq \beta \tag{27}$$

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

Proof: It suffices to show that for any nonzero $\{\boldsymbol{H}_i\}_{i=1}^r$ with $\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_i) = 0$,

$$\sum_{i=1}^r \langle \boldsymbol{H}_i, \boldsymbol{h}_i \boldsymbol{x}_i^T - \mathcal{A}_i^* \boldsymbol{\lambda} \rangle + \|\boldsymbol{H}_{i,T_i^\perp}\|_* > 0.$$

By decomposing the equation on T_i and $T_i^\perp$ for each i, we have

$$\sum_{i=1}^r \langle \boldsymbol{H}_{i,T_i}, \boldsymbol{h}_i \boldsymbol{x}_i^T - (\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i} \rangle - \langle \boldsymbol{H}_{i,T_i^\perp}, (\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i^\perp} \rangle + \|\boldsymbol{H}_{i,T_i^\perp}\|_* > 0.$$

Then, by applying Cauchy-Schwarz inequality and the fact that $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$, we only need to show that the following expression holds:

$$\sum_{i=1}^r -\|\boldsymbol{H}_{i,T_i}\|_F \|\boldsymbol{h}_i \boldsymbol{x}_i^T - (\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i}\|_F + \|\boldsymbol{H}_{i,T_i^\perp}\|_* (1 - \|(\mathcal{A}_i^* \boldsymbol{\lambda})_{T_i^\perp}\|) > 0. \tag{28}$$

In the following part, we will show that

$$\frac{1}{2}(\sum_{i=1}^r \|\boldsymbol{H}_{i,T_i}\|_F) \leq \gamma(\sum_{i=1}^r \|\boldsymbol{H}_{i,T_i^\perp}\|_F) \leq \gamma(\sum_{i=1}^r \|\boldsymbol{H}_{i,T_i^\perp}\|_*)$$

in order to achieve (28). We start with $\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_i) = 0$. By decomposing $\boldsymbol{H}_i$ on T_i and $T_i^\perp$ and using linearity, we have

$$\|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_{i,T_i})\|_F = \|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_{i,T_i^\perp})\|_F.$$

It is easy to bound the quantity on the right hand side by using $\|\mathcal{A}_i\| \leq \gamma$ and the triangle inequality,

$$\|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_{i,T_i^\perp})\|_F \leq \gamma(\sum_{i=1}^r \|\boldsymbol{H}_{i,T_i^\perp}\|_F). \tag{29}$$

The difficulty is to establish the lower bound. There holds

$$
\begin{aligned}
\|\sum_{i=1}^{r}\mathcal{A}_i(\boldsymbol{H}_{i,T_i})\|_F^2 &\geq \sum_{i=1}^{r}\|\mathcal{A}_i(\boldsymbol{H}_{i,T_i})\|^2 + 2\sum_{j\neq k}\langle\mathcal{A}_j(\boldsymbol{H}_{j,T_j}),\mathcal{A}_k(\boldsymbol{H}_{k,T_k})\rangle \\
&\geq \frac{3}{4}\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i}\|_F^2 - 2\mu\sum_{j\neq k}\|\boldsymbol{H}_{j,T_j}\|_F\|\boldsymbol{H}_{k,T_k}\|_F \\
&= \begin{bmatrix}\|\boldsymbol{H}_{1,T_1}\|_F \\ \vdots \\ \|\boldsymbol{H}_{r,T_r}\|_F\end{bmatrix}^T \begin{bmatrix}\frac{3}{4} & -\mu & \cdots & -\mu \\ -\mu & \frac{3}{4} & \cdots & -\mu \\ \vdots & \vdots & \ddots & \vdots \\ -\mu & -\mu & \cdots & \frac{3}{4}\end{bmatrix}\begin{bmatrix}\|\boldsymbol{H}_{1,T_1}\|_F \\ \vdots \\ \|\boldsymbol{H}_{r,T_r}\|_F\end{bmatrix},
\end{aligned}
$$

where the second inequality uses $\|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_i\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \leq \frac{1}{4}$ and $\|\mathcal{P}_{T_2}\mathcal{A}_2^*\mathcal{A}_1\mathcal{P}_{T_1}\| \leq \mu \leq \frac{1}{4r}$. It is easy to see that the coefficient matrix inside the quadratic form has its smallest eigenvalue $\frac{3}{4} - (r-1)\mu \geq \frac{3}{4} - \frac{r-1}{4r} > \frac{1}{2}$ and all the other eigenvalues are $\frac{3}{4}$. Now we have

$$
\|\sum_{i=1}^{r}\mathcal{A}_i(\boldsymbol{H}_{i,T_i})\|_F \geq \sqrt{\frac{1}{2}\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i}\|_F^2} \geq \frac{1}{2r}\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i}\|_F. \tag{30}
$$

Combining (30) and (29) leads to

$$
\frac{1}{2r}\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i}\|_F \leq \gamma\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i^\perp}\|_F. \tag{31}
$$

The expression on the left side of (28) has its lower bound as follows:

$$
\begin{aligned}
&\sum_{i=1}^{r} -\|\boldsymbol{H}_{i,T_i}\|_F\|\boldsymbol{h}_i\boldsymbol{x}_i^T - (\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i}\|_F + \|\boldsymbol{H}_{i,T_i^\perp}\|_*(1 - \|(\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i^\perp}\|) \\
&\geq \sum_{i=1}^{r} -\|\boldsymbol{H}_{i,T_i}\|_F\|\boldsymbol{h}_i\boldsymbol{x}_i^T - (\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i}\|_F + \|\boldsymbol{H}_{i,T_i^\perp}\|_F(1 - \|(\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i^\perp}\|) \\
&\geq -\alpha\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i}\|_F + (1-\beta)\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i^\perp}\|_F \\
&\geq -2r\gamma\alpha\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i^\perp}\|_F + (1-\beta)\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i^\perp}\|_F \\
&\geq (-2r\gamma\alpha + (1-\beta))\sum_{i=1}^{r}\|\boldsymbol{H}_{i,T_i^\perp}\|_F \geq 0,
\end{aligned}
$$

where the first inequality uses $\|\cdot\|_* \geq \|\cdot\|_F$, the second one follows from the assumption (27), and the third one follows from (31). Under the condition $-2r\gamma\alpha + (1-\beta) > 0$, (28) holds if at least one of the terms $\|\boldsymbol{H}_{i,T_i^\perp}\|_F$ is nonzero. If $\boldsymbol{H}_{i,T_i^\perp} = 0$ for all $1 \leq i \leq r$, then $\boldsymbol{H}_i = 0$ via (31). ■

5 Local Isometry Property

Our goal in this section is to prove that the first assumption in (26) of Lemma 4.2 holds with high probability if L is large enough. Instead of studying $\|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_i\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\|$ directly, we will focus on the more general expression $\|\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\|$, where $\mathcal{A}_{i,p}$ and $\boldsymbol{S}_{i,p}$ are defined in (17) and (19) respectively.

5.1 An explicit formula for $\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i}$

For each fixed pair of (i,p) where $1 \le i \le r$ and $1 \le p \le P$, the proof of $\|\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}$ is actually the same. Therefore, for simplicity of notation, we omit the subscript i and denote $\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i}$ by $\mathcal{P}_T\mathcal{A}_p^*\mathcal{A}_p\boldsymbol{S}_p\mathcal{P}_T$ throughout the proof of Proposition 5.1. By definition, $\mathcal{A}_p\boldsymbol{S}_p\mathcal{P}_T(\boldsymbol{Z}) = \{\boldsymbol{b}_l^*\boldsymbol{S}_p\mathcal{P}_T(\boldsymbol{Z})\boldsymbol{a}_l\}_{l\in\Gamma_p}$ for any $\boldsymbol{Z} \in \mathbb{C}^{K\times N}$. Using (23) gives us an explicit expression of $\boldsymbol{b}_l^*\boldsymbol{S}_p\mathcal{P}_T(\boldsymbol{Z})\boldsymbol{a}_l$, i.e.,

$$\begin{aligned}
\boldsymbol{b}_l^*\boldsymbol{S}_p\mathcal{P}_T(\boldsymbol{Z})\boldsymbol{a}_l &= \boldsymbol{b}_l^*\boldsymbol{S}_p\left[\boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z} + (\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*\right]\boldsymbol{a}_l \\
&= \langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{h}^*\boldsymbol{Z}\boldsymbol{a}_l + \langle \boldsymbol{a}_l, \boldsymbol{x}\rangle \boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x} \\
&= \boldsymbol{h}^*\boldsymbol{Z}\widetilde{\boldsymbol{v}}_l + \widetilde{\boldsymbol{u}}_l^*\boldsymbol{Z}\boldsymbol{x}, \quad l \in \Gamma_p,
\end{aligned}$$

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

$$\boldsymbol{b}_l^*\mathcal{P}_T(\boldsymbol{Z})\boldsymbol{a}_l = \boldsymbol{h}^*\boldsymbol{Z}\boldsymbol{v}_l + \boldsymbol{u}_l^*\boldsymbol{Z}\boldsymbol{x}, \quad l \in \Gamma_p$$

where

$$\begin{aligned}
\boldsymbol{v}_l &:= \langle \boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{a}_l, && (32)\\
\boldsymbol{u}_l &:= \langle \boldsymbol{a}_l, \boldsymbol{x}\rangle (\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l, && (33)\\
\widetilde{\boldsymbol{v}}_l &:= \langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{a}_l, && (34)\\
\widetilde{\boldsymbol{u}}_l &:= \langle \boldsymbol{a}_l, \boldsymbol{x}\rangle (\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l. && (35)
\end{aligned}$$

Now we have

$$\mathcal{A}\mathcal{P}_T(\boldsymbol{S}_p\boldsymbol{Z}) = \{\langle \boldsymbol{Z}, \boldsymbol{h}\widetilde{\boldsymbol{v}}_l^* + \widetilde{\boldsymbol{u}}_l\boldsymbol{x}^*\rangle\}_{l\in\Gamma_p} \quad \mathcal{P}_T\mathcal{A}^*(\boldsymbol{z}) = \sum_{l\in\Gamma_p} z_l(\boldsymbol{h}\boldsymbol{v}_l^* + \boldsymbol{u}_l\boldsymbol{x}^*).$$

By combining the terms we arrive at

$$\mathcal{P}_T\mathcal{A}_p^*\mathcal{A}_p\boldsymbol{S}_p\mathcal{P}_T(\boldsymbol{Z}) = \sum_{l\in\Gamma_p}\left[\boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}\widetilde{\boldsymbol{v}}_l\boldsymbol{v}_l^* + \boldsymbol{h}\widetilde{\boldsymbol{u}}_l^*\boldsymbol{Z}\boldsymbol{x}\boldsymbol{v}_l^* + \boldsymbol{u}_l\boldsymbol{h}^*\boldsymbol{Z}\widetilde{\boldsymbol{v}}_l\boldsymbol{x}^* + \boldsymbol{u}_l\widetilde{\boldsymbol{u}}_l^*\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*\right]. \quad (36)$$

The explicit form of each component in this summation is

$$\begin{aligned}
\boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}\widetilde{\boldsymbol{v}}_l\boldsymbol{v}_l^* &= \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}\boldsymbol{a}_l\boldsymbol{a}_l^*, \\
\boldsymbol{h}\widetilde{\boldsymbol{u}}_l^*\boldsymbol{Z}\boldsymbol{x}\boldsymbol{v}_l^* &= \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*\boldsymbol{a}_l\boldsymbol{a}_l^*, \\
\boldsymbol{u}_l\boldsymbol{h}^*\boldsymbol{Z}\widetilde{\boldsymbol{v}}_l\boldsymbol{x}^* &= \langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle (\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{h}^*\boldsymbol{Z}\boldsymbol{a}_l\boldsymbol{a}_l^*\boldsymbol{x}\boldsymbol{x}^*, \\
\boldsymbol{u}_l\widetilde{\boldsymbol{u}}_l^*\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^* &= |\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*.
\end{aligned}$$

It is easy to compute the expectation of those random matrices by using $\mathbb{E}(\boldsymbol{a}_l\boldsymbol{a}_l^*) = \boldsymbol{I}_N$ and $\mathbb{E}\,|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 = \|\boldsymbol{x}\|^2 = 1$. Our goal here is to estimate the operator norm of $\mathcal{P}_T\mathcal{A}_p^*\mathcal{A}_p\boldsymbol{S}_p\mathcal{P}_T - \mathcal{P}_T$ which is the sum of four components, i.e.,

$$\mathcal{P}_T\mathcal{A}_p^*\mathcal{A}_p\boldsymbol{S}_p\mathcal{P}_T - \mathcal{P}_T = \sum_{s=1}^{4}\mathcal{M}_s$$

where each $\mathcal{M}_i$ is a random linear operator with zero mean. More precisely, each of $\mathcal{M}_s$ is given by

$$\begin{aligned}
\mathcal{M}_1(\boldsymbol{Z}) &= \sum_{l\in\Gamma_p} \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I}), && (37)\\
\mathcal{M}_2(\boldsymbol{Z}) &= \sum_{l\in\Gamma_p} \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I}), && (38)\\
\mathcal{M}_3(\boldsymbol{Z}) &= \sum_{l\in\Gamma_p} \langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\boldsymbol{x}^*, && (39)\\
\mathcal{M}_4(\boldsymbol{Z}) &= \sum_{l\in\Gamma_p} (|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*. && (40)
\end{aligned}$$

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

5.2 Main result in this section

Now we present the main result in this section.

Proposition 5.1. *Under the assumption of* (14) *and* (10) *and that* $\{\boldsymbol{a}_{i,l}\}$ *are standard Gaussian random vectors of length* N_i,

$$\|\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}, \quad 1 \le i \le r, 1 \le p \le P \tag{41}$$

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

The following corollary, which is a special case of Proposition 5.1 (simply set $Q = L$ and $\boldsymbol{S}_{i,p} = \boldsymbol{I}_{K_i}$), indicates the first condition in (26) holds with high probability.

Corollary 5.2. *Under the assumption of* (14) *and* (10) *and that* $\{\boldsymbol{a}_{i,l}\}$ *are standard Gaussian random vectors of length* N_i,

$$\|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_i\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}, \quad 1 \le i \le r \tag{42}$$

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

Remark 5.3. *Although Proposition 5.1 and Corollary 5.2 are quite similar to Lemma 3 in [2] at the first glance, we include* $\boldsymbol{S}_{i,p}$ *and the new definition of* μ_h^2 *in our result. The purpose is to resolve the issue mentioned in Remark 2.1 by making* $\mathbb{E}(\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i}) = \mathcal{P}_{T_i}$. *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 5.1 is given as follows.

Proof: To prove Proposition 5.1, it suffices to show that $\|\mathcal{M}_s\| \le \frac{1}{16}$ for $1 \le s \le 4$ and then take the union bound over all $1 \le p \le P$ and $1 \le i \le r$. For each fixed pair of (i,p), it is shown in Lemmata 5.5–5.8 that

$$\|\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}$$

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

$$\mathbb{P}\left(\|\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}, \quad \forall 1 \le i \le r, 1 \le p \le P\right) \ge 1 - 4PrL^{-\alpha} \ge 1 - 4rL^{-\alpha+1}$$

if $Q \geq C_\alpha \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L$ 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, $Q \geq C_\alpha \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L \log(r+1)$.

■

5.3 Main tools

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

Theorem 5.4. *Consider a finite sequence of $\mathcal{Z}_l$ of independent centered random matrices with dimension $M_1 \times M_2$. Assume that $\|\mathcal{Z}_l\|_{\psi_1} \leq R$ where the norm $\|\cdot\|_{\psi_1}$ of a matrix is defined as*

$$\|\boldsymbol{Z}\|_{\psi_1} := \inf_{u\geq 0}\{\mathbb{E}[\exp(\|\boldsymbol{Z}\|/u)] \leq 2\}. \tag{43}$$

and introduce the random matrix

$$\boldsymbol{S} = \sum_{l=1}^{Q} \mathcal{Z}_l. \tag{44}$$

Compute the variance parameter

$$\sigma^2 = \max\Big\{\|\sum_{l=1}^{Q}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\|, \|\sum_{l=1}^{Q}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\|\Big\}, \tag{45}$$

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

$$\|\boldsymbol{S}\| \leq C_0 \max\{\sigma\sqrt{t + \log(M_1 + M_2)}, R\log\left(\frac{\sqrt{Q}R}{\sigma}\right)(t + \log(M_1+M_2))\} \tag{46}$$

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

5.4 Estimation of four summations

5.4.1 Estimation of $\mathcal{M}_1$ in (37)

Lemma 5.5. *Under the assumption of* (14), (10) *and* (12) *and that $\boldsymbol{a}_l \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ independently, then*

$$\|\mathcal{M}_1\| \leq \frac{1}{16}$$

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

Proof: By definition of $\mathcal{M}_1$ in (37),

$$\mathcal{M}_1(\boldsymbol{Z}) := \sum_{l\in\Gamma_p}\mathcal{Z}_l(\boldsymbol{Z}), \quad \mathcal{Z}_l(\boldsymbol{Z}) := \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle \boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})$$

Each $\mathcal{Z}_l$ is a rank-1 matrix and can be viewed as a $KN \times KN$ matrix since it applies to $\mathbb{C}^{K\times N}$. Moreover, $\|\mathcal{Z}_l\| = |\overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle|\|\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I}\|$ is a random variable with an exponential tail. In order to apply Theorem 5.4, we need to know R and the upper bound of σ^2. Following from (32), (14) and Lemma 12.1,

$$\|\mathcal{Z}_l\|_{\psi_1} \leq |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle|\|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\|_{\psi_1} \leq C\frac{\mu_h}{\sqrt{L}}\cdot\frac{\sqrt{L}\mu_h}{Q}\cdot N = C\frac{\mu_h^2 N}{Q}$$

Thus $R := \max_{l\in\Gamma_p} \|\mathcal{Z}_l\|_{\psi_1} \leq C\frac{\mu_h^2 N}{Q}$. Note that $\mathcal{Z}^*(\boldsymbol{Z}) = \langle \boldsymbol{h}, \boldsymbol{b}_l\rangle \overline{\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle} \boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})$. We can express $\mathcal{Z}^*\mathcal{Z} = \mathcal{Z}\mathcal{Z}^*$ as

$$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) = |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})^2.$$

Then we continue to compute its variance,

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &= (N+1)\sum_{l\in\Gamma_p}|\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2|\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \\
&\leq \frac{2\mu_h^2 N}{L}\sum_{l\in\Gamma_p}|\langle \boldsymbol{S}_p\boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \\
&\leq \frac{2\mu_h^2 N}{L}\cdot\|\boldsymbol{S}_p\| \\
&\leq \frac{2\mu_h^2 N}{L}\cdot\frac{4L}{3Q} = \frac{8\mu_h^2 N}{3Q}.
\end{aligned}$$

where $\mathbb{E}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})^2 = (N+1)\boldsymbol{I}$ follows from (94). Thus the variance σ^2 is bounded by

$$\sigma^2 \leq \frac{8\mu_h^2 N}{3Q}.$$

$\log\left(\frac{\sqrt{Q}R}{\sigma}\right) \leq C_1 \log L$ for some positive constant C_1 since $\sqrt{Q}R/\sigma$ is at most of poly-L order. Applying (46) immediately by choosing $t = \alpha\log L$ and $Q \geq C_\alpha \mu_h^2 N\log^2 L/\delta^2$ gives us

$$\mathcal{M}_1 \leq C\max\{\sqrt{\frac{\mu_h^2 N}{Q}(\alpha\log L + \log(2KN))}, \frac{\mu_h^2 N}{Q}(\alpha\log L + \log(2KN))\log L\} \leq \delta,$$

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

$$\|\mathcal{M}_1\| \leq \frac{1}{16}$$

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

■

5.4.2 Estimation of $\mathcal{M}_2$ in (38)

Lemma 5.6. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_l \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ *independently, then*

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

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

Proof: By definition of $\mathcal{M}_2$ in (38),

$$\mathcal{M}_2(\boldsymbol{Z}) := \sum_{l\in\Gamma_p}\mathcal{Z}_l(\boldsymbol{Z}), \quad \mathcal{Z}_l(\boldsymbol{Z}) = \overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle}\boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})$$

Immediately, we have $\|\mathcal{Z}_l\| = \|\overline{\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle} \boldsymbol{h}\boldsymbol{b}_l^* \boldsymbol{S}_p\| \cdot \|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\|$ and $\mathcal{Z}_l$ is actually a $KN \times KN$ matrix. Then we estimate $\|\mathcal{Z}_l\|_{\psi_1}$ as follows:

$$\begin{aligned}
\|\mathcal{Z}_l\|_{\psi_1} &= |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle| \|\boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p\| \cdot \|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\|_{\psi_1} \\
&= \frac{\mu_h}{\sqrt{L}} \cdot \|\boldsymbol{S}_p\boldsymbol{b}_l\| \cdot \|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\|_{\psi_1} \\
&\leq \frac{\mu_h}{\sqrt{L}} \cdot \frac{4L}{3Q} \frac{\mu_{\max}\sqrt{K}}{\sqrt{L}} \cdot \|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\|_{\psi_1} \\
&\leq C\frac{\mu_{\max}\mu_h\sqrt{KN}}{Q} \\
&\leq C\frac{\max\{\mu_{\max}^2 K, \mu_h^2 N\}}{Q},
\end{aligned}$$

where the first equality uses the fact that $\|\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*\| = 1$, $\|\boldsymbol{x}\boldsymbol{x}^*\| = 1$ and $|\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle| \leq \frac{\mu_h}{L}$. The third inequality uses $\|\boldsymbol{S}_p\boldsymbol{b}_l\| \leq \|\boldsymbol{S}_p\| \cdot \|\boldsymbol{b}_l\|$ and the fourth inequality follows from $\|(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\|_{\psi_1} \leq C\sqrt{N}$ in (98). Therefore we have $R := \max_{l\in\Gamma_p} \|\mathcal{Z}_l\|_{\psi_1} \leq C\frac{\max\{\mu_{\max}^2 K, \mu_h^2 N\}}{Q}$. Now we proceed to estimate σ^2. By definition, the adjoint of $\mathcal{Z}_l^*$ can be represented as

$$\mathcal{Z}_l^*(\boldsymbol{Z}) = \langle \boldsymbol{h}, \boldsymbol{b}_l\rangle (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\boldsymbol{x}^*.$$

Then $\mathcal{Z}^*\mathcal{Z}$ and $\mathcal{Z}\mathcal{Z}^*$ are easily obtained

$$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) = |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})^2\boldsymbol{x}\boldsymbol{x}^*$$

and

$$\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z}) = |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I}).$$

The expectation of $\mathcal{Z}_l^*\mathcal{Z}_l$ and $\mathcal{Z}_l\mathcal{Z}_l^*$ are computed via

$$\begin{aligned}
\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})) &= \mathbb{E}\, |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})^2\boldsymbol{x}\boldsymbol{x}^* \\
&= (N+1)|\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*
\end{aligned}$$

where $\mathbb{E}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})^2 = (N+1)\boldsymbol{I}$ follows from (94). Similarly,

$$\begin{aligned}
\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})) &= \mathbb{E}(|\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \boldsymbol{h}\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})) \\
&= |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 \boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{h}\boldsymbol{h}^*\boldsymbol{Z}(\boldsymbol{I} + \boldsymbol{x}\boldsymbol{x}^*)
\end{aligned}$$

where $\mathbb{E}[(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\boldsymbol{x}\boldsymbol{x}^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})] = \|\boldsymbol{x}\|^2\boldsymbol{I} + \boldsymbol{x}\boldsymbol{x}^*$ from (99) and the fact that $\boldsymbol{x}$ is real. Taking the sum of $\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)$ and $\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)$ over $l \in \Gamma_p$ leads to

$$\begin{aligned}
\|\sum_{l\in\Gamma_p} \mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &= (N+1)\|\sum_{l\in\Gamma_p} |\langle \boldsymbol{h}, \boldsymbol{b}_l\rangle|^2 (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\| \\
&\leq \frac{2\mu_h^2 N}{L}\|(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\| \\
&\leq \frac{2\mu_h^2 N}{L} \cdot \frac{4L}{3Q} = \frac{8\mu_h^2 N}{3Q}
\end{aligned}$$

and

$$
\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\| &= \|\sum_{l\in\Gamma_p}|\langle \boldsymbol{h},\boldsymbol{b}_l\rangle|^2\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\|\cdot\|\boldsymbol{I}+\boldsymbol{x}\boldsymbol{x}^*\| \\
&\leq 2\max_{l\in\Gamma_p}\{\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I}-\boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\}\sum_{l\in\Gamma_p}|\langle \boldsymbol{h},\boldsymbol{b}_l\rangle|^2 \\
&\leq 2\left[\|\boldsymbol{S}_p\|^2\max_{l\in\Gamma_p}\|\boldsymbol{b}_l\|^2\right]\cdot\|\boldsymbol{T}_p\| \\
&\leq \frac{32L^2}{9Q^2}\cdot\frac{\mu_{\max}^2K}{L}\cdot\frac{5Q}{4L} \\
&\leq \frac{40\mu_{\max}^2K}{9Q}.
\end{aligned}
$$

Thus the variance σ^2 is bounded above by

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

and $\log\left(\frac{\sqrt{Q}R}{\sigma}\right)\leq C_1\log L$ for some constant C_1. Then we just use (46) to estimate the deviation of $\mathcal{M}_2$ from $\mathbf{0}$ by choosing $t=\alpha\log L$ and $Q\geq C_\alpha\max\{\mu_{\max}^2K,\mu_h^2N\}\log^2 L/\delta^2$ and it gives us

$$
\begin{aligned}
\mathcal{M}_2 \leq\ & C\max\left\{\sqrt{\frac{\max\{\mu_{\max}^2K,\mu_h^2N\}}{Q}(\alpha\log L+\log(2KN))}\right. \\
&\left.,\frac{\max\{\mu_{\max}^2K,\mu_h^2N\}}{Q}(\alpha\log L+\log(2KN))\log L\right\}\leq\delta.
\end{aligned}
$$

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 $1-L^{-\alpha}$.

■

5.4.3 Estimation of $\mathcal{M}_3$ in (39)

The proof of Lemma 5.7 is quite similar to that of Lemma 5.6. For simplicity, we just state the result without proving it in details.

Lemma 5.7. *Under the assumption of* (14)*,* (10) *and* (12) *and that* $\boldsymbol{a}_l\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_N)$ *independently, then*

$$
\|\mathcal{M}_3\|\leq\frac{1}{16}
$$

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

5.4.4 Estimation of $\mathcal{M}_4$ in (40)

Lemma 5.8. *Under the assumption of* (14)*,* (10) *and* (12) *and that* $\boldsymbol{a}_l\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_N)$ *independently, then*

$$
\|\mathcal{M}_4\|\leq\frac{1}{16}
$$

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

Proof: From the definition of $\mathcal{M}_4$ in (40),

$$\mathcal{M}_4 := \sum_{l\in\Gamma_p} \mathcal{Z}_l, \quad \mathcal{Z}_l(\boldsymbol{Z}) = (|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*$$

Note that $\mathcal{Z}_l$ can be regarded as a $KN \times KN$ matrix and $\|\mathcal{Z}_l\| = |(|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)| \cdot \|\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p\|$. $\|\mathcal{Z}_l\|_{\psi_1}$ is estimated as

$$\begin{aligned}
\|\mathcal{Z}_l\|_{\psi_1} &\leq \|(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\| \cdot \|(|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)\|_{\psi_1} \\
&\leq C\frac{\mu_{\max}^2 K}{L} \cdot \frac{4L}{3Q} = C\frac{\mu_{\max}^2 K}{Q},
\end{aligned}$$

where $\|(|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)\|_{\psi_1} \leq C\|\boldsymbol{x}\|^2 = C$ follows from (97). The second step is to estimate σ^2. Note $\mathcal{Z}_l^*(\boldsymbol{Z}) = (|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*$ and $\mathcal{Z}_l^*\mathcal{Z}_l$ and $\mathcal{Z}_l\mathcal{Z}_l^*$ are in the following forms:

$$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) = (|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)^2 \cdot \boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*$$

and

$$\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z}) = (|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)^2 \cdot \boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*.$$

Consider its expectation,

$$\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})) = 2\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*$$

and

$$\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})) = 2\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^*$$

where $\mathbb{E}(|\langle \boldsymbol{a}_l, \boldsymbol{x}\rangle|^2 - 1)^2 = 2$. By taking the sum over $l \in \Gamma_p$, we have an estimation of σ^2 :

$$\begin{aligned}
\|\sum_{l\in\Gamma_p} \mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &= 2\|\sum_{l\in\Gamma_p} [\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)]\| \\
&\leq 2\max_{l\in\Gamma_p}\{\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\} \cdot \|\sum_{l\in\Gamma_p} [(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)]\| \\
&\leq 2\frac{\mu_{\max}^2 K}{L} \cdot \|\boldsymbol{S}_p\| = \frac{\mu_{\max}^2 K}{L} \cdot \frac{8L}{3Q} \\
&\leq C\frac{\mu_{\max}^2 K}{Q},
\end{aligned}$$

and

$$\begin{aligned}
\|\sum_{l\in\Gamma_p} \mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\| &= 2\|\sum_{l\in\Gamma_p} [\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l \cdot (\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{b}_l\boldsymbol{b}_l^*(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)]\| \\
&\leq 2\max_{l\in\Gamma_p}\{\boldsymbol{b}_l^*\boldsymbol{S}_p(\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^*)\boldsymbol{S}_p\boldsymbol{b}_l\} \cdot \|\boldsymbol{T}_p\| \\
&\leq 2\left(\frac{4L}{3Q}\right)^2 \cdot \frac{\mu_{\max}^2 K}{L} \cdot \frac{5Q}{4L} \\
&= \frac{40\mu_{\max}^2 K}{9Q}.
\end{aligned}$$

Therefore we have

$$\sigma^2 \leq C\frac{\mu_{\max}^2 K}{Q}$$

and $\log\left(\frac{\sqrt{Q}R}{\sigma}\right) \leq C_1 \log L$ for some constant C_1. Now we are ready to apply Bernstein inequality (46) by choosing $t = \alpha \log L$ and $Q \geq C_\alpha \mu_{\max}^2 K \log^2 L/\delta^2$ gives us

$$\begin{aligned}\mathcal{M}_4 \quad \leq \quad & C \max\left\{\sqrt{\frac{\mu_{\max}^2 K}{Q}(\alpha \log L + \log(2KN))}\right. \\ & \left., \frac{\mu_{\max}^2 K}{Q}(\alpha \log L + \log(2KN)) \log L\right\} \leq \delta.\end{aligned}$$

with probability at least $1 - L^{-\alpha}$ where K and N are properly assumed to be smaller than L. In particular, $\|\mathcal{M}_4\| \leq \frac{1}{16}$ if one can choose $\delta = \frac{1}{16}$.

■

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

In this section, we aim to show that $\mu \leq \frac{1}{4r}$, where μ is defined in (25), i.e., the second condition in (26) holds with high probability. The main idea here is first to show that a more general and stronger version of incoherent property,

$$\|\mathcal{P}_{T_j} \mathcal{A}_{j,p}^* \mathcal{A}_{k,p} \boldsymbol{S}_{k,p} \mathcal{P}_{T_k}\| \leq \frac{1}{4r}$$

holds with high probability for any $1 \leq p \leq P$ and $j \neq k$. Since the derivation is exactly 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).

6.1 An explicit formulation of $\mathcal{P}_{T_2} \mathcal{A}_{2,p}^* \mathcal{A}_{1,p} \boldsymbol{S}_{1,p} \mathcal{P}_{T_1}$

Following the same procedures as the previous sections, we have explicit expressions for $\mathcal{A}_{1,p}\mathcal{P}_{T_1}$ and $\mathcal{P}_{T_2}\mathcal{A}_{2,p}^*$,

$$\mathcal{A}_{1,p} \boldsymbol{S}_{1,p} \mathcal{P}_{T_1}(\boldsymbol{Z}) = \{\langle \boldsymbol{Z}, \boldsymbol{h}_1 \widetilde{\boldsymbol{v}}_{1,l}^* + \widetilde{\boldsymbol{u}}_{1,l} \boldsymbol{x}_1^* \rangle\}_{l \in \Gamma_p} \quad \mathcal{P}_{T_2}\mathcal{A}_2^*(\boldsymbol{z}) = \sum_{l \in \Gamma_p} z_l (\boldsymbol{h}_2 \boldsymbol{v}_{2,l}^* + \boldsymbol{u}_{2,l} \boldsymbol{x}_2^*).$$

where $\widetilde{\boldsymbol{u}}_{1,l}$, $\widetilde{\boldsymbol{v}}_{1,l}$, $\boldsymbol{u}_{2,l}$ and $\boldsymbol{v}_{2,l}$ are defined in (32) except the notation, i.e., we omit subscript i in the previous section. By combining $\mathcal{P}_{T_2}\mathcal{A}_{2,p}^*$ and $\mathcal{A}_{1,p}\boldsymbol{S}_{1,p}\mathcal{P}_{T_1}$, we arrive at

$$\mathcal{P}_{T_2}\mathcal{A}_{2,p}^*\mathcal{A}_{1,p}\boldsymbol{S}_{1,p}\mathcal{P}_{T_1}(\boldsymbol{Z}) = \sum_{l \in \Gamma_p} \left[\boldsymbol{h}_2 \boldsymbol{h}_1^* \boldsymbol{Z} \widetilde{\boldsymbol{v}}_{1,l} \boldsymbol{v}_{2,l}^* + \boldsymbol{h}_2 \widetilde{\boldsymbol{u}}_{1,l}^* \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{v}_{2,l}^* + \boldsymbol{u}_{2,l} \boldsymbol{h}_1^* \boldsymbol{Z} \widetilde{\boldsymbol{v}}_{1,l} \boldsymbol{x}_2^* + \boldsymbol{u}_{2,l} \widetilde{\boldsymbol{u}}_{1,l}^* \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{x}_2^*\right]. \tag{47}$$

Note that the expectations of all terms are equal to $\boldsymbol{0}$ because $\{\boldsymbol{u}_{1,l}, \boldsymbol{v}_{1,l}\}$ is independent of $\{\boldsymbol{u}_{2,l}, \boldsymbol{v}_{2,l}\}$ and both $\boldsymbol{u}_{i,l}$ and $\boldsymbol{v}_{i,l}$ are of zero mean. Define $\mathcal{M}_{s,\text{mix}}$ as

$$\mathcal{M}_{1,\text{mix}}(\boldsymbol{Z}) := \sum_{l \in \Gamma_p} \boldsymbol{h}_2 \boldsymbol{h}_1^* \boldsymbol{Z} \widetilde{\boldsymbol{v}}_{1,l} \boldsymbol{v}_{2,l}^* = \sum_{l \in \Gamma_p} \langle \boldsymbol{S}_{1,p} \boldsymbol{h}_1, \boldsymbol{b}_{1,l} \rangle \overline{\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l} \rangle} \boldsymbol{h}_2 \boldsymbol{h}_1^* \boldsymbol{Z} \boldsymbol{a}_{1,l} \boldsymbol{a}_{2,l}^*, \tag{48}$$

$$\mathcal{M}_{2,\text{mix}}(\boldsymbol{Z}) := \sum_{l \in \Gamma_p} \boldsymbol{h}_2 \widetilde{\boldsymbol{u}}_{1,l}^* \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{v}_{2,l}^* = \sum_{l \in \Gamma_p} \langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \overline{\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l} \rangle} \boldsymbol{h}_2 \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p} (\boldsymbol{I} - \boldsymbol{h}_1 \boldsymbol{h}_1^*) \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{a}_{2,l}^*, \tag{49}$$

$$\mathcal{M}_{3,\text{mix}}(\boldsymbol{Z}) := \sum_{l \in \Gamma_p} \boldsymbol{u}_{2,l} \boldsymbol{h}_1^* \boldsymbol{Z} \widetilde{\boldsymbol{v}}_{1,l} \boldsymbol{x}_2^* = \sum_{l \in \Gamma_p} \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle \langle \boldsymbol{S}_{1,p} \boldsymbol{h}_1, \boldsymbol{b}_{1,l} \rangle (\boldsymbol{I} - \boldsymbol{h}_2 \boldsymbol{h}_2^*) \boldsymbol{b}_{2,l} \boldsymbol{h}_1^* \boldsymbol{Z} \boldsymbol{a}_{1,l} \boldsymbol{x}_2^*, \tag{50}$$

$$\mathcal{M}_{4,\text{mix}}(\boldsymbol{Z}) := \sum_{l \in \Gamma_p} \boldsymbol{u}_{2,l} \widetilde{\boldsymbol{u}}_{1,l}^* \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{x}_2^* = \sum_{l \in \Gamma_p} \langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle (\boldsymbol{I} - \boldsymbol{h}_2 \boldsymbol{h}_2^*) \boldsymbol{b}_{2,l} \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p} (\boldsymbol{I} - \boldsymbol{h}_1 \boldsymbol{h}_1^*) \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{x}_2^* \tag{51}$$

and the sum of $\mathcal{M}_{s,\text{mix}}$ simply follows:

$$\mathcal{P}_{T_2}\mathcal{A}_{2,p}^*\mathcal{A}_{1,p}\boldsymbol{S}_{1,p}\mathcal{P}_{T_1} = \sum_{s=1}^{4}\mathcal{M}_{s,\text{mix}}.$$

Each $\mathcal{M}_{s,\text{mix}}$ can be treated as a $K_2N_2 \times K_1N_1$ matrix because it is a linear operator from $\mathbb{C}^{K_1\times N_1}$ to $\mathbb{C}^{K_2\times N_2}$.

6.2 Main result in this section

Here is the main result in this section.

Proposition 6.1. *Under the assumption of* (14) *and* (10) *and that* $\{\boldsymbol{a}_{i,l}\}$ *are standard Gaussian random vectors of length* N_i,

$$\|\mathcal{P}_{T_j}\mathcal{A}_{j,p}^*\mathcal{A}_{k,p}\boldsymbol{S}_{k,p}\mathcal{P}_{T_k}\| \leq \frac{1}{4r}, \quad 1\leq j\neq k\leq r, 1\leq p\leq P \tag{52}$$

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

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

Corollary 6.2. *Under the assumption of* (14) *and* (10) *and that* $\{\boldsymbol{a}_{i,l}\}$ *are standard Gaussian random vectors of length* N_i,

$$\|\mathcal{P}_{T_j}\mathcal{A}_{j}^*\mathcal{A}_{k}\mathcal{P}_{T_k}\| \leq \frac{1}{4r}, \quad 1\leq j\neq k\leq r, 1\leq p\leq P \tag{53}$$

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

The proof of Proposition 6.1 follows two steps. First we will show each $\|\mathcal{M}_{s,\text{mix}}\|\leq\frac{1}{16r}$ 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 6.3–6.6, that

$$\|\mathcal{P}_{T_j}\mathcal{A}_{j,p}^*\mathcal{A}_{k,p}\boldsymbol{S}_{k,p}\mathcal{P}_{T_k}\| \leq \frac{1}{4r}$$

with probability at least $1-4L^{-\alpha}$ if $Q\geq C_\alpha r^2\max\{\mu_{\max}^2K_i,\mu_h^2N_i\}\log^2 L$. 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

$$\mathbb{P}\left(\|\mathcal{P}_{T_j}\mathcal{A}_{j,p}^*\mathcal{A}_{k,p}\boldsymbol{S}_{k,p}\mathcal{P}_{T_k}\| \leq \frac{1}{4r}, \quad \forall j\neq k, 1\leq p\leq P\right)\geq 1-4L^{-\alpha}Pr^2/2\geq 1-2L^{-\alpha+1}r^2$$

if $Q\geq C_\alpha r^2\max\{\mu_{\max}^2K,\mu_h^2N\}\log^2 L$ 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, $Q\geq C_\alpha r^2\max\{\mu_{\max}^2K,\mu_h^2N\}\log^2 L\log(r+1)$.

■

6.3 Estimation of $\|\mathcal{M}_{s,\text{mix}}\|$

The idea of the proof is simple but the actual proof itself involves quite a few calculations, i.e., computing the $\|\cdot\|_{\psi_1}$ and the variance σ^2 and then applying Bernstein inequality.

6.3.1 Proof of $\|\mathcal{M}_{1,\text{mix}}\| \le \frac{1}{16r}$

Lemma 6.3. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_{i,l} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N_i})$ *independently for* $i = 1, 2$ *and* $l \in \Gamma_p$*, then*

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

holds with probability $1 - L^{-\alpha}$ *if* $Q \ge C_\alpha r^2 \mu_h^2 N \log^2 L$.

Proof: By definition of $\mathcal{M}_{1,\text{mix}}$ in (48),

$$\mathcal{M}_{1,\text{mix}}(\boldsymbol{Z}) := \sum_{l\in\Gamma_p} \mathcal{Z}_l(\boldsymbol{Z}), \quad \mathcal{Z}_l(\boldsymbol{Z}) := \langle \boldsymbol{S}_{1,p}\boldsymbol{h}_1, \boldsymbol{b}_{1,l}\rangle \overline{\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle} \boldsymbol{h}_2 \boldsymbol{h}_1^* \boldsymbol{Z} \boldsymbol{a}_{1,l} \boldsymbol{a}_{2,l}^*$$

and $\|\mathcal{Z}_l\| = |\langle \boldsymbol{S}_{1,p}\boldsymbol{h}_1, \boldsymbol{b}_{1,l}\rangle \overline{\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle}| \|\boldsymbol{a}_{1,l}\boldsymbol{a}_{2,l}^*\|$. Following from (32) and (14) gives

$$\begin{aligned}
\|\mathcal{Z}_l\|_{\psi_1} &\le |\langle \boldsymbol{S}_{1,p}\boldsymbol{h}_1, \boldsymbol{b}_{1,l}\rangle \langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle| \|(\|\boldsymbol{a}_{1,l}\| \cdot \|\boldsymbol{a}_{2,l}\|)\|_{\psi_1} \\
&\le \frac{\sqrt{L}\mu_h}{Q} \cdot \frac{\mu_h}{\sqrt{L}} \|(\|\boldsymbol{a}_{1,l}\| \cdot \|\boldsymbol{a}_{2,l}\|)\|_{\psi_1} \\
&\le C\frac{\mu_h^2 \sqrt{N_1 N_2}}{Q} \le C\frac{\mu_h^2 N}{Q}.
\end{aligned}$$

where the last inequality follows from Lemma 12.4 and the fact that $\boldsymbol{a}_{i,l}$ is a $N_i \times 1$ Gaussian random vector and therefore $\|\boldsymbol{a}_{i,l}\|$ is the square root of a χ^2 random variable of freedom N_i. Now let us move to the estimation of σ^2. By the definition of the adjoint operator,

$$\mathcal{Z}^*(\boldsymbol{Z}) = \overline{\langle \boldsymbol{S}_{1,p}\boldsymbol{h}_1, \boldsymbol{b}_{1,l}\rangle} \langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle \boldsymbol{h}_1 \boldsymbol{h}_2^* \boldsymbol{Z} \boldsymbol{a}_{2,l} \boldsymbol{a}_{1,l}^*.$$

We can express $\mathcal{Z}^*\mathcal{Z}$ as

$$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) = |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \|\boldsymbol{a}_{2,l}\|^2 \boldsymbol{h}_1 \boldsymbol{h}_1^* \boldsymbol{Z} \boldsymbol{a}_{1,l} \boldsymbol{a}_{1,l}^*,$$

and $\mathcal{Z}\mathcal{Z}^*$ as

$$\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z}) = |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \|\boldsymbol{a}_{1,l}\|^2 \boldsymbol{h}_1 \boldsymbol{h}_1^* \boldsymbol{Z} \boldsymbol{a}_{2,l} \boldsymbol{a}_{2,l}^*.$$

Their expectations are

$$\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})) = N_2 |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \boldsymbol{h}_1 \boldsymbol{h}_1^* \boldsymbol{Z}$$

and

$$\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})) = N_1 |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \boldsymbol{h}_2 \boldsymbol{h}_2^* \boldsymbol{Z}.$$

We proceed to computing their variance.

$$\begin{aligned}
\|\sum_{l\in\Gamma_p} \mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &= N_2 \|\sum_{l\in\Gamma_p} |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \boldsymbol{h}_1 \boldsymbol{h}_1^*\| \\
&= N_2 \sum_{l\in\Gamma_p} |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \\
&\le \frac{N_2\mu_h^2}{L} \sum_{l\in\Gamma_p} |\langle \boldsymbol{h}_1, \boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\rangle|^2 \\
&\le \frac{N_2\mu_h^2}{L} \cdot \|\boldsymbol{S}_{1,p}\| \le \frac{4\mu_h^2 N_2}{3Q}.
\end{aligned}$$

where the third and fourth inequalities follow from (14) and (10). Similarly, one can have

$$\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\| \leq \frac{4\mu_h^2 N_1}{3Q}.$$

Thus the variance σ^2 is bounded by

$$\sigma^2 := \max\{\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\|, \|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\|\} \leq \frac{4\mu_h^2\max\{N_1,N_2\}}{3Q} \leq C\frac{\mu_h^2 N}{Q}.$$

Applying (46) by choosing $t=\alpha\log L$ and $Q\geq C_\alpha\mu_h^2 N\log^2 L/\delta^2$ immediately gives us

$$\begin{aligned}\mathcal{M}_{1,\text{mix}} &\leq C\max\left\{\sqrt{\frac{\mu_h^2 N}{Q}(\alpha\log L+\log(2KN))}\right.\\ &\left., \frac{\mu_h^2 N}{Q}(\alpha\log L+\log(2KN))\log L\right\} \leq \delta,\end{aligned}$$

with probability at least $1-L^{-\alpha}$ where K and N are properly assumed to be smaller than L. By choosing $\delta=\frac{1}{16r}$ and let $Q\geq C_\alpha r^2\mu_h^2 N\log^2 L$,

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

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

■

6.3.2 Proof of $\|\mathcal{M}_{2,\text{mix}}\|\leq\frac{1}{16r}$

Lemma 6.4. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_{i,l}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{N_i})$ *independently for* $i=1,2$ *and* $l\in\Gamma_p$*, then*

$$\|\mathcal{M}_{2,\text{mix}}\| \leq \frac{1}{16r}$$

holds with probability $1-L^{-\alpha}$ *if* $Q\geq C_\alpha r^2\max\{\mu_{\max}^2 K,\mu_h^2 N\}\log^2 L$.

Proof: Following from the definition in (49),

$$\mathcal{M}_{2,\text{mix}} := \sum_{l\in\Gamma_p}\mathcal{Z}_l(\boldsymbol{Z}), \quad \mathcal{Z}_l(\boldsymbol{Z}) = \langle\boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle\overline{\langle\boldsymbol{h}_2,\boldsymbol{b}_{2,l}\rangle}\boldsymbol{h}_2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{Z}\boldsymbol{x}_1\boldsymbol{a}_{2,l}^*$$

and $\|\mathcal{Z}_l\| = |\langle\boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle\overline{\langle\boldsymbol{h}_2,\boldsymbol{b}_{2,l}\rangle}|\|\boldsymbol{h}_2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}\|\|\boldsymbol{x}_1\boldsymbol{a}_{2,l}^*\|$. By using Lemma 12.1 and 12.4,

$$\begin{aligned}\|\mathcal{Z}_l\|_{\psi_1} &\leq |\langle\boldsymbol{h}_2,\boldsymbol{b}_{2,l}\rangle|\|\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\|\cdot\|(|\langle\boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle|\cdot\|\boldsymbol{a}_{2,l}\|)\|_{\psi_1}\\ &\leq C\frac{\mu_h}{\sqrt{L}}\frac{\sqrt{K_1}\mu_{\max}}{\sqrt{L}}\frac{4L}{3Q}\|(|\langle\boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle|\cdot\|\boldsymbol{a}_{2,l}\|)\|_{\psi_1}\\ &\leq C\frac{4\mu_{\max}\mu_h\sqrt{K_1}}{3Q}\|(|\langle\boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle|\cdot\|\boldsymbol{a}_{2,l}\|)\|_{\psi_1}\\ &\leq C\frac{\mu_{\max}\mu_h\sqrt{K_1N_2}}{Q}\\ &\leq C\frac{\max\{\mu_{\max}^2K_1,\mu_h^2N_2\}}{Q}\end{aligned}$$

where $\|\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\| \leq \|\boldsymbol{S}_{1,p}\|\|\boldsymbol{b}_{1,l}\|$ and Lemma 12.4 gives

$$\|(|\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1\rangle| \cdot \|\boldsymbol{a}_{2,l}\|)\|_{\psi_1} \leq C\sqrt{N_2}.$$

since $\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1\rangle$ is a standard Gaussian random variable and $\|\boldsymbol{a}_{2,l}\|^2$ is a χ^2 random variable of degree N_2. We proceed to estimate σ^2 by first finding $\mathcal{Z}_l^*(\boldsymbol{Z})$,

$$\mathcal{Z}_l^*(\boldsymbol{Z}) = \overline{\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1\rangle}\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{h}_2^*\boldsymbol{Z}\boldsymbol{a}_{2,l}\boldsymbol{x}_1^*.$$

$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})$ and $\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})$ have the following forms:

$$\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) = |\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1\rangle\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2\|\boldsymbol{a}_{2,l}\|^2(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{Z}\boldsymbol{x}_1\boldsymbol{x}_1^*$$

and

$$\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z}) = |\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1\rangle\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{h}_2\boldsymbol{h}_2^*\boldsymbol{Z}\boldsymbol{a}_{2,l}\boldsymbol{a}_{2,l}^*.$$

The expectations of $\mathcal{Z}_l^*\mathcal{Z}_l$ and $\mathcal{Z}_l\mathcal{Z}_l^*$ are

$$\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})) = N_2|\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{Z}\boldsymbol{x}_1\boldsymbol{x}_1^*.$$

and

$$\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})) = |\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l} \cdot \boldsymbol{h}_2\boldsymbol{h}_2^*\boldsymbol{Z}.$$

where $\mathbb{E}(\boldsymbol{a}_{i,l}\boldsymbol{a}_{i,l}^*) = \boldsymbol{I}_{N_i}$ and $\mathbb{E}\,\|\boldsymbol{a}_{i,l}\|^2 = N_i$. Taking the sum over $l \in \Gamma_p$ leads to

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &= N_2\|\sum_{l\in\Gamma_p}|\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\| \\
&\leq \frac{\mu_h^2N_2}{L}\|\sum_{l\in\Gamma_p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\| \\
&\leq \frac{\mu_h^2N_2}{L}\cdot\|\boldsymbol{S}_{1,p}\| \leq \frac{\mu_h^2N_2}{L}\cdot\frac{4L}{3Q} = \frac{4\mu_h^2N_2}{3Q}.
\end{aligned}$$

and

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\| &= \sum_{l\in\Gamma_p}|\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\,\cdot \\
&= \sum_{l\in\Gamma_p}\|(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\|^2|\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \\
&\leq \max_{l\in\Gamma_p}\|(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\|^2\sum_{l\in\Gamma_p}|\langle \boldsymbol{h}_2, \boldsymbol{b}_{2,l}\rangle|^2 \\
&\leq \max_{l\in\Gamma_p}\{\|\boldsymbol{S}_{1,p}\|^2\|\boldsymbol{b}_{1,l}\|^2\}\cdot\|\boldsymbol{T}_{2,p}\| \\
&\leq \frac{16L^2}{9Q^2}\cdot\frac{\mu_{\max}^2K_1}{L}\cdot\frac{5Q}{4L} \\
&= \frac{20\mu_{\max}^2K_1}{9Q}.
\end{aligned}$$

Thus the variance σ^2 is bounded by

$$\sigma^2 \leq C\frac{\max\{\mu_{\max}^2K_1, \mu_h^2N_2\}}{Q} \leq C\frac{\max\{\mu_{\max}^2K, \mu_h^2N\}}{Q}.$$

Then we just apply (46) to estimate the deviation of $\mathcal{M}_{2,\text{mix}}$ from $\mathbf{0}$ by choosing $t = \alpha \log L$ and $Q \geq C_\alpha \max\{\mu^2_{\max} K, \mu^2_h N\} \log^2 L/\delta^2$ immediately gives us

$$\begin{aligned}\mathcal{M}_{2,\text{mix}} \quad &\leq \quad C \max\left\{ \sqrt{\frac{\max\{\mu^2_{\max} K, \mu^2_h N\}}{Q}(\alpha \log L + \log(2KN))} \right. \\ &\left. , \frac{\max\{\mu^2_{\max} K, \mu^2_h N\}}{Q}(\alpha \log L + \log(2KN)) \log L \right\} \leq \delta.\end{aligned}$$

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 $Q \geq C_\alpha r^2 \max\{\mu^2_{\max} K, \mu^2_h N\} \log^2 L$,

$$\|\mathcal{M}_{2,\text{mix}}\| \leq \frac{1}{16r}$$

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

■

6.3.3 Proof of $\|\mathcal{M}_{3,\text{mix}}\| \leq \frac{1}{16r}$

The estimation of $\mathcal{M}_{3,\text{mix}}$ is actually the same as $\mathcal{M}_{2,\text{mix}}$ by slightly changing the subscript of $\mathcal{M}_{3,\text{mix}}$. Therefore, we only give the statement of lemma without proofs.

Lemma 6.5. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_{i,l} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{N_i})$ *independently for* $i = 1, 2$ *and* $l \in \Gamma_p$, *then*

$$\|\mathcal{M}_{3,\text{mix}}\| \leq \frac{1}{16r}$$

holds with probability $1 - L^{-\alpha}$ *if* $Q \geq C_\alpha r^2 \max\{\mu^2_{\max} K, \mu^2_h N\} \log^2 L$.

6.3.4 Proof of $\|\mathcal{M}_{4,\text{mix}}\| \leq \frac{1}{16r}$

Lemma 6.6. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_{i,l} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{N_i})$ *independently for* $i = 1, 2$ *and* $l \in \Gamma_p$, *then*

$$\|\mathcal{M}_{4,\text{mix}}\| \leq \frac{1}{16r}$$

holds with probability $1 - L^{-\alpha}$ *if* $Q \geq C_\alpha r^2 \mu^2_{\max} K \log^2 L$.

Proof: By definition of $\mathcal{M}_{4,\text{mix}}$ in (51),

$$\mathcal{M}_{4,\text{mix}}(\boldsymbol{Z}) := \sum_{l \in \Gamma_p} \mathcal{Z}_l(\boldsymbol{Z}), \quad \mathcal{Z}_l(\boldsymbol{Z}) = \langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle (\boldsymbol{I} - \boldsymbol{h}_2 \boldsymbol{h}_2^*) \boldsymbol{b}_{2,l} \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p} (\boldsymbol{I} - \boldsymbol{h}_1 \boldsymbol{h}_1^*) \boldsymbol{Z} \boldsymbol{x}_1 \boldsymbol{x}_2^*$$

and $\|\mathcal{Z}_l\| = |\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle| \|\boldsymbol{b}_{2,l} \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p}\|$. As usual, we first give an upper bound of $\|\mathcal{Z}_l\|_{\psi_1}$,

$$\begin{aligned}\|\mathcal{Z}_l\|_{\psi_1} \quad &= \quad \|\boldsymbol{b}_{2,l} \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p}\| \cdot \|\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle\|_{\psi_1} \\ &\leq \quad \frac{\mu^2_{\max} \sqrt{K_1 K_2}}{L} \cdot \frac{4L}{3Q} \|\langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle\|_{\psi_1} \\ &\leq \quad C \frac{\mu^2_{\max} \sqrt{K_1 K_2}}{Q} \leq C \frac{\mu^2_{\max} K}{Q}\end{aligned}$$

where $\prod_{i=1}^2 |\langle \boldsymbol{a}_{i,l}, \boldsymbol{x}_i \rangle|$ is the product of two standard Gaussian random variables and its ψ_1-norm is bounded by a constant. Thus $R \leq C \frac{\mu^2_{\max} K}{Q}$. The next step is to estimate σ^2.

$$\mathcal{Z}_l^*(\boldsymbol{Z}) = \langle \boldsymbol{a}_{1,l}, \boldsymbol{x}_1 \rangle \langle \boldsymbol{a}_{2,l}, \boldsymbol{x}_2 \rangle (\boldsymbol{I} - \boldsymbol{h}_1 \boldsymbol{h}_1^*) \boldsymbol{S}_{1,p} \boldsymbol{b}_{1,l} \boldsymbol{b}_{2,l}^* (\boldsymbol{I} - \boldsymbol{h}_2 \boldsymbol{h}_2^*) \boldsymbol{Z} \boldsymbol{x}_2 \boldsymbol{x}_1^*$$

It is easy to verify that

$$\begin{aligned}\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z}) &= |\langle \boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle\langle \boldsymbol{a}_{2,l},\boldsymbol{x}_2\rangle|^2\|(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{b}_{2,l}\|^2(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{Z}\boldsymbol{x}_1\boldsymbol{x}_1^*\\ \mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z}) &= |\langle \boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle\langle \boldsymbol{a}_{2,l},\boldsymbol{x}_2\rangle|^2\|(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\|^2(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{b}_{2,l}\boldsymbol{b}_{2,l}^*(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{Z}\boldsymbol{x}_2\boldsymbol{x}_2^*.\end{aligned}$$

Taking the expectation and using the fact that $\mathbb{E}\,|\langle \boldsymbol{a}_{1,l},\boldsymbol{x}_1\rangle\langle \boldsymbol{a}_{2,l},\boldsymbol{x}_2\rangle|^2 = 1$ lead to

$$\begin{aligned}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l(\boldsymbol{Z})) &= \|(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{b}_{2,l}\|^2(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{Z}\boldsymbol{x}_1\boldsymbol{x}_1^*\\ \mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*(\boldsymbol{Z})) &= \|(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\|^2(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{b}_{2,l}\boldsymbol{b}_{2,l}^*(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{Z}\boldsymbol{x}_2\boldsymbol{x}_2^*.\end{aligned}$$

By taking the sum over $l\in\Gamma_p$, we have an estimation of σ^2.

$$\begin{aligned}\Big\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\Big\| &= \Big\|\sum_{l\in\Gamma_p}\|(\boldsymbol{I}-\boldsymbol{h}_2\boldsymbol{h}_2^*)\boldsymbol{b}_{2,l}\|^2(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\Big\|\\ &\le \max_{l\in\Gamma_p}\|\boldsymbol{b}_{2,l}\|^2\cdot\Big\|\sum_{l\in\Gamma_p}\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}\Big\|\\ &\le \frac{\mu_{\max}^2K_2}{L}\cdot\|\boldsymbol{S}_{1,p}\|\le\frac{\mu_{\max}^2K_2}{L}\cdot\frac{4L}{3Q}=\frac{4\mu_{\max}^2K_2}{3Q}\end{aligned}$$

and

$$\Big\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\Big\|\le\|\boldsymbol{S}_{1,p}\|^2\max_{l\in\Gamma_p}\|\boldsymbol{b}_{1,l}\|^2\cdot\|\boldsymbol{T}_{2,p}\|\le\frac{20\mu_{\max}^2K_1}{9Q}.$$

Therefore

$$\sigma^2\le C\frac{Q\mu_{\max}^2\max\{K_1,K_2\}}{L^2}\le C\frac{\mu_{\max}^2K}{Q}.$$

Now we are ready to apply Bernstein inequality: by choosing $t=\alpha\log L$ and $Q\ge C_\alpha\mu_{\max}^2K\log^2 L/\delta^2$ immediately gives us

$$\begin{aligned}\mathcal{M}_{4,\mathrm{mix}} &\le C\max\Big\{\sqrt{\frac{\mu_{\max}^2K}{Q}(\alpha\log L+\log(2KN))}\\ &\quad,\frac{\mu_{\max}^2K}{Q}(\alpha\log L+\log(2KN))\log L\Big\}\le\delta.\end{aligned}$$

with probability at least $1-L^{-\alpha}$ where K and N are properly assumed to be smaller than L. If we let $\delta=\frac{1}{16r}$ and $Q\ge C_\alpha r^2\mu_{\max}^2K\log^2 L$, then

$$\|\mathcal{M}_{4,\mathrm{mix}}\|\le\frac{1}{16r}$$

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

■

7 Constructing a dual certificate

In this section, we will finish the proof of the main theorem by constructing a $\boldsymbol{\lambda}$ such that

$$\|\boldsymbol{h}_i\boldsymbol{x}_i^*-(\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i}\|_F\le(5r\gamma)^{-1},\quad\|(\mathcal{A}_i^*\boldsymbol{\lambda})_{T_i^\perp}\|\le\frac{1}{2}\tag{54}$$

holds simultaneously for all $1\le i\le r$. If such a $\boldsymbol{\lambda}$ exists, then solving (8) yields exact recovery according to Lemma 4.2. The difficulty of this mission is obvious since we require all $\mathcal{A}_i^*\boldsymbol{\lambda}$ to be close to $\boldsymbol{h}_i\boldsymbol{x}_i^*$ and "small" on $T_i^\perp$. However, it becomes possible with help of the incoherence between $\mathcal{A}_i$ and $\mathcal{A}_j$. The method to achieve that is to apply a well-known and widely used technique called golfing scheme, developed by Gross in [19].

7.1 Golfing scheme

The approximate dual certificate $\{\boldsymbol{Y}_i := \mathcal{A}_i^*\boldsymbol{\lambda}\}_{i=1}^r$ satisfying Lemma 4.2 is constructed via a sequence of random matrices, following from the philosophy of golfing scheme. The constructed sequence $\{\boldsymbol{Y}_{i,p}\}_{p=1}^P$ would approach $\boldsymbol{h}_i\boldsymbol{x}_i^T$ on T_i exponentially fast while keeping $\boldsymbol{Y}_{i,p}$ "small" on $T_i^\perp$ at the same time. Initialize $\boldsymbol{Y}_{i,0} = \boldsymbol{0}_{K_i\times N_i}$ for all $1 \leq i \leq r$ and

$$\boldsymbol{\lambda}_0 := \sum_{j=1}^r \mathcal{A}_{j,1}(\boldsymbol{S}_{j,1}\boldsymbol{h}_j\boldsymbol{x}_j^*) \in \mathbb{C}^L$$

Then for p from 1 to P (where P will be specified later in Lemma 7.1), we define the following recursive formula:

$$\boldsymbol{\lambda}_{p-1} := \sum_{j=1}^r \mathcal{A}_{j,p}\left(\boldsymbol{S}_{j,p}(\boldsymbol{h}_j\boldsymbol{x}_j^* - \mathcal{P}_{T_j}(\boldsymbol{Y}_{j,p-1}))\right) \tag{55}$$

$$\boldsymbol{Y}_{i,p} := \boldsymbol{Y}_{i,p-1} + \mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1}, \quad 1 \leq i \leq r. \tag{56}$$

$\boldsymbol{Y}_{i,p}$ denotes the result after p-th iteration and let $\boldsymbol{Y}_i = \boldsymbol{Y}_{i,P}$, i.e., the final outcome for each i. Denote $\boldsymbol{W}_{i,p}$ as the difference between $\boldsymbol{Y}_{i,p}$ and $\boldsymbol{h}_i\boldsymbol{x}_i^*$ on T_i, i.e.,

$$\boldsymbol{W}_{i,p} = \boldsymbol{h}_i\boldsymbol{x}_i^* - \mathcal{P}_{T_i}(\boldsymbol{Y}_{i,p}) \in T_i, \quad \boldsymbol{W}_{i,0} = \boldsymbol{h}_i\boldsymbol{x}_i^* \tag{57}$$

and (55) can be simplified into

$$\boldsymbol{\lambda}_{p-1} = \sum_{i=1}^r \mathcal{A}_{i,p}(\boldsymbol{S}_{i,p}\boldsymbol{W}_{i,p-1}).$$

Moreover, $\boldsymbol{W}_{i,p}$ yields the following equation:

$$\boldsymbol{W}_{i,p} = \boldsymbol{W}_{i,p-1} - \sum_{j=1}^r \mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{j,p}(\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1}) \tag{58}$$

from (56) and (57). An important observation here is that each $\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1}$ is an *unbiased* estimator of $\boldsymbol{W}_{i,p-1}$, i.e.,

$$\mathbb{E}(\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1}) = \sum_{j=1}^r \mathbb{E}(\mathcal{A}_{i,p}^*\mathcal{A}_{j,p}(\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1})) = \boldsymbol{W}_{i,p-1}. \tag{59}$$

where $\mathbb{E}(\mathcal{A}_{i,p}^*\mathcal{A}_{j,p}(\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1})) = \boldsymbol{0}$ for all $j \neq i$ due to the independence between $\mathcal{A}_{j,p}$ and $\mathcal{A}_{i,p}$ and $\mathbb{E}(\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{S}_{i,p}\boldsymbol{W}_{i,p-1})) = \boldsymbol{W}_{i,p-1}$. Remember that $\{\boldsymbol{W}_{j,p-1}\}_{j=1}^r$ are independent of $\{\mathcal{A}_{i,p}\}_{i=1}^r$ based on the construction of sequences in (55) and (56). Therefore, more precisely the expectation above should be treated as the conditional expectation of $\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1}$ given $\{\boldsymbol{W}_{j,p-1}\}_{j=1}^r$ are known.

7.2 $\|\mathcal{P}_{T_i}(\boldsymbol{Y}_i) - \boldsymbol{h}_i\boldsymbol{x}_i^*\|_F$ decays exponentially fast

Lemma 7.1. *Conditioned on* (41) *and* (53), *the golfing scheme* (55) *and* (56) *generate a sequence of* $\{\boldsymbol{Y}_{i,p}\}_{p=1}^P$ *such that*

$$\|\boldsymbol{W}_{i,p}\|_F = \|\mathcal{P}_{T_i}(\boldsymbol{Y}_{i,p}) - \boldsymbol{h}_i\boldsymbol{x}_i^*\|_F \leq 2^{-p}$$

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

$$\|\mathcal{P}_{T_i}(\boldsymbol{Y}_i) - \boldsymbol{h}_i\boldsymbol{x}_i^*\| \leq 2^{-\log_2(5r\gamma)} \leq \frac{1}{5r\gamma}$$

where $\boldsymbol{Y}_i := \boldsymbol{Y}_{i,P}$. *In other words, the first condition in* (54) *holds.*

Proof: Directly following from (58) leads to

$$\begin{aligned} \boldsymbol{W}_{i,p} &= \boldsymbol{W}_{i,p-1} - \mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}(\boldsymbol{S}_{i,p}\boldsymbol{W}_{i,p-1}) - \sum_{j\neq i}^{r}\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{j,p}(\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1}) \quad (60)\\ &= \boldsymbol{W}_{i,p-1} - \mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{i,p}\boldsymbol{S}_{i,p}\mathcal{P}_{T_i}(\boldsymbol{W}_{i,p-1}) - \sum_{j\neq i}^{r}\mathcal{P}_{T_i}\mathcal{A}_{i,p}^*\mathcal{A}_{j,p}\boldsymbol{S}_{j,p}\mathcal{P}_{T_j}(\boldsymbol{W}_{j,p-1}). \quad (61)\end{aligned}$$

where $\boldsymbol{W}_{j,p-1} \in T_j$ and thus $\boldsymbol{W}_{j,p-1} = \mathcal{P}_{T_j}(\boldsymbol{W}_{j,p-1})$. By using triangle inequality and applying (41) and (53),

$$\|\boldsymbol{W}_{i,p}\|_F \leq \frac{1}{4}\|\boldsymbol{W}_{i,p-1}\|_F + \frac{1}{4r}\sum_{j\neq i}\|\boldsymbol{W}_{j,p-1}\|_F, \quad 1 \leq i \leq r.$$

From the formula above, it is easy to see that

$$\max_{1\leq i\leq r}\|\boldsymbol{W}_{i,p}\|_F \leq \frac{1}{2}\max_{1\leq i\leq r}\|\boldsymbol{W}_{i,p-1}\|_F,$$

Recall that $\|\boldsymbol{W}_{i,0}\|_F = \|\boldsymbol{h}_i\boldsymbol{x}_i^*\|_F = 1$ for all $1 \leq i \leq r$ and by the induction above, we prove that

$$\|\boldsymbol{W}_{i,p}\|_F \leq 2^{-p}, \quad 1 \leq p \leq P, \quad 1 \leq i \leq r.$$

■

7.3 Proof of $\|\mathcal{P}_{T_i^\perp}(\boldsymbol{Y}_{i,P})\| \leq \frac{1}{2}$

In the previous section, we have already shown that $\mathcal{P}_{T_i}(\boldsymbol{Y}_{i,p})$ approaches $\boldsymbol{h}_i\boldsymbol{x}_i^*$ exponentially fast with respect to p. The only missing piece of the proof is to show that $\|\mathcal{P}_{T_i^\perp}(\boldsymbol{Y}_{i,P})\|$ is bounded by $\frac{1}{2}$ for all $1 \leq i \leq r$, i.e., the second condition in (27) holds. Without loss of generality, we set $i = 1$. Following directly from (55) and (56),

$$\boldsymbol{Y}_{1,P} = \sum_{p=1}^{P}\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1}.$$

Simply applying the triangle inequality leads to

$$\begin{aligned} \|\mathcal{P}_{T_1^\perp}(\boldsymbol{Y}_{1,P})\| &= \|\mathcal{P}_{T_1^\perp}\left(\sum_{p=1}^{P}\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{P-1}\right)\| \\ &= \|\mathcal{P}_{T_1^\perp}\left(\sum_{p=1}^{P}(\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1})\right)\| \\ &\leq \sum_{p=1}^{P}\|\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1}\|, \end{aligned}$$

where the second equation follows from $\mathcal{P}_{T_1^\perp}(\boldsymbol{W}_{1,p-1}) = \boldsymbol{0}$. It suffices to demonstrate that $\|\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1}\| \leq 2^{-p-1}$ in order to make $\|\boldsymbol{Y}_{1,P}\| \leq \frac{1}{2}$ since

$$\|\mathcal{P}_{T_1}(\boldsymbol{Y}_{1,P})\| \leq \sum_{p=1}^{P}2^{-p-1} < \frac{1}{2}.$$

Before moving to the proof, we first define the quantity μ_p which will be useful in the proof,

$$\mu_p := \frac{Q}{\sqrt{L}} \max_{1\leq i\leq r, l\in\Gamma_{p+1}} \|\boldsymbol{W}_{i,p}^* \boldsymbol{S}_{i,p+1} \boldsymbol{b}_{i,l}\|. \tag{62}$$

In particular, $\mu_0 \leq \mu_h$ because of

$$\mu_0 = \frac{Q}{\sqrt{L}} \max_{i,l\in\Gamma_1} \|\boldsymbol{x}_i \boldsymbol{h}_i^* \boldsymbol{S}_{i,1} \boldsymbol{b}_{i,l}\| = \frac{Q}{\sqrt{L}} \max_{i,l\in\Gamma_1} \|\boldsymbol{h}_i^* \boldsymbol{S}_{i,1} \boldsymbol{b}_{i,l}\| \leq \mu_h.$$

and the definition of μ_h in (13). Also we define $\boldsymbol{w}_{i,l}$ as

$$\boldsymbol{w}_{i,l} := \boldsymbol{W}_{i,p-1}^* \boldsymbol{S}_{i,p} \boldsymbol{b}_{i,l}, \quad l \in \Gamma_p \tag{63}$$

and we have

$$\max_{i,l\in\Gamma_p} \|\boldsymbol{w}_{i,l}\| \leq \frac{\sqrt{L}}{Q} \mu_{p-1}. \tag{64}$$

Remark 7.2. *The definition of μ_p is a little complicated but the idea behind it is simple. Since we have already shown in Lemma 7.1 that $\boldsymbol{W}_{i,p}$ is very close to $\boldsymbol{h}_i\boldsymbol{x}_i^*$ for large p, μ_p can be viewed as a measure of the incoherence between $\boldsymbol{W}_{i,p}$ (an approximation of $\boldsymbol{h}_i\boldsymbol{x}_i^*$) and $\boldsymbol{b}_{i,l}$ in the $p+1$th block (i.e., Γ_{p+1}). We would like to have "small" μ_p, i.e., $\mu_p \leq \|\boldsymbol{W}_{i,p}\|\mu_h \leq 2^{-p}\mu_h$ which would guarantee that $\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1}$ concentrates well around $\boldsymbol{W}_{i,p-1}$ for all i and p. This insight leads us to the following lemma.*

Lemma 7.3. *Let μ_p be defined in (62) and $\boldsymbol{W}_{i,p}$ satisfy*

$$\mu_p \leq 2^{-p}\mu_h, \quad \|\boldsymbol{W}_{i,p}\|_F \leq 2^{-p}, \quad 1 \leq p \leq P, 1 \leq i \leq r.$$

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

$$\|\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{i,p-1}\| \leq 2^{-p-1},$$

simultaneously for (p,i) with probability at least $1 - L^{-\alpha+1}$. Thus, the second condition in (54),

$$\|\mathcal{P}_{T_i^\perp}(\boldsymbol{Y}_{i,P})\| \leq \frac{1}{2}$$

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

Remark 7.4. *The validity of the assumption $\mu_p \leq 2^{-p}\mu_h$ is assured in Lemma 7.5.*

Proof: Without loss of generality, we start with $i = 1$. It is shown in (59) that

$$\mathbb{E}\left(\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1}\right) = 0.$$

First we rewrite $\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1}$ into the sum of rank-1 matrices with mean $\mathbf{0}$ by using (55) and (17),

$$\mathcal{A}_{1,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{1,p-1} = \sum_{l\in\Gamma_p} \left[\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p} \boldsymbol{W}_{1,p-1} \left(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}\right) + \sum_{j\neq 1} \boldsymbol{b}_{1,l}\boldsymbol{b}_{j,l}^* \boldsymbol{S}_{j,p} \boldsymbol{W}_{j,p-1} \boldsymbol{a}_{j,l}\boldsymbol{a}_{1,l}^* \right]. \tag{65}$$

Denote $\mathcal{Z}_l$ by

$$\mathcal{Z}_l := \boldsymbol{b}_{1,l}\boldsymbol{w}_{1,l}^* \left(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}\right) + \sum_{j\neq 1} \boldsymbol{b}_{1,l}\boldsymbol{w}_{j,l}^* \boldsymbol{a}_{j,l}\boldsymbol{a}_{1,l}^* \in \mathbb{C}^{K_i\times N_i} \tag{66}$$

where $\boldsymbol{w}_{j,l}$ is defined in (63). The goal is to bound the operator norm of (65), i.e, $\|\sum_{l\in\Gamma_p}\mathcal{Z}_l\|$, by 2^{-p-1}. An important fact here is that μ_{p-1} is independent of all $\boldsymbol{a}_{i,l}$ with $l \in \Gamma_p$ because

μ_{p-1} is a function of $\{\boldsymbol{a}_{i,k}\}_{k\in\Gamma_s, s<p}$. Following from (62) and the assumption $\mu_p \leq 2^{-p}\mu_h$, we have

$$\|\boldsymbol{w}_{i,l}\| \leq \frac{\sqrt{L}}{Q}\mu_{p-1} \leq \frac{\sqrt{L}}{Q}2^{-p+1}\mu_h, \quad \forall l \in \Gamma_p. \tag{67}$$

The proof is more or less a routine: estimate $\|\mathcal{Z}_l\|_{\psi_1}$, σ^2 and apply (46). For any fixed $l \in \Gamma_p$,

$$\begin{aligned}\|\mathcal{Z}_l\| &\leq \|\boldsymbol{b}_{1,l}\boldsymbol{w}_{1,l}^*\left(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}\right)\| + \sum_{j\neq 1}\|\boldsymbol{b}_{1,l}\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}\boldsymbol{a}_{1,l}^*\| \\ &\leq \frac{\mu_{\max}\sqrt{K_1}}{\sqrt{L}}\left[\|\boldsymbol{w}_{1,l}^*\left(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}\right)\| + \sum_{j\neq 1}\|\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}\boldsymbol{a}_{1,l}^*\|\right]\end{aligned}$$

Note that for $j \neq 1$, $\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l} \sim \mathcal{N}(0, \|\boldsymbol{w}_{j,l}\|^2)$ and $\|\boldsymbol{a}_{1,l}\|^2 \sim \chi^2_{N_1}$. From (67) and Lemma 12.4,

$$\|(|\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}| \cdot \|\boldsymbol{a}_{1,l}\|)\|_{\psi_1} \leq C\sqrt{N_1}\|\boldsymbol{w}_{j,l}\| \leq C\frac{2^{-p+1}\mu_h\sqrt{LN_1}}{Q}.$$

On the other hand,

$$\|\boldsymbol{w}_{1,l}^*(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\|_{\psi_1} \leq C\sqrt{N_1}\|\boldsymbol{w}_{1,l}\| \leq C\frac{2^{-p+1}\mu_h\sqrt{LN_1}}{Q}.$$

follows from (98) and (67). Taking the sum over j, from 1 to r, gives

$$\|\mathcal{Z}_l\|_{\psi_1} \leq C\frac{2^{-p+1}r\mu_{\max}\mu_h\sqrt{K_1N_1}}{Q} \leq C\frac{2^{-p+1}r\max\{\mu^2_{\max}K, \mu_h^2 N\}}{Q}, \quad l \in \Gamma_p.$$

Thus we have $R := \max_{l\in\Gamma_p}\|\mathcal{Z}_l\|_{\psi_1} \leq C\frac{2^{-p+1}r\max\{\mu^2_{\max}K, \mu_h^2 N\}}{Q}$. Now let's move on to the estimation of σ^2.

$$\mathcal{Z}_l = \boldsymbol{b}_{1,l}\boldsymbol{w}_{1,l}^*(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}) + \sum_{j\neq 1}\boldsymbol{b}_{1,l}\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}\boldsymbol{a}_{1,l}^*.$$

and

$$\mathcal{Z}_l^* = (\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\boldsymbol{w}_{1,l}\boldsymbol{b}_{1,l}^* + \sum_{j\neq 1}\boldsymbol{a}_{1,l}\boldsymbol{a}_{j,l}^*\boldsymbol{w}_{j,l}\boldsymbol{b}_{1,l}^*.$$

The corresponding $\mathcal{Z}_l^*\mathcal{Z}_l$ and $\mathcal{Z}_l\mathcal{Z}_l^*$ have quite complicated expressions. However, all the cross terms have zero expectation, which simplifies $\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)$ and $\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)$ a lot. Their expectations are

$$\begin{aligned}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l) &= \mathbb{E}\left(\|\boldsymbol{b}_{1,l}\|^2(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\boldsymbol{w}_{1,l}\boldsymbol{w}_{1,l}^*(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I}) + \|\boldsymbol{b}_{1,l}\|^2\sum_{j\neq 1}|\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}|^2\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^*\right) \\ &= \|\boldsymbol{b}_{1,l}\|^2\left(\sum_{j=1}^r\|\boldsymbol{w}_{j,l}\|^2\right)\boldsymbol{I} + \|\boldsymbol{b}_{1,l}\|^2\bar{\boldsymbol{w}}_{1,l}\bar{\boldsymbol{w}}_{1,l}^*.\end{aligned}$$

where $\mathbb{E}\left[(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\boldsymbol{w}_{1,l}\boldsymbol{w}_{1,l}^*(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\right] = \|\boldsymbol{w}_{1,l}\|^2\boldsymbol{I} + \bar{\boldsymbol{w}}_{1,l}\bar{\boldsymbol{w}}_{1,l}^*$ and $\mathbb{E}\,|\boldsymbol{w}_{j,l}^*\boldsymbol{a}_{j,l}|^2 = \|\boldsymbol{w}_{j,l}\|^2$.

$$\begin{aligned}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*) &= \mathbb{E}\left(\|(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\boldsymbol{w}_{1,l}\|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^* + \sum_{j\neq 1}\|\boldsymbol{a}_{1,l}\|^2|\langle\boldsymbol{w}_{j,l}, \boldsymbol{a}_{j,l}\rangle|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\right) \\ &= N_1\sum_{j=1}^r\|\boldsymbol{w}_{j,l}\|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^* + \|\boldsymbol{w}_{1,l}\|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\end{aligned}$$

where $\mathbb{E}\|(\boldsymbol{a}_{1,l}\boldsymbol{a}_{1,l}^* - \boldsymbol{I})\boldsymbol{w}_{1,l}\|^2 = (N_1+1)\|\boldsymbol{w}_{1,l}\|^2$ in (94) and $\mathbb{E}\|\boldsymbol{a}_{1,l}\|^2 = N_1$.

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l^*\mathcal{Z}_l)\| &\leq 2\sum_{l\in\Gamma_p}\left[\|\boldsymbol{b}_{1,l}\|^2\left(\sum_{j=1}^r\|\boldsymbol{w}_{j,l}\|^2\right)\right] \\
&\leq \frac{2\mu_{\max}^2K_1}{L}\sum_{j=1}^r\sum_{l\in\Gamma_p}\operatorname{Tr}(\boldsymbol{W}_{j,p-1}^*\boldsymbol{S}_{j,p}\boldsymbol{b}_{j,l}\boldsymbol{b}_{j,l}^*\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1}) \\
&\leq \frac{2\mu_{\max}^2K_1}{L}\sum_{j=1}^r\|\boldsymbol{W}_{j,p-1}\boldsymbol{W}_{j,p-1}^*\|_*\|\boldsymbol{S}_{j,p}\| \\
&\leq \frac{2\mu_{\max}^2K_1}{L}\frac{4L}{3Q}\sum_{j=1}^r\|\boldsymbol{W}_{j,p-1}\|_F^2 \\
&\leq C\frac{4^{-p+1}r\mu_{\max}^2K_1}{Q}.
\end{aligned}$$

where the last inequality follows from $\|\boldsymbol{W}_{i,p-1}\|_F \leq 2^{-p+1}$ and we also use the fact that $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathbb{E}(\mathcal{Z}_l\mathcal{Z}_l^*)\| &= \left\|\sum_{l\in\Gamma_p}\left[N_1\sum_{j=1}^r\|\boldsymbol{w}_{j,l}\|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^* + \|\boldsymbol{w}_{1,l}\|^2\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\right]\right\| \\
&\leq \max_{j,l}\|\boldsymbol{w}_{j,l}\|^2\cdot\left\|\sum_{l\in\Gamma_p}\left[rN_1\boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^* + \boldsymbol{b}_{1,l}\boldsymbol{b}_{1,l}^*\right]\right\| \\
&\leq \frac{\mu_{p-1}^2L}{Q^2}\cdot 2rN_1\|\boldsymbol{T}_{1,p}\| = \frac{5r\mu_{p-1}^2N_1}{2Q} \\
&\leq C\frac{4^{-p+1}r\mu_h^2N_1}{Q}
\end{aligned}$$

where $\|\boldsymbol{w}_{i,l}\| \leq \frac{\sqrt{L}\mu_{p-1}}{Q} \leq \frac{2^{-p+1}\sqrt{L}\mu_h}{Q}$ and $\|\boldsymbol{T}_{1,p}\| \leq \frac{5Q}{4L}$. Finally we have an upper bound of σ^2 as

$$\sigma^2 \leq C\frac{4^{-p+1}r\max\{\mu_{\max}^2K_1, \mu_h^2N_1\}}{Q} \leq C\frac{4^{-p+1}r\max\{\mu_{\max}^2K, \mu_h^2N\}}{Q}.$$

By using Bernstein inequality (46) with $t = \alpha\log L$ and $\log\left(\frac{\sqrt{Q}R}{\sigma}\right) \leq C_1\log L$, we have

$$\begin{aligned}
\|\sum_{l\in\Gamma_p}\mathcal{Z}_l\| \leq\ & C_0 2^{-p+1}\max\Big\{\sqrt{\alpha\frac{r\max\{\mu_{\max}^2K, \mu_h^2N\}}{Q}\log L}, \\
& \alpha\frac{r\max\{\mu_{\max}^2K, \mu_h^2N\}\log^2 L}{Q}\Big\}
\end{aligned}$$

In order to let $\|\sum_{l\in\Gamma_p}\mathcal{Z}_l\| \leq 2^{-p+1}$ with probability at least $1 - L^{-\alpha}$, it suffices to let $Q \geq C_\alpha r\max\{\mu_{\max}^2K, \mu_h^2N\}\log^2 L$. This finishes the proof for case when $i = 1$. Then we take the union bound over all p and $1 \leq i \leq r$, i.e., totally rP events and then

$$\|\mathcal{A}_{i,p}^*\boldsymbol{\lambda}_{p-1} - \boldsymbol{W}_{i,p-1}\| \leq 2^{-p-1}$$

holds simultaneously for all $1 \leq p \leq P$ and $1 \leq i \leq r$ 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

$\alpha' = \alpha + \log r$. In other words, $Q \geq C_\alpha r \max\{\mu_{\max}^2 K, \mu_h^2 N\} \log^2 L \log(r+1)$ makes

$$\|\mathcal{P}_{T_i^\perp}(\boldsymbol{Y}_{i,p})\| < \frac{1}{2}$$

hold simultaneously for $1 \leq i \leq r$ and $1 \leq p \leq P$ with probability at least $1 - L^{-\alpha+1}$.

■

7.4 Proof of $\mu_p \leq \frac{1}{2}\mu_{p-1}$

Recall that μ_p is defined in (62) as $\mu_p = \frac{Q}{\sqrt{L}} \max_{1\leq i\leq r, l\in\Gamma_{p+1}}(\|\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p+1}\boldsymbol{W}_{i,p}\|)$. 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 7.5. *Under the assumption of* (14), (10) *and* (12) *and that* $\boldsymbol{a}_{i,l} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N_i})$ *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* $Q \geq C_\alpha r^2 \max\{\mu_{\max}^2 K, \mu_h^2 N\} \log^2 L \log(r+1)$.

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

$$\frac{Q}{\sqrt{L}}\|\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p+1}\boldsymbol{W}_{i,p}\| \leq \frac{1}{2}\mu_{p-1} \tag{68}$$

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 $\frac{Q}{\sqrt{L}}\|\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p+1}\boldsymbol{W}_{i,p}\| \leq \frac{1}{2}\mu_{p-1}$ holds with high probability. Then taking the union bound over (i,l) completes the proof. Following from (60) and the definition of $\mathcal{A}_{j,p}$ in (17) give us

$$\begin{aligned}
-\boldsymbol{W}_{1,p} &= \mathcal{P}_{T_1}\left(\sum_{k\in\Gamma_p} \boldsymbol{b}_{1,k}\boldsymbol{b}_{1,k}^*\boldsymbol{S}_{1,p}\boldsymbol{W}_{1,p-1}(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\right) + \sum_{j\neq 1}\mathcal{P}_{T_1}\left(\sum_{k\in\Gamma_p} \boldsymbol{b}_{1,k}\boldsymbol{b}_{j,k}^*\boldsymbol{S}_{j,p}\boldsymbol{W}_{j,p-1}\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\right) \\
&= \mathcal{P}_{T_1}\left(\sum_{k\in\Gamma_p} \boldsymbol{b}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\right) + \sum_{j\neq 1}\mathcal{P}_{T_1}\left(\sum_{k\in\Gamma_p} \boldsymbol{b}_{1,k}\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\right) \\
&=: \Pi_1 + \Pi_2.
\end{aligned}$$

where $\boldsymbol{w}_{j,k} := \boldsymbol{W}_{j,p-1}^*\boldsymbol{S}_{j,p}\boldsymbol{b}_{j,k}$ defined in (63). By triangle inequality, it suffices to show

$$\|\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p+1}\Pi_1\| \leq \frac{\sqrt{L}}{4Q}\mu_{p-1}, \quad \|\boldsymbol{b}_{i,l}^*\boldsymbol{S}_{i,p+1}\Pi_2\| \leq \frac{\sqrt{L}}{4Q}\mu_{p-1} \tag{69}$$

so that (68) holds.

Step 1: proof of $\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_1\| \leq \frac{\sqrt{L}\mu_{p-1}}{4Q}$ For a fixed $l \in \Gamma_{p+1}$,

$$\begin{aligned}
\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_1 &= \sum_{k\in\Gamma_p} \boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Big[\boldsymbol{h}_1\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I}) \\
&\quad + (\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\boldsymbol{x}_1^*\Big]
\end{aligned}$$

where $\mathcal{P}_{T_1}$ has an explicit form in (23). Define

$$\boldsymbol{z}_k := (\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\boldsymbol{b}_{1,k}^*\boldsymbol{h}_1\boldsymbol{h}_1^*\boldsymbol{S}_{1,p+1}\boldsymbol{b}_{1,l} \in \mathbb{C}^{N_1} \tag{70}$$

and

$$z_k := \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}(\boldsymbol{I} - \boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1. \tag{71}$$

Then by the triangle inequality,

$$\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_1\| \le \|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\| + |\sum_{k\in\Gamma_k} z_k|. \tag{72}$$

Our goal now is to bound both $\|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\|$ and $|\sum_{k\in\Gamma_k} z_k|$ by $\frac{\sqrt{L}\mu_{p-1}}{8Q}$. First we take a look at $\sum_{k\in\Gamma_p}\boldsymbol{z}_k$. For each k,

$$\begin{aligned}
\|\boldsymbol{z}_k\|_{\psi_1} &= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|\cdot|\langle\boldsymbol{h}_1,\boldsymbol{b}_{1,k}\rangle|\cdot\|(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\|_{\psi_1}\\
&\le C\frac{\sqrt{L}\mu_h}{Q}\frac{\mu_h}{\sqrt{L}}\sqrt{N_1}\|\boldsymbol{w}_{1,k}\| = C\frac{\mu_h^2\sqrt{N_1}\|\boldsymbol{w}_{1,k}\|}{Q}.
\end{aligned}$$

which follows from (14) and $\|(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\|_{\psi_1} \le C\sqrt{N_1}\|\boldsymbol{w}_{1,k}\|$ in (98). The expectation of $\mathbb{E}(\boldsymbol{z}_k^*\boldsymbol{z}_k)$ and $\mathbb{E}(\boldsymbol{z}_k\boldsymbol{z}_k^*)$ can be easily computed,

$$\begin{aligned}
\mathbb{E}(\boldsymbol{z}_k^*\boldsymbol{z}_k) &= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\,\mathbb{E}[\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})^2\boldsymbol{w}_{1,k}]\\
&= (N_1+1)|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\|\boldsymbol{w}_{1,k}\|^2\\
\mathbb{E}(\boldsymbol{z}_k\boldsymbol{z}_k^*) &= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\,\mathbb{E}[(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})]\\
&= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2(\|\boldsymbol{w}_{1,k}\|^2\boldsymbol{I} + \bar{\boldsymbol{w}}_{1,k}\bar{\boldsymbol{w}}_{1,k}^*).
\end{aligned}$$

where $\mathbb{E}[\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})^2\boldsymbol{w}_{1,k}] = (N_1+1)\|\boldsymbol{w}_{1,k}\|^2$ and $\mathbb{E}[(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})] = \|\boldsymbol{w}_{1,k}\|^2\boldsymbol{I} + \bar{\boldsymbol{w}}_{1,k}\bar{\boldsymbol{w}}_{1,k}^*$ follow from (94) and (99).

$$\begin{aligned}
\|\sum_{k\in\Gamma_p}\mathbb{E}(\boldsymbol{z}_k^*\boldsymbol{z}_k)\| &\le (N_1+1)|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2\max_{k\in\Gamma_p}\{\|\boldsymbol{w}_{1,k}\|^2\}\sum_{k\in\Gamma_p}|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\\
&\le \frac{(N_1+1)L\mu_h^2}{Q^2}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2\|\boldsymbol{T}_{1,p}\|\\
&\le \frac{(N_1+1)L\mu_h^2}{Q^2}\frac{5Q}{4L}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2 = \frac{5\mu_h^2N_1\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2}{2Q}.
\end{aligned}$$

The estimation of $\|\sum_{k\in\Gamma_p}\mathbb{E}(\boldsymbol{z}_k\boldsymbol{z}_k^*)\|$ is quite similar to that of $\|\sum_{k\in\Gamma_p}\mathbb{E}(\boldsymbol{z}_k^*\boldsymbol{z}_k)\|$ and thus we give the result directly without going to the details,

$$\|\sum_{k\in\Gamma_p}\mathbb{E}(\boldsymbol{z}_k\boldsymbol{z}_k^*)\| \le \frac{5\mu_h^2\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2}{2Q}.$$

Therefore,

$$R := \max_{k\in\Gamma_p}\|\boldsymbol{z}_k\|_{\psi_1} \le C\frac{\mu_h^2\sqrt{N}}{Q}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|$$

and similarly, we have

$$\sigma^2 \le C\frac{\mu_h^2N\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2}{Q}.$$

Then we just apply (46) with $t = \alpha\log L$ and $\log(\sqrt{Q}R/\sigma) \le C_1\log L$ to estimate $\|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\|$,

$$\|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\| \le C\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2\max\left\{\sqrt{\frac{\alpha\mu_h^2N}{Q}\log L}, \frac{\alpha\mu_h^2\sqrt{N}}{Q}\log^2 L)\right\}. \tag{73}$$

Note that $\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\| \leq \frac{\sqrt{L}\mu_{p-1}}{Q}$ in (64) and thus it suffices to let $Q \geq C_\alpha \mu_h^2 N \log^2 L$ to ensure that $\|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\| \leq \frac{\sqrt{L}\mu_{p-1}}{8Q}$ holds with probability at least $1 - L^{-\alpha}$.

Now consider z_k in (71) by first computing its $\|z_k\|_{\psi_1}$,

$$
\begin{aligned}
\|z_k\|_{\psi_1} &= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}| \cdot \|\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\|_{\psi_1} \\
&= \|\boldsymbol{b}_{1,l}\|\|\boldsymbol{S}_{1,p+1}\|\|\boldsymbol{b}_{1,k}\|\|\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\|_{\psi_1} \\
&\leq C\frac{\mu_{\max}^2 K_1}{L}\cdot\frac{4L}{3Q}\|\boldsymbol{w}_{1,k}\| \\
&\leq C\frac{\mu_{\max}^2 K_1}{Q}\|\boldsymbol{w}_{1,k}\|
\end{aligned}
$$

where $\|\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\|_{\psi_1} \leq C\|\boldsymbol{w}_{1,k}\|$ in (100). Thus $R := \max\{\|\boldsymbol{z}_k\|_{\psi_1}\} \leq C\frac{\mu_{\max}^2 K_1}{Q}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|$.

$$
\begin{aligned}
\mathbb{E}\, z_k^2 &= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}|^2\,\mathbb{E}\left[\boldsymbol{w}_{1,k}^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\boldsymbol{x}_1^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{w}_{1,k}\right] \\
&= |\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}|^2\boldsymbol{w}_{1,k}^*(\boldsymbol{I}+\boldsymbol{x}_1\boldsymbol{x}_1^*)\boldsymbol{w}_{1,k}
\end{aligned}
$$

where $\mathbb{E}((\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})\boldsymbol{x}_1\boldsymbol{x}_1^*(\boldsymbol{a}_{1,k}\boldsymbol{a}_{1,k}^* - \boldsymbol{I})) = \boldsymbol{I}+\boldsymbol{x}_1\boldsymbol{x}_1^*$ follows from (99). The variance $\sum_{k\in\Gamma_p}|z_k|^2$ is bounded by

$$
\begin{aligned}
\sigma^2 &\leq \boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1)\boldsymbol{T}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p}\boldsymbol{b}_{1,l}\max_{k\in\Gamma_p}\boldsymbol{w}_{1,k}^*(\boldsymbol{I}+\boldsymbol{x}_1\boldsymbol{x}_1^*)\boldsymbol{w}_{1,k} \\
&\leq \|\boldsymbol{b}_{1,l}\|^2\|\boldsymbol{S}_{1,p}\|^2\|\boldsymbol{T}_{1,p}\|\max_{k\in\Gamma_p}\boldsymbol{w}_{1,k}^*(\boldsymbol{I}+\boldsymbol{x}_1\boldsymbol{x}_1^*)\boldsymbol{w}_{1,k} \\
&\leq 2\frac{\mu_{\max}^2 K_1}{L}\frac{16L^2}{9Q^2}\frac{5Q}{4L}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2 \\
&= \frac{40\mu_{\max}^2 K_1}{9Q}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2.
\end{aligned}
$$

Similar to what we have done in (73),

$$
|\sum_{k\in\Gamma_p} z_k| \leq C\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\|^2\max\left\{\sqrt{\frac{\alpha\mu_{\max}^2 K}{Q}\log L}, \frac{\alpha\mu_{\max}^2 K}{Q}\log^2 L)\right\} \tag{74}
$$

Note that $\max_{k\in\Gamma_p}\|\boldsymbol{w}_{1,k}\| \leq \frac{\sqrt{L}\mu_{p-1}}{Q}$ and thus $Q \geq C_\alpha\mu_{\max}^2 K\log^2 L$ guarantees that $|\sum_{k\in\Gamma_p} z_k| \leq \frac{\sqrt{L}\mu_{p-1}}{8Q}$ holds with probability at least $1 - L^{-\alpha}$. Combining (73) and (74) gives

$$
\mathbb{P}\left(\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p}\Pi_1\| \geq \frac{\sqrt{L}\mu_{p-1}}{4Q}\right) \leq \mathbb{P}\left(\|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\| \geq \frac{\sqrt{L}\mu_{p-1}}{8Q}\right) + \mathbb{P}\left(|\sum_{k\in\Gamma_p} z_k| \geq \frac{\sqrt{L}\mu_{p-1}}{8Q}\right) \leq 2L^{-\alpha}, \tag{75}
$$

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

Step 2: proof of $\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_2\| \leq \frac{\sqrt{L}\mu_{p-1}}{4Q}$ For any fixed $l \in \Gamma_{p+1}$,

$$
\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_2 = \boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\sum_{j\neq 1}\mathcal{P}_{T_1}\left(\sum_{k\in\Gamma_p}\boldsymbol{b}_{1,k}\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\right).
$$

Now we rewrite $\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\Pi_2$ into

$$\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\Pi_2 = \sum_{j\neq 1}\left(\sum_{k\in\Gamma_p} \boldsymbol{z}_{j,k}^* + z_{j,k}\boldsymbol{x}_1^*\right)$$

where

$$\begin{aligned} \boldsymbol{z}_{j,k} &:= \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\boldsymbol{h}_1\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k} && (76)\\ z_{j,k} &:= \boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\boldsymbol{x}_1 && (77)\end{aligned}$$

By triangle inequality,

$$\|\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\Pi_2\| \leq \sum_{j\neq 1, j\leq r}\left[\|\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}\| + |\sum_{k\in\Gamma_p} z_{j,k}|\right]. \tag{78}$$

In order to bound $\|\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\Pi_2\|$ by $\frac{\sqrt{L}\mu_{p-1}}{4Q}$, it suffices to prove that for all $1\leq j\leq r$,

$$\|\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}\| \leq \frac{\sqrt{L}\mu_{p-1}}{8rQ}, \quad |\sum_{k\in\Gamma_p} z_{j,k}| \leq \frac{\sqrt{L}\mu_{p-1}}{8rQ}. \tag{79}$$

For $\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}$,

$$\begin{aligned} \|\boldsymbol{z}_{j,k}\|_{\psi_1} &\leq |\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\boldsymbol{h}_1||\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|(|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}|\cdot\|\boldsymbol{a}_{1,k}^*\|)_{\psi_1}\\ &\leq C\frac{\sqrt{L}\mu_h}{Q}\frac{\mu_h}{\sqrt{L}}\sqrt{N_1}\|\boldsymbol{w}_{j,k}\|\\ &\leq C\frac{\mu_h^2\sqrt{N_1}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|}{Q}. \end{aligned}$$

where $(|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}|\cdot\|\boldsymbol{a}_{1,k}^*\|)_{\psi_1} \leq C\sqrt{N_1}\|\boldsymbol{w}_{j,k}\|$ follows from Lemma 12.4. Now we move on to the estimation of σ^2.

$$\begin{aligned} \|\sum_{k\in\Gamma_p}\mathbb{E}\,\boldsymbol{z}_{j,k}^*\boldsymbol{z}_{j,k}\| &= \sum_{k\in\Gamma_p}|\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\boldsymbol{h}_1\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\,\mathbb{E}\left[|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}|^2\|\boldsymbol{a}_{1,k}\|^2\right]\\ &= N_1\sum_{k\in\Gamma_p}|\boldsymbol{b}_{1,l}^* \boldsymbol{S}_{1,p+1}\boldsymbol{h}_1|^2|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2|\|\boldsymbol{w}_{j,k}\|^2\\ &\leq N_1\frac{L\mu_h^2}{Q^2}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\sum_{k\in\Gamma_p}|\boldsymbol{h}_1^*\boldsymbol{b}_{1,k}|^2\\ &\leq N_1\frac{L\mu_h^2}{Q^2}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\|\boldsymbol{T}_{1,p}\|\\ &\leq \frac{5\mu_h^2 N_1\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2}{4Q} \end{aligned}$$

and similarly,

$$\|\sum_{k\in\Gamma_p}\mathbb{E}\,\boldsymbol{z}_{j,k}\boldsymbol{z}_{j,k}^*\| \leq \frac{5\mu_h^2\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2}{4Q}.$$

Thus $\sigma^2 \leq C\frac{\mu_h^2 N_1\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2}{Q}$. By applying Bernstein inequality (46), we have

$$\|\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}\| \leq C\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\max\left\{\sqrt{\frac{\alpha\mu_h^2 N}{Q}\log L}, \frac{\alpha\mu_h^2 N}{Q}\log^2 L)\right\}$$

where $\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\| \leq \frac{\sqrt{L}\mu_{p-1}}{Q}$. Choosing $Q \geq C_\alpha r^2\mu_h^2 N\log^2 L$ leads to

$$\|\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}\| \leq \frac{\sqrt{L}\mu_{p-1}}{8rQ} \tag{80}$$

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 (77) and fixed j,

$$\begin{aligned} R : &= \max|z_{j,k}| \leq \max_{k\in\Gamma_p}\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}\| \cdot \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\boldsymbol{x}_1\|_{\psi_1} \\ &\leq C\frac{\mu_{\max}^2 K_1}{L}\frac{4L}{3Q}\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\| = C\frac{\mu_{\max}^2 K_1 \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|}{Q} \end{aligned}$$

where $\|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\boldsymbol{x}_1\|_{\psi_1} \leq C\|\boldsymbol{w}_{j,k}\|$ follows from Lemma 12.4. Now we proceed to compute the variance by

$$\begin{aligned} \sigma^2 &:= \sum_{k\in\Gamma_p}\mathbb{E}\,|z_{j,k}|^2 = \sum_{k\in\Gamma_p}|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}|^2\,\mathbb{E}\,|\boldsymbol{w}_{j,k}^*\boldsymbol{a}_{j,k}\boldsymbol{a}_{1,k}^*\boldsymbol{x}_1|^2 \\ &= \sum_{k\in\Gamma_p}|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}|^2\|\boldsymbol{w}_{j,k}\|^2 \\ &\leq \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\sum_{k\in\Gamma_p}|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{b}_{1,k}|^2 \\ &\leq \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{T}_{1,p}(\boldsymbol{I}-\boldsymbol{h}_1\boldsymbol{h}_1^*)\boldsymbol{S}_{1,p+1}\boldsymbol{b}_{1,l} \\ &\leq \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\|\boldsymbol{S}_{1,p+1}\|^2\|\boldsymbol{T}_{1,p}\|\|\boldsymbol{b}_{1,l}\|^2 \\ &\leq \max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\frac{16L^2}{9Q^2}\frac{5Q}{4L}\frac{\mu_{\max}^2 K_1}{L} \leq C\frac{\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\mu_{\max}^2 K_1}{Q}. \end{aligned}$$

Then we apply Bernstein inequality to get an upper bound of $|\sum_k z_{j,k}|$ for fixed j,

$$|\sum_{k\in\Gamma_p} z_{j,k}| \leq C\max_{k\in\Gamma_p}\|\boldsymbol{w}_{j,k}\|^2\max\left\{\sqrt{\frac{\alpha\mu_{\max}^2 K}{Q}\log L}, \frac{\alpha\mu_{\max}^2 K}{Q}\log^2 L)\right\} \leq \frac{\sqrt{L}\mu_{p-1}}{8rQ}.$$

with probability $1 - L^{-\alpha}$ if $Q \geq C_\alpha r^2\mu_{\max}^2 K\log^2 L$. Thus combined with (80), we have proven that for fixed j,

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

holds with probability at least $1 - 2L^{-\alpha}$. By taking union bound over $1 \leq j \leq r$ and using (78), we can conclude that

$$\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\Pi_2\| \leq \frac{\sqrt{L}\mu_{p-1}}{4Q}$$

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

Final step: Proof of (68) To sum up, we have already shown that for fixed $l \in \Gamma_p$ and $i = 1$,

$$\frac{Q}{\sqrt{L}}\|\boldsymbol{b}_{1,l}^*\boldsymbol{S}_{1,p+1}\boldsymbol{W}_{1,p}\| \leq \|\sum_{k\in\Gamma_p}\boldsymbol{z}_k\| + |\sum_{k\in\Gamma_k} z_k| + \sum_{j\neq 1}\left[\|\sum_{k\in\Gamma_p}\boldsymbol{z}_{j,k}\| + |\sum_{k\in\Gamma_p} z_{j,k}|\right] \leq \frac{1}{2}\mu_{p-1}$$

with probability at least $1-(r+2)L^{-\alpha}$ if $Q \geq C_\alpha r^2 \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L$. 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

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

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

■

8 Proof of the Main Theorem

We now assemble the various intermediate and auxiliary results to establish Theorem 3.1. We recall that Theorem 3.1 follows immediately from Lemma 4.2, which in turn hinges on the validity of the conditions (26) and (27). Let us focus on condition (26) first, i.e., we need to show that

$$\max_i \|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_i\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \leq \frac{1}{4}, \tag{81}$$

$$\max_{j\neq k} \|\mathcal{P}_{T_j}\mathcal{A}_j^*\mathcal{A}_k\mathcal{P}_{T_k}\| \leq \frac{1}{4r}, \tag{82}$$

$$\max_i \|\mathcal{A}_i\| \leq \gamma \tag{83}$$

Under the assumptions of Theorem 3.1, Proposition 5.1 ensures that condition (81) holds with probability at least $1-L^{-\alpha+1}$ if $Q \geq C_\alpha \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L\log(r+1)$ where $K := \max K_i$ and $N := \max N_i$. Moving on to the incoherence condition (82), Proposition 6.1 implies that this condition holds with probability at least $1-L^{-\alpha+1}$ if $Q \geq C_\alpha r^2 \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L\log(r+1)$. Furthermore, γ in condition (83) 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 (27). Under the assumption that properties (41) and (53) hold, Lemma 7.1 implies the first part of condition (27). The two properties (41) and (53) have been established in Propositions 5.1 and 6.1, respectively. The second part of the approximate dual certificate condition in (27) is established in Lemma 7.3 with the aid of Lemma 7.5, with probability at least $1-2L^{-\alpha+1}$ if $Q \geq C_\alpha r^2 \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L\log(r+1)$.
By "summing up" all the probabilities of failure in each substep,

$$\mathbb{P}(\hat{\boldsymbol{X}}_i = \boldsymbol{X}_i, \forall 1 \leq i \leq r) \geq 1 - 5L^{-\alpha+1}$$

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

$$L \geq C_\alpha r^2 \max\{\mu^2_{\max}K, \mu^2_h N\}\log^2 L\log\gamma\log(r+1)$$

with $\gamma \leq \sqrt{N\log(NL/2) + \alpha\log L}$. Thus, the sufficient conditions stated in Lemma 4.2 are fulfilled with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$, hence Theorem 3.1 follows now directly from Lemma 4.2.

9 Stability theory – Proof of Theorem 3.3

9.1 Notation

Since we do not assume $\{\boldsymbol{X}_i\}_{i=1}^r$ are of the same size, notation will be an issue during the discussion. We introduce a few useful notations in order to make the derivations easier. Recall

$\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{Z}_i)$ 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 $\boldsymbol{\Phi} := [\boldsymbol{\Phi}_1|\cdots|\boldsymbol{\Phi}_r]$ with $\boldsymbol{\Phi}_i \in \mathbb{C}^{L\times K_iN_i}$ and $\boldsymbol{\Phi}\in\mathbb{C}^{L\times\sum_{i=1}^r K_iN_i}$ as

$$\boldsymbol{\Phi}_i \operatorname{vec}(\boldsymbol{Z}_i) := \operatorname{vec}(\mathcal{A}_i(\boldsymbol{Z}_i)), \quad \boldsymbol{\Phi}\begin{bmatrix}\operatorname{vec}(\boldsymbol{Z}_1)\\ \vdots \\ \operatorname{vec}(\boldsymbol{Z}_r)\end{bmatrix} := \operatorname{vec}(\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{Z}_i)).$$

where $\boldsymbol{Z}_i\in\mathbb{C}^{K_i\times N_i}$. The operation "vec" vectorizes a matrix into a column vector. $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}_i$ are well-defined and can verified with a little knowledge of block matrix. It could be be shown by slightly modifying the proof of Lemma 2 in [2] that

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^* = \sum_{i=1}^r \boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^* \in\mathbb{C}^{L\times L}$$

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

$$0.48\mu^2_{\min}\frac{\sum_{i=1}^r K_iN_i}{L} \le \lambda^2_{\min} \le \lambda^2_{\max} \le 4.5\mu^2_{\max}\frac{\sum_{i=1}^r K_iN_i}{L} \tag{84}$$

with probability at least $1-O(L^{-\alpha+1})$ if $\sum_{i=1}^r K_iN_i \ge \frac{C_\alpha}{\mu^2_{\min}} L\log^2 L$ with $\mu^2_{\min}$ defined in (9). Note that $\sum_{i=1}^r K_iN_i$ is usually much larger than L in applications.

Let $\boldsymbol{E}_i = \hat{\boldsymbol{X}}_i - \boldsymbol{X}_i \in\mathbb{C}^{K_i\times N_i}, 1\le i\le r$ be the difference between $\hat{\boldsymbol{X}}_i$ and $\boldsymbol{X}_i$. Define

$$\boldsymbol{e}_i := \operatorname{vec}(\boldsymbol{E}_i), \quad \boldsymbol{e} := \begin{bmatrix}\boldsymbol{e}_1\\ \vdots\\ \boldsymbol{e}_r\end{bmatrix} \in \mathbb{C}^{(\sum_{i=1}^r K_iN_i)\times 1},$$

where $\boldsymbol{e}$ is a long vector consisting of all $\boldsymbol{e}_i$, $1\le i\le r$. We also consider $\boldsymbol{e}$ being projected on $\operatorname{Ran}(\boldsymbol{\Phi}^*)$, denoted by $\boldsymbol{e}_{\boldsymbol{\Phi}}$,

$$\boldsymbol{e}_{\boldsymbol{\Phi}} := \boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{e}$$

where $\boldsymbol{\Phi}\boldsymbol{e} = \sum_{i=1}^r \boldsymbol{\Phi}_i\boldsymbol{e}_i = \sum_{i=1}^r \mathcal{A}_i(\boldsymbol{E}_i)$. From (21), we know that

$$\|\boldsymbol{\Phi}\boldsymbol{e}\|_F = \|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{E}_i)\|_F \le \|\sum_{i=1}^r \mathcal{A}_i(\hat{\boldsymbol{X}}_i) - \hat{\boldsymbol{y}}\|_F + \|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{X}_i) - \hat{\boldsymbol{y}}\|_F \le 2\eta. \tag{85}$$

since both $\{\hat{\boldsymbol{X}}_i\}_{i=1}^r$ and $\{\boldsymbol{X}_i\}_{i=1}^r$ are inside the feasible set. Similarly, define $\boldsymbol{e}_{\boldsymbol{\Phi}^\perp} = \boldsymbol{e} - \boldsymbol{e}_{\boldsymbol{\Phi}} \in \operatorname{Null}(\boldsymbol{\Phi})$ and denote $\boldsymbol{H}_i\in\mathbb{C}^{K_i\times N_i}$ and $\boldsymbol{J}_i\in\mathbb{C}^{K_i\times N_i}$, $1\le i\le r$, as matrices satisfying

$$\boldsymbol{e}_{\boldsymbol{\Phi}^\perp} := \begin{bmatrix}\operatorname{vec}(\boldsymbol{H}_1)\\ \vdots\\ \operatorname{vec}(\boldsymbol{H}_r)\end{bmatrix}, \quad \boldsymbol{e}_{\boldsymbol{\Phi}} := \begin{bmatrix}\operatorname{vec}(\boldsymbol{J}_1)\\ \vdots\\ \operatorname{vec}(\boldsymbol{J}_r)\end{bmatrix} \tag{86}$$

where $\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_i) = \boldsymbol{\Phi}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp} = \boldsymbol{0}$ and $\boldsymbol{H}_i + \boldsymbol{J}_i = \boldsymbol{E}_i$ follows from the definition of $\boldsymbol{H}_i$ and $\boldsymbol{J}_i$. Define $\boldsymbol{P}_{T_i}$ as the projection matrix from $\operatorname{vec}(\boldsymbol{Z})$ to $\operatorname{vec}(\mathcal{P}_{T_i}(\boldsymbol{Z}))$, as

$$\boldsymbol{P}_{T_i}\operatorname{vec}(\boldsymbol{Z}) = \operatorname{vec}(\mathcal{P}_{T_i}(\boldsymbol{Z})), \quad \boldsymbol{P}_{T_i}\in\mathbb{C}^{(K_iN_i)\times(K_iN_i)}$$

and

$$\boldsymbol{P}_T := \begin{bmatrix}\boldsymbol{P}_{\boldsymbol{T}_1} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{P}_{\boldsymbol{T}_r}\end{bmatrix} \quad \boldsymbol{P}_{T^\perp} := \begin{bmatrix}\boldsymbol{I}_{K_1N_1} - \boldsymbol{P}_{\boldsymbol{T}_1} & \cdots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{I}_{K_rN_r} - \boldsymbol{P}_{\boldsymbol{T}_r}\end{bmatrix}$$

Actually the definitions above immediately give the following equations:

$$\boldsymbol{P}_T\boldsymbol{e} = \begin{bmatrix}\boldsymbol{P}_{T_1}\boldsymbol{e}_1\\ \vdots \\ \boldsymbol{P}_{T_r}\boldsymbol{e}_r\end{bmatrix}, \quad \boldsymbol{P}_T\boldsymbol{e}_{\boldsymbol{\Phi}^\perp} = \begin{bmatrix}\mathrm{vec}(\boldsymbol{H}_{1,T_1})\\ \vdots \\ \mathrm{vec}(\boldsymbol{H}_{r,T_r})\end{bmatrix}, \quad \boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp} = \begin{bmatrix}\mathrm{vec}(\boldsymbol{H}_{1,T_1^\perp})\\ \vdots \\ \mathrm{vec}(\boldsymbol{H}_{r,T_r^\perp})\end{bmatrix} \tag{87}$$

9.2 Proof of Theorem 3.3

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

$$\|\boldsymbol{e}\| \leq C\frac{r\lambda_{\max}\sqrt{\max\{K,N\}}}{\lambda_{\min}(1-\beta-2r\gamma\alpha)}\eta.$$

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

$$\|\boldsymbol{e}\|_F^2 = \|\boldsymbol{e}_{\boldsymbol{\Phi}}\|^2 + \|\boldsymbol{P}_T\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 + \|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 \tag{88}$$

Following from (87), (30) and (29) gives an estimate of the second term in (88),

$$\begin{aligned}
\|\boldsymbol{P}_T\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 &= \sum_{i=1}^r \|\boldsymbol{H}_{i,T_i}\|_F^2 \leq 2\left\|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_{i,T_i})\right\|_F^2 \\
&= 2\left\|\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_{i,T_i^\perp})\right\|_F^2 \leq 2\gamma^2\left(\sum_{i=1}^r\|\boldsymbol{H}_{i,T_i^\perp}\|_F\right)^2 \\
&\leq 2r\gamma^2\sum_{i=1}^r\|\boldsymbol{H}_{i,T_i^\perp}\|_F^2 \leq 2r\gamma^2\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 \\
&\leq 2r\lambda_{\max}^2\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2.
\end{aligned}$$

where $\max\|\mathcal{A}_i\| \leq \gamma$, $\lambda_{\max}^2$ is largest eigenvalue of $\boldsymbol{\Phi}\boldsymbol{\Phi}^*$ and obviously $\gamma \leq \lambda_{\max}$. The second equality holds since $\sum_{i=1}^r \mathcal{A}_i(\boldsymbol{H}_i) = \boldsymbol{0}$. For the third term in (88), by reversing the arguments in the proof of Lemma 4.2, we have

$$\begin{aligned}
\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F &= \sqrt{\sum_{i=1}^r\|\boldsymbol{H}_{i,T_i^\perp}\|_F^2} \leq \sum_{i=1}^r\|\boldsymbol{H}_{i,T_i^\perp}\|_F \\
&\leq \frac{1}{1-\beta-2r\gamma\alpha}\sum_{i=1}^r\langle\boldsymbol{H}_i,\boldsymbol{h}_i\boldsymbol{x}_i^T\rangle + \|\boldsymbol{H}_{i,T_i^\perp}\|_* \\
&\leq \frac{1}{1-\beta-2r\gamma\alpha}\sum_{i=1}^r[\|\boldsymbol{X}_i+\boldsymbol{H}_i\|_* - \|\boldsymbol{X}_i\|_*] \\
&\leq \frac{1}{1-\beta-2r\gamma\alpha}\sum_{i=1}^r\left[\|\boldsymbol{X}_i+\boldsymbol{H}_i\|_* - \|\hat{\boldsymbol{X}}_i\|_*\right]
\end{aligned}$$

where the first equality comes from (87), the third inequality is due to Lemma 4.1 and the last inequality follows from $\sum_{i=1}^r\|\hat{\boldsymbol{X}}_i\|_* \leq \sum_{i=1}^r\|\boldsymbol{X}_i\|_*$ in (22). From the definition of $\boldsymbol{H}_i$ and $\boldsymbol{J}_i$ in (86), $\hat{\boldsymbol{X}}_i = \boldsymbol{X}_i + \boldsymbol{E}_i = \boldsymbol{X}_i + \boldsymbol{H}_i + \boldsymbol{J}_i$ and triangle inequality gives,

$$\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F \leq \frac{1}{1-\beta-2r\gamma\alpha}\sum_{i=1}^r\|\boldsymbol{J}_i\|_* \leq \frac{\sqrt{\max\{K,N\}}}{1-\beta-2r\gamma\alpha}\sum_{i=1}^r\|\boldsymbol{J}_i\|_F.$$

In other words,

$$\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 \leq \frac{r\max\{K,N\}}{(1-\beta-2r\gamma\alpha)^2}\sum_{i=1}^r \|\boldsymbol{J}_i\|_F^2 \leq \frac{r\max\{K,N\}}{(1-\beta-2r\gamma\alpha)^2}\|\boldsymbol{e}_{\boldsymbol{\Phi}}\|_F^2 \tag{89}$$

where $\|\boldsymbol{e}_{\boldsymbol{\Phi}}\|_F^2 = \sum_{i=1}^r \|\boldsymbol{J}_i\|_F^2$ follows from (86). By combining all those estimations together, i.e., $\|\boldsymbol{P}_T\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 \leq 4r\lambda_{\max}^2\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2$, (89) and (88), we arrive at

$$\begin{aligned}\|\boldsymbol{e}\|_F^2 &\leq \|\boldsymbol{e}_{\boldsymbol{\Phi}}\|_F^2 + (2r\lambda_{\max}^2+1)\|\boldsymbol{P}_{T^\perp}\boldsymbol{e}_{\boldsymbol{\Phi}^\perp}\|_F^2 \\ &\leq C\frac{r^2\lambda_{\max}^2\max\{K,N\}}{(1-\beta-2r\gamma\alpha)^2}\|\boldsymbol{e}_{\boldsymbol{\Phi}}\|_F^2\end{aligned}$$

Note that $\boldsymbol{e}_{\boldsymbol{\Phi}} := \boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{e}$,

$$\|\boldsymbol{e}_{\boldsymbol{\Phi}}\|_F \leq \frac{1}{\lambda_{\min}}\|\boldsymbol{\Phi}\boldsymbol{e}\|_F$$

where $\lambda_{\min}^2$ is the smallest eigenvalue of $\boldsymbol{\Phi}\boldsymbol{\Phi}^*$. By applying $\|\boldsymbol{\Phi}\boldsymbol{e}\| \leq 2\eta$ in (85), we have

$$\|\boldsymbol{e}\|_F \leq C\frac{r\lambda_{\max}\sqrt{\max\{K,N\}}}{\lambda_{\min}(1-\beta-2r\gamma\alpha)}\|\boldsymbol{\Phi}\boldsymbol{e}\|_F \leq C\frac{r\lambda_{\max}\sqrt{\max\{K,N\}}}{\lambda_{\min}(1-\beta-2r\gamma\alpha)}\eta.$$

In particular, if we choose $\alpha = (5r\gamma)^{-1}$ and $\beta = \frac{1}{2}$ according to Lemma 4.2, then $\frac{1}{1-\beta-2r\gamma\alpha} = 10$. This completes the proof of Theorem 3.3.
■

10 Numerical Simulations

10.1 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 $\boldsymbol{B}_i$ is the first K_i columns of an $L \times L$ DFT matrices with $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_{K_i}$ and each $\boldsymbol{A}_i$ is an $L \times N_i$ Gaussian random matrix. $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ yield $\mathcal{N}(0, \boldsymbol{I}_{K_i})$ and $\mathcal{N}(0, \boldsymbol{I}_{N_i})$ respectively. We denote $\boldsymbol{X}_i = \boldsymbol{h}_i\boldsymbol{x}_i^T$, the "lifted" matrix and solve (8) to recover $\boldsymbol{X}_i$. For each pair of (L, r), 10 experiments are performed and the recovery is regarded as a success if

$$\frac{\sqrt{\sum_{i=1}^r \|\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\|_F^2}}{\sqrt{\sum_{i=1}^r \|\boldsymbol{X}_i\|_F^2}} < 10^{-3} \tag{90}$$

where each $\hat{\boldsymbol{X}}_i$, given by solving (8) via CVX package [18] on MATLAB, serves as an approximation of $\boldsymbol{X}_i$. 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 $\boldsymbol{X}_i$ 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 performance if $\boldsymbol{A}_i = \boldsymbol{D}_i\boldsymbol{H}_i$ where $\boldsymbol{D}_i$ is a diagonal matrix with Bernoulli random variables (taking value ± 1 with equal probabilities) on the diagonal and $\boldsymbol{H}_i$ is fixed as the first N_i columns of a non-random Hadamard matrix. In other words, the only randomness of $\boldsymbol{A}_i$ comes from $\boldsymbol{D}_i$. Both $\boldsymbol{H}_i$ and $\boldsymbol{D}_i$ are matrices of ± 1 entries and can be easily generated in many applications. By using the same settings on L, r, $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ as before and $K_i = N_i = 15$, we apply (8) to recover $(\boldsymbol{h}_i, \boldsymbol{x}_i)_{i=1}^r$. Since the existence of Hadamard matrices of order $4k$ with positive integer k is still an open problem [20], we only test $L = 2^s$ with $s = 6, 7, 8$ and 9. Surprisingly, Figure 1 (the bottom one) also demonstrates that the minimal L scales linearly with r and our algorithm almost reaches the information theoretic optimum even if all $\boldsymbol{A}_i$ are partial Hadamard matrices.

Figure 2 shows the performance of recovery via solving (8) under the assumption that L is fixed and K_i and N_i are changing. The results are presented for two cases: (i) the $\boldsymbol{A}_i$ are Gaussian random matrices, and (ii) the $\boldsymbol{A}_i$ are Hadamard matrices premultiplied by a binary diagonal matrix as explained above. In the simulations, we assume there exist two sources ($r = 2$) with $K_1 = K_2$ and $N_1 = N_2$. We fix $L = 128$ and let K_i and N_i vary from 5 to 50. $\boldsymbol{B}_i$ consists of the first K_i columns of an $L \times L$ DFT matrix. Both $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ are random Gaussian vectors. The boundary between success and failure in the phase transition plot is well approximated by a line, which matches the relationship between L, K_i, and N_i stated in Theorem 3.1. More precisely, the probability of success is quite satisfactory if $L = 128 \geq 1.5r(K_i + N_i)$ in this case.

10.2 Number of measurements L vs. the incoherence parameter μ_h^2

Theorem 3.1 indicates that L scales with μ_h^2 defined in (13) and μ_h^2 also plays an important role in the proof. Moreover, Figure 3 implies that μ_h^2 is not only necessary for "technical reasons" but also related to the numerical performance. In the experiment, we fix $r = 1$ and $K = N = 30$. $\boldsymbol{A}$ is a Gaussian random matrix, and $\boldsymbol{B}$ is a low-frequency Fourier matrix, while L and μ_h^2 vary. Thanks to the properties of low-frequency Fourier matrices, we are able to construct a vector $\boldsymbol{h}$ whose associated incoherence parameter μ_h^2 in (13) is equal to a particular number. In particular, we choose $\boldsymbol{h}$ to be one of those vectors whose first $3, 6, \cdots, 27, 30$ entries are 1 and the others are zero. The advantage of those choices is that $\max_{1\leq l\leq L} L|\langle \boldsymbol{b}_l, \boldsymbol{h}\rangle|^2/\|\boldsymbol{h}\|^2$ will not change with L and can be computed explicitly. We can see in Figure 3 that the minimal L required for exact recovery seems strongly linearly associated with $\mu_h^2 = L \max |\langle \boldsymbol{b}_l, \boldsymbol{h}\rangle|^2/\|\boldsymbol{h}\|^2$.

10.3 Robustness

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 $\frac{\sqrt{\sum_{i=1}^r \|\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\|_F^2}}{\sqrt{\sum_{i=1}^r \|\boldsymbol{X}_i\|_F^2}}$ 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 $\boldsymbol{B}_i$ are as outlined in Section 10.1 and the $\boldsymbol{A}_i$ are Gaussian random matrices. In the simulation, we choose $\boldsymbol{\epsilon}_i$ to be a normalized Gaussian random vector. Namely, we first sample $\boldsymbol{\epsilon}_i$ from a multivariate Gaussian distribution and then normalize $\|\boldsymbol{\epsilon}_i\|_F = \sigma\sqrt{\sum_{i=1}^r \|\boldsymbol{X}_i\|_F^2}$ where $\sigma = 1, 0.5, 0.1, 0.05, 0.01, \cdots$ and 0.0001. For each σ, 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 are equal to 10, and all K_i are equal to 15) and the $\boldsymbol{A}_i$ are the "random" Hadamard matrices described above. For both experiments, Figure 4 indicates that the average relative error (dB) is linearly correlated with $\text{SNR} = 10\log_{10}(\sum_{i=1}^r \|\boldsymbol{X}_i\|_F^2/\|\boldsymbol{\epsilon}\|_F^2)$, as one would wish.

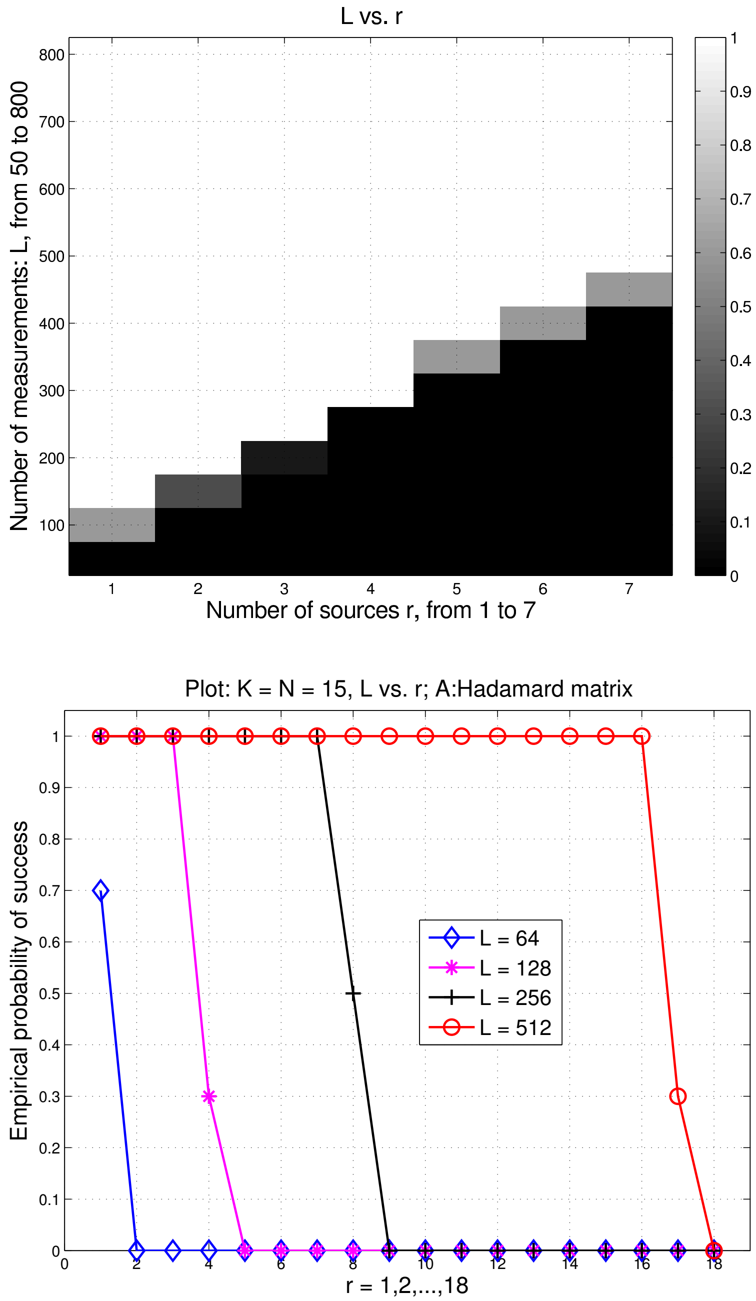

Figure 1: Phase transition plot: performance of (8) for different pairs of (L, r). White: 100% success and black: 0% success. Top: $\boldsymbol{A}_i : L \times N_i$ Gaussian random matrices. $K_i = 30$ and $N_i = 25$. $1 \le r \le 7$ and $L = 50, 100, \cdots, 800$; Bottom: $\boldsymbol{A}_i = \boldsymbol{D}_i \boldsymbol{H}_i$ where $\boldsymbol{A}$ is the first N_i columns of an $L \times L$ Hadamard matrix and $\boldsymbol{D}_i$ is a diagonal matrix with i.i.d. random entries taking ± 1 with equal probability. $K_i = N_i = 15$ with $r = 1, \cdots, 18$ and $L = 64, 128, 256, 512$.

11 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 area of multiuser wireless communications. Our theory provides a bound for the number of measurements needed to guarantee successful recovery. While this bound scales quadratically

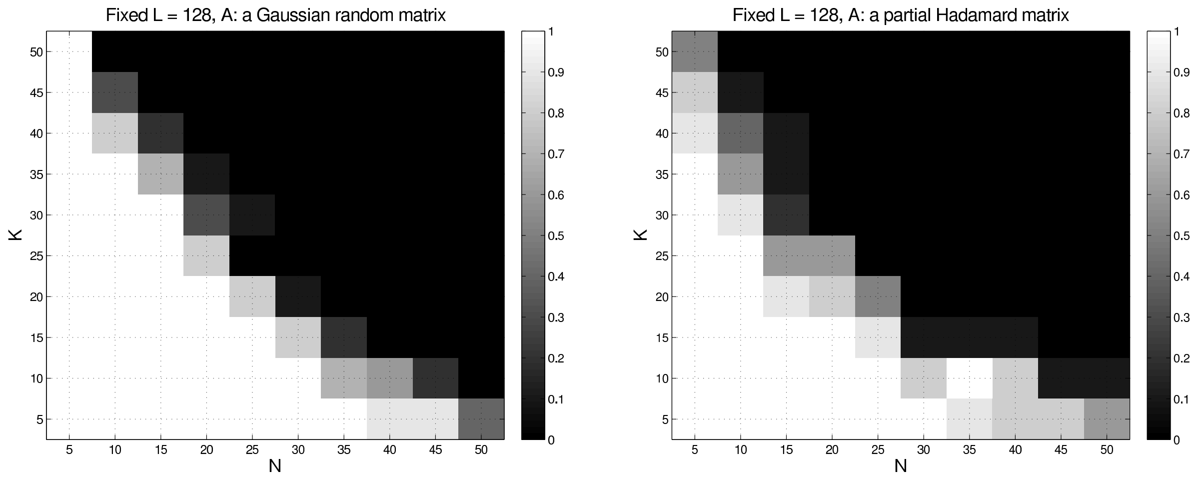

Figure 2: Phase transition plot: empirical probability of recovery success for (K_i, N_i) where K_i and N_i both vary from 5 to 50 and $L = 128$ is fixed. White: 100% success and black: 0% success. Left: each $\boldsymbol{A}_i$ is a $L \times N_i$ Gaussian random matrix; Right: $\boldsymbol{A}_i = \boldsymbol{D}_i \boldsymbol{H}_i$ with $\boldsymbol{H}_i$ being the first N_i columns of the $L \times L$ Hadamard matrix and $\boldsymbol{D}_i$ a diagonal matrix with entries taking value on ± 1 with equal probabilities.

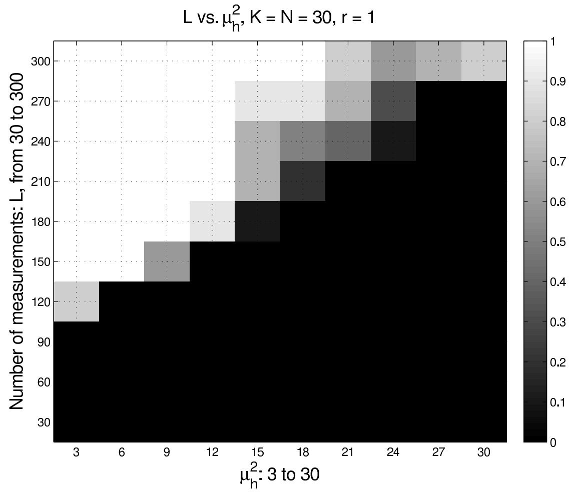

Figure 3: Phase transition plot: Empirical probability of recovery success for $(L, \max L|\langle \boldsymbol{b}_l, \boldsymbol{h}\rangle|^2/\|\boldsymbol{h}\|^2)$ where $r = 1$, $K = N = 30$. White: 100% success and black: 0% success.

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 $\boldsymbol{A}_i$? (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 $\boldsymbol{x}_i$ belong to some non-linear subspace,

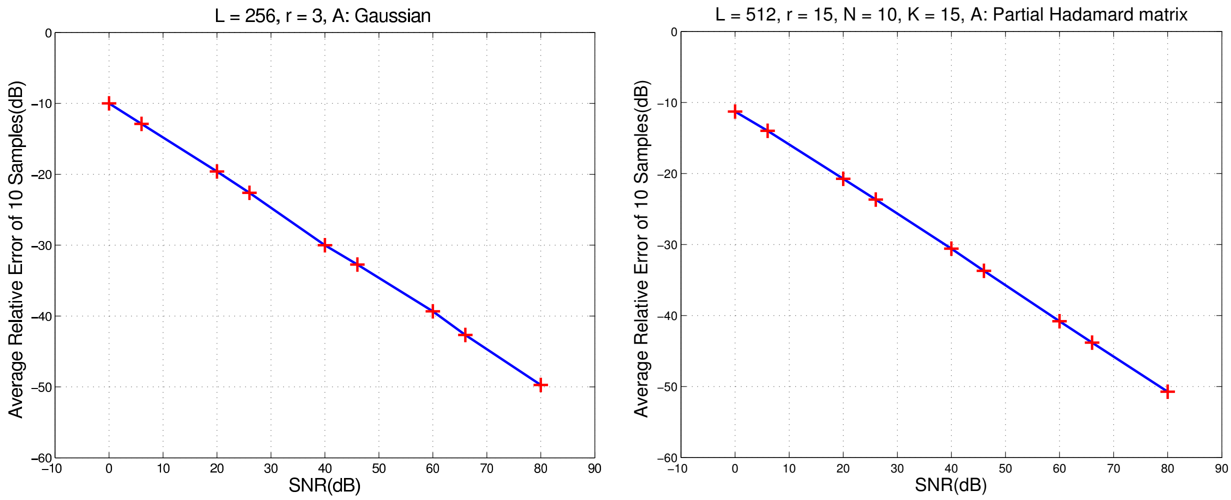

Figure 4: Performance of (21) under different SNR. Left: $\{\boldsymbol{A}_i\}$ are Gaussian and there are 3 sources and $L = 256$; Right: $\boldsymbol{A}_i = \boldsymbol{D}_i\boldsymbol{H}_i$ where $\boldsymbol{H}_i$ is a partial Hadamard matrix and $\boldsymbol{D}_i$ is a diagonal matrix with random ± 1 entries. Here there are 15 sources in total and $L = 512$.

e.g. for sparse $\boldsymbol{x}_i$? (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 Internet-of-Things.

Acknowledgement

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

12 Appendix

12.1 Useful lemmas

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 (43), i.e.,

$$c_1 \sup_{q\geq 1} q^{-1}(\mathbb{E}\,|Z|^q)^{1/q} \leq \|Z\|_{\psi_1} \leq c_2 \sup_{q\geq 1} q^{-1}(\mathbb{E}\,|Z|^q)^{1/q}, \tag{91}$$

where c_1 and c_2 are two universal positive constants, see Section 5.2.4 in [40]. Therefore, $\sup_{q\geq 1} q^{-1}(\mathbb{E}\,|Z|^q)^{1/q}$ will be used to quantify $\|Z\|_{\psi_1}$ in this section since it is easier to use in explicit calculations.

Lemma 12.1. *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 [38]. 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 $\boldsymbol{w}$ with an exponential tail

$$\|z + w\|_{\psi_1} \leq C(\|z\|_{\psi_1} + \|w\|_{\psi_1}) \tag{92}$$

for some universal contant C.

Proof: We only prove (92) by using the equivalent quantity introduced in (91).

$$\begin{aligned}\|z+w\|_{\psi_1} &\leq c_2 \sup_{q\geq 1} q^{-1}(\mathbb{E}\,|z+w|^q)^{1/q} \\ &\leq c_2 \sup_{q\geq 1} q^{-1}\left[(\mathbb{E}\,|z|)^{1/q} + (\mathbb{E}\,|w|)^{1/q}\right] \\ &\leq c_1 c_2(\|z\|_{\psi_1} + \|w\|_{\psi_1}),\end{aligned}$$

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

Lemma 12.2. *Let $\boldsymbol{u} \in \mathbb{R}^n \sim \mathcal{N}(0, \boldsymbol{I}_n)$, then $\|\boldsymbol{u}\|^2 \sim \chi_n^2$ and*

$$\|\|\boldsymbol{u}\|^2\|_{\psi_1} = \|\langle \boldsymbol{u}, \boldsymbol{u}\rangle\|_{\psi_1} \leq 2n. \tag{93}$$

Furthemore,

$$\mathbb{E}\left[(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I}_n)^2\right] = (n+1)\boldsymbol{I}_n. \tag{94}$$

Lemma 12.3 (Lemma 10-13 in [2])**.** *Let $\boldsymbol{u} \in \mathbb{R}^n \sim \mathcal{N}(0, \boldsymbol{I}_n)$ and $\boldsymbol{q} \in \mathbb{C}^n$ be any deterministic vector, then the following properties hold*

$$|\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2 \sim \|\boldsymbol{q}\|^2\chi_1^2, \tag{95}$$

$$\||\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2\|_{\psi_1} \leq C\|\boldsymbol{q}\|^2, \tag{96}$$

$$\||\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2 - \|\boldsymbol{q}\|^2\|_{\psi_1} \leq C\|\boldsymbol{q}\|^2, \tag{97}$$

$$\|(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I})\boldsymbol{q}\|_{\psi_1} \leq C\sqrt{n}\|\boldsymbol{q}\|, \tag{98}$$

$$\mathbb{E}\left[(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I})\boldsymbol{q}\boldsymbol{q}^*(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I})\right] = \|\boldsymbol{q}\|^2\boldsymbol{I} + \bar{\boldsymbol{q}}\bar{\boldsymbol{q}}^*. \tag{99}$$

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

$$\|\langle \boldsymbol{u}, \boldsymbol{q}\rangle\langle \boldsymbol{p}, \boldsymbol{u}\rangle - \langle \boldsymbol{q}, \boldsymbol{p}\rangle\|_{\psi_1} \leq \|\boldsymbol{q}\|\|\boldsymbol{p}\|. \tag{100}$$

Proof: (95) to (98) and (100) directly follow from Lemma 10-13 in [2], except for small differences in the constants. We only prove property (99)

$$\mathbb{E}\left[(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I})\boldsymbol{q}\boldsymbol{q}^*(\boldsymbol{u}\boldsymbol{u}^* - \boldsymbol{I})\right] = \mathbb{E}\left[|\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2\boldsymbol{u}\boldsymbol{u}^*\right] - \boldsymbol{q}\boldsymbol{q}^*.$$

For each (i,j)-th entry of $R_{ij} = |\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2 u_i u_j = \boldsymbol{q}^*\left[u_i u_j \boldsymbol{u}\boldsymbol{u}^*\right]\boldsymbol{q}$.

$$\mathbb{E}\left[u_i u_j \boldsymbol{u}\boldsymbol{u}^*\right] = \begin{cases} \boldsymbol{E}_{ij} + \boldsymbol{E}_{ji} & i \neq j \\ \boldsymbol{I} + \boldsymbol{E}_{ii} & i = j \end{cases}$$

where $\boldsymbol{E}_{ij}$ 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}

$$\mathbb{E}\,R_{ij} = \begin{cases} q_i^* q_j + q_j^* q_i & i \neq j \\ \|\boldsymbol{q}\|^2 + |q_i|^2 & i = j \end{cases}$$

and

$$\mathbb{E}\left[|\langle \boldsymbol{u}, \boldsymbol{q}\rangle|^2\boldsymbol{u}\boldsymbol{u}^*\right] - \boldsymbol{q}\boldsymbol{q}^* = \|\boldsymbol{q}\|^2\boldsymbol{I} + \boldsymbol{q}\boldsymbol{q}^* + \bar{\boldsymbol{q}}\bar{\boldsymbol{q}}^* - \boldsymbol{q}\boldsymbol{q}^* = \|\boldsymbol{q}\|^2\boldsymbol{I} + \bar{\boldsymbol{q}}\bar{\boldsymbol{q}}^*.$$

where $\bar{\boldsymbol{q}}$ is the complex conjugate of $\boldsymbol{q}$. ■

Lemma 12.4. *Assume $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I}_n)$ and $\boldsymbol{v} \sim \mathcal{N}(0, \boldsymbol{I}_m)$ are two independent Gaussian random vectors, then*

$$\left\|\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2\right\|_{\psi_1} \leq n + m$$

and

$$\left\|\|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|\right\|_{\psi_1} \leq C\sqrt{mn}.$$

Proof: Let us start with the first one.

$$\left\| \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 \right\|_{\psi_1} \leq \| \|\boldsymbol{u}\|^2 \|_{\psi_1} + \| \|\boldsymbol{u}\|^2 \|_{\psi_1} \leq n + m,$$

which directly follows from (92) and (93). Following from independence,

$$\| \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\| \|_{\psi_1} \leq c_2 \sup_{q \geq 1} q^{-1} (\mathbb{E}\, \|\boldsymbol{u}\|^q \|\boldsymbol{v}\|^q)^{1/q} \leq c_2 \sup_{q} q^{-1} (\mathbb{E}\, \|\boldsymbol{u}\|^q)^{1/q} (\mathbb{E}\, \|\boldsymbol{v}\|^q)^{1/q}.$$

Let $t = q/2$,

$$\begin{aligned} \| \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\| \|_{\psi_1} &\leq c_2 \sup_{t \geq 1} \frac{1}{2t} (\mathbb{E}\, \|\boldsymbol{u}\|^{2t})^{1/2t} (\mathbb{E}\, \|\boldsymbol{v}\|^{2t})^{1/2t} \\ &\leq \frac{c_2}{2} \left(\sup_{t \geq 1} \frac{1}{t} (\mathbb{E}\, \|\boldsymbol{u}\|^{2t})^{1/t} \right)^{1/2} \left(\sup_{t \geq 1} \frac{1}{t} (\mathbb{E}\, \|\boldsymbol{v}\|^{2t})^{1/t} \right)^{1/2} \\ &\leq \frac{c_1 c_2}{2} \sqrt{\|\boldsymbol{u}\|_{\psi_1} \cdot \|\boldsymbol{v}\|_{\psi_1}} \\ &\leq C\sqrt{mn}, \end{aligned}$$

where $\|\boldsymbol{u}\|^2 \sim \chi_n^2$ and $\|\boldsymbol{v}\|^2 \sim \chi_m^2$ and $\|\boldsymbol{u}\|_{\psi_1}$ and $\|\boldsymbol{v}\|_{\psi_1}$ are given by (93). ■

12.2 An Important fact about "low-frequency" DFT matrix

Suppose that $\boldsymbol{B}$ is a "low-frequency" Fourier matrix, i.e.,

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

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 $\Gamma_p = \{p, P + p, \cdots, (Q-1)P + p\}$ 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 $\boldsymbol{B}_p$ be the $Q \times K$ matrix by choosing its rows from those of $\boldsymbol{B}$ with indices in Γ_p. Then we can rewrite $\boldsymbol{B}_p$ as

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

and it actually equals

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

Therefore

$$\boldsymbol{B}_p = \sqrt{\frac{Q}{L}} \boldsymbol{F}_Q \operatorname{diag}(e^{2\pi i (P-p)/(PQ)}, \cdots, e^{2\pi i K (P-p)/(PQ)})$$

where $\boldsymbol{F}_Q$ is the first K columns of a $Q \times Q$ DFT matrix with $\boldsymbol{F}_Q^* \boldsymbol{F}_Q = \boldsymbol{I}_K$. Therefore

$$\sum_{l \in \Gamma_p} \boldsymbol{b}_l \boldsymbol{b}_l^* = \boldsymbol{B}_p^* \boldsymbol{B}_p = \frac{Q}{L} \boldsymbol{I}_K.$$

where $\boldsymbol{b}_l$ is the l-th column of $\boldsymbol{B}^*$.

References

[1] A. Ahmed, A. Cosse, and L. Demanet. A convex approach to blind deconvolution with diverse inputs. *Preprint*, 2015.

[2] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *Information Theory, IEEE Transactions on*, 60(3):1711–1732, 2014.

[3] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, page iau005, 2014.

[4] C. Bilen, G. Puy, R. Gribonval, and L. Daudet. Convex optimization approaches for blind sensor calibration using sparsity. *Signal Processing, IEEE Transactions on*, 62(18):4847–4856, 2014.

[5] E. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[6] E. J. Candes, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *Information Theory, IEEE Transactions on*, 61(4):1985–2007, 2015.

[7] E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[8] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[9] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[10] E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[11] E. Candès and Y. Plan. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

[12] J.-F. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon. Blind separation of noisy Gaussian stationary sources. application to cosmic microwave background imaging. *arXiv preprint astro-ph/0209466*, 2002.

[13] S. Chaudhuri, R. Velmurugan, and R. Rameshan. Blind deconvolution methods: A review. In *Blind Image Deconvolution*, pages 37–60. Springer, 2014.

[14] Y. Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *arXiv preprint arXiv:1506.02751*, 2015.

[15] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[16] B. Friedlander and A. J. Weiss. Self-calibration for high-resolution array processing. In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing, Vol. II*, chapter 10, pages 349–413. Prentice-Hall, 1991.

[17] D. Gesbert, B. C. Ng, and A. J. Paulraj. Blind space-time receivers for CDMA communications. In *CDMA Techniques for Third Generation Mobile Systems*, pages 285–302. Springer, 1999.

[18] M. Grant, S. Boyd, and Y. Ye. Cvx: Matlab software for disciplined convex programming, 2008.

[19] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.

[20] A. Hedayat, W. Wallis, et al. Hadamard matrices and their applications. *The Annals of Statistics*, 6(6):1184–1238, 1978.

[21] V. Koltchinskii et al. Von Neumann entropy penalization and low-rank matrix estimation. *The Annals of Statistics*, 39(6):2936–2973, 2011.

[22] V. Koltchinskii, K. Lounici, A. B. Tsybakov, et al. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011.

[23] X. Li and H. H. Fan. Direct blind multiuser detection for CDMA in multipath without channel estimation. *Signal Processing, IEEE Transactions on*, 49(1):63–73, 2001.

[24] Y. Li, K. Lee, and Y. Bresler. Identifiability in blind deconvolution with subspace or sparsity constraints. *arXiv preprint arXiv:1505.03399*, 2015.

[25] S. Ling and T. Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015.

[26] J. Liu, J. Xin, Y. Qi, F.-G. Zheng, et al. A time domain algorithm for blind separation of convolutive sound mixtures and L_1 constrained minimization of cross correlations. *Communications in Mathematical Sciences*, 7(1):109–128, 2009.

[27] M. B. McCoy, V. Cevher, Q. T. Dinh, A. Asaei, and L. Baldassarre. Convexity in source separation: Models, geometry, and algorithms. *Signal Processing Magazine, IEEE*, 31(3):87–95, 2014.

[28] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex demixing, with applications. *Foundations of Computational Mathematics*, 14(3):503–567, 2014.

[29] B. Recht. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 12:3413–3430, 2011.

[30] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[31] N. Shamir, Z. Zalevsky, L. Yaroslavsky, and B. Javidi. Blind source separation of images based on general cross correlation of linear operators. *Journal of Electronic Imaging*, 20(2):023017–023017, 2011.

[32] C. Shin, R. W. Heath Jr, and E. J. Powers. Blind channel estimation for MIMO-OFDM systems. *Vehicular Technology, IEEE Transactions on*, 56(2):670–685, 2007.

[33] S. Shwartz and M. Zibulevsky. *Efficient blind separation of convolutive image mixtures.* Technion-IIT, Department of Electrical Engineering, 2005.

[34] G. W. Stewart. Perturbation theory for the singular value decomposition. *technical report CS-TR 2539, university of Maryland*, September 1990.

[35] P. Sudhakar, S. Arberet, and R. Gribonval. Double sparsity: Towards blind estimation of multiple channels. In *Latent Variable Analysis and Signal Separation*, pages 571–578. Springer, 2010.

[36] M. N. Syed, P. G. Georgiev, and P. M. Pardalos. Blind Signal Separation Methods in Computational Neuroscience. In *Modern Electroencephalographic Assessment Techniques*, pages 291–322. Springer, 2015.

[37] I. Toumi, S. Caldarelli, and B. Torrésani. A review of blind source separation in NMR spectroscopy. *Progress in nuclear magnetic resonance spectroscopy*, 81:37–64, 2014.

[38] A. van der Vaart and J. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[39] S. Verdu. *Multiuser Detection*. Cambridge University Press, 1998.

[40] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2010. To Appear. Preprint available at `http://www-personal.umich.edu/~romanv/papers/papers.html`.

[41] X. Wang and H. V. Poor. Blind equalization and multiuser detection in dispersive cdma channels. *Communications, IEEE Transactions on*, 46(1):91–103, 1998.

[42] P.-Å. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.

[43] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Information and Inference*, 2(1):32–68, 2013.

[44] G. Wunder, H. Boche, T. Strohmer, and P. Jung. Sparse signal processing concepts for efficient 5g system design. *IEEE Access*, accepted, 2014. Preprint available at arXiv.org/abs/1411.0435.