

Fast and Guaranteed Blind Multichannel Deconvolution Under a Bilinear System Model

Kiryung Lee[✉], *Member, IEEE*, Ning Tian, and Justin Romberg, *Fellow, IEEE*

Abstract—We consider the multichannel blind deconvolution problem where we observe the output of multiple channels that are all excited with the same unknown input. From these observations, we wish to estimate the impulse responses of each of the channels. We show that this problem is well-posed if the channels follow a bilinear model where the ensemble of channel responses is modeled as lying in a low-dimensional subspace but with each channel modulated by an independent gain. Under this model, we show how the channel estimates can be found by minimizing a quadratic function over a non-convex set. We analyze two methods for solving this non-convex program, and provide performance guarantees for each. The first is a method of alternating eigenvectors that breaks the program down into a series of eigenvalue problems. The second is a truncated power iteration, which can roughly be interpreted as a method for finding the largest eigenvector of a symmetric matrix with the additional constraint that it adheres to our bilinear model. As with most non-convex optimization algorithms, the performance of both of these algorithms is highly dependent on having a good starting point. We show how such a starting point can be constructed from the channel measurements. Our performance guarantees are non-asymptotic, and provide a sufficient condition on the number of samples observed per channel in order to guarantee channel estimates of certain accuracy. Our analysis uses a model with a “generic” subspace that is drawn at random, and we show the performance bounds hold with high probability. Mathematically, the key estimates are derived by quantifying how well the eigenvectors of certain random matrices approximate the eigenvectors of their mean. We also present a series of numerical results demonstrating that the empirical performance is consistent with the presented theory.

Index Terms—Blind deconvolution, non-convex optimization, eigenvalue decomposition, sensitivity analysis.

I. INTRODUCTION

BLIND deconvolution, where we estimate two unknown signals from an observation of their convolution, is a classical problem in signal processing. It is ubiquitous, appearing in applications including channel estimation in communications, image deblurring and restoration, seismic data analysis,

speech dereverberation, sensor calibration in medical imaging, and convolutive dictionary learning. While algorithms based on heuristics for particular applications have existed for decades, it is not until recently that a rich mathematical theory has developed around this problem. The fundamental identifiability of solutions to this problem has been studied from an information theoretic perspective [3]–[9]. Practical algorithms with provable performance guarantees that make the problem well-posed by imposing structural constraints on the signals have arisen based on ideas from compressed sensing and low-rank matrix recovery. These include methods based on convex programming [10]–[12], alternating minimization [13], and gradient descent [14]. More recent works studied the more challenging problem of blind deconvolution with off-the-grid sparsity models [15], [16].

In this paper, we consider the multichannel blind deconvolution problem: we observe a single unknown signal (the “source”) convolved with a number of different “channels”. The fact that the input is shared makes this problem better-posed than in the single channel case. Mathematical theory for the multichannel problem under various constraints has existed since the 1990s (see [17], [18] for surveys). One particular strand of this research detailed in [19]–[21] gives concrete results under the very loose assumption that the channel responses are time-limited. These works show how with this model in place, the channel responses can be estimated by forming a cross-correlation matrix from the channel outputs and then computing its smallest eigenvector. This estimate is consistent in that it is guaranteed to converge to the true channel responses as the number of observations gets infinitely large. However, no performance guarantees were given for a finite number of samples, and the method tends to be unstable for moderate sample sizes in even modest noise. Recent work [22] has shown that this spectral method can indeed be stabilized by introducing a more restrictive linear (subspace) model on the channel responses.

Our main contributions in this paper are methods for estimating the channel responses when the ensemble has a certain kind of bilinear structure. In particular, we model the ensemble of channel responses as lying in a low-dimensional subspace, but with each channel modulated by an independent constant; we will discuss in the next section an application in which this model is relevant. Our estimation framework again centers on constructing a cross-correlation matrix and minimizing a quadratic function involving this matrix over the unit sphere, but with the additional constraint that the solution

Manuscript received October 19, 2016; revised March 9, 2018; accepted March 19, 2018. Date of publication May 25, 2018; date of current version June 20, 2018. This work was supported in part by NSF under Grants CCF 14-22540, IIS 14-47879, and CCF 17-18771, and in part by ONR under Grant N00014-13-1-0632. This paper was presented in part at the 2016 IEEE Information Theory Workshop [1] and in part by the 22nd International Conference on Digital Signal Processing [2].

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: kiryung@ece.gatech.edu; ningtian@gatech.edu; jrom@ece.gatech.edu).

Communicated by H. Rauhut, Associate Editor for Signal Processing.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2018.2840711

can be written as the Kronecker product of two shorter vectors. This optimization program, which might be interpreted as a kind of structured eigenvalue problem, is inherently non-convex. We propose two iterative methods for solving it, each with very simple, computationally efficient iterations. The first is a method of alternating eigenvectors, where we alternate between fixing a subset of the unknowns and estimating the other by solving a standard eigenvalue problem. The second method is a truncated power iteration, where we repeatedly apply the cross-correlation matrix to an initial point, but project the result after each application to enforce the structural constraints. We derive performance guarantees for both of these algorithms when the low-dimensional subspace is generic (i.e. generated at random).

A. Related Work

Closely related to the problem of multichannel blind deconvolution is the problem of blind calibration. Here we observe the product of an unknown weighting vector applied to a series of other unknown vectors. Non-convex optimization algorithms for blind calibration have been studied and analyzed in [23] and [24].

Multichannel blind deconvolution can also be approached by linearizing the problem in the Fourier domain. This has been proposed for various applications, including the calibration of a sensor network [25], computational relighting in inverse rendering [26], and auto-focus in synthetic aperture radar [27]. Under a generic condition that the unknown impulse responses belong to random subspaces, necessary and sufficient conditions for the unique identification of the solution have been put forth in [6], and a rigorous analysis of a least-squares method has been studied [28].

More recently, performance guarantees for spectral methods for both subspace and sparsity models have been developed in [29]. As in this paper, these methods are estimating the channel by solving a structured eigenvalue problem. The structural model, however, is very different than the one considered here.

Algorithms for solving non-convex quadratic and bilinear problems have recently been introduced for solving problems closely related to blind deconvolution. In [30], it is shown that a non-convex optimization over matrix manifolds provides a guaranteed solution for matrix completion [30]. Alternating minimization, another non-convex optimization algorithm for matrix completion that provides a provable performance guarantee, was analyzed in [31]–[33]. A different suite of gradient-based algorithms with a specially designed regularizer within the conventional Euclidean geometry have also been studied recently [34]. Wirtinger flow [35]–[38] and alternating minimization [39], [40] are non-convex optimization algorithms for the phase retrieval problem. Alternating minimization has also been recently analyzed for the closely related problem of blind Ptychographic diffraction imaging [41]. Dictionary learning is another bilinear problem arising in numerous applications; convergence of a Riemannian trust-region method for this problem has been studied with a thorough geometric analysis in [42] and [43].

B. Organization

The rest of the paper is organized as follows. The multichannel blind deconvolution problem is formulated under a bilinear channel model in Section II. After we review relevant previous methods for multichannel blind deconvolution in Section III, we present two iterative algorithms for multichannel blind deconvolution under the bilinear channel model in Section IV, which are obtained by modifying the classical cross-convolution method. Our main results on non-asymptotic stable recovery are presented in Section V with an outline of the proofs. Detailed analysis of the spectral initialization and the two iterative algorithms are derived in Sections VII, VIII, and IX. We demonstrate numerical results that support our theory in Section VI, and summarize in Section X.

II. PROBLEM STATEMENT

In the classic multichannel blind deconvolution problem, we observe an unknown signal $\mathbf{x} \in \mathbb{C}^L$ that has been convolved with M different unknown channel responses $\mathbf{h}_1, \dots, \mathbf{h}_M \in \mathbb{C}^L$:

$$\mathbf{y}_m = \mathbf{h}_m \circledast \mathbf{x} + \mathbf{w}_m, \quad m = 1, \dots, M, \quad (1)$$

where \circledast denotes circular convolution¹ modulo L and $\mathbf{w}_m \in \mathbb{C}^L$ is additive noise. Given the outputs $\{\mathbf{y}_m\}_{m=1}^M$, and working without knowledge of the common input \mathbf{x} , we want to recover the unknown channel impulse responses $\{\mathbf{h}_m\}_{m=1}^M$.

We will show how we can solve this problem when the channels are time-limited and obey a *bilinear model*. By time-limited, we mean that only the first K entries in the \mathbf{h}_m can be non-zero; we can write

$$\mathbf{h}_m = \mathbf{S}^\top \underline{\mathbf{h}}_m, \quad \text{where } \mathbf{S} := [\mathbf{I}_K \quad \mathbf{0}_{K, L-K}]. \quad (2)$$

In addition, the shapes of $\mathbf{h}_1, \dots, \mathbf{h}_M$ are *jointly* modeled as lying in a D -dimensional subspace of \mathbb{C}^K , but are multiplied by unknown channel gains $a_1, \dots, a_M > 0$. This means that

$$\underline{\mathbf{h}}_m = a_m \Phi_m \mathbf{b}, \quad \forall m = 1, \dots, M, \quad (3)$$

where Φ_1, \dots, Φ_M are complex $K \times D$ matrices, whose columns are the parts of the basis vectors corresponding to channel m , and $\mathbf{b} \in \mathbb{C}^D$ is the common set of basis coefficients. Stacking up the channel responses into a single vector $\underline{\mathbf{h}} \in \mathbb{C}^{MK}$ and the gains into $\mathbf{a} \in \mathbb{C}^M$, an equivalent way to write (3) is

$$\underline{\mathbf{h}} = \Phi(\mathbf{a} \otimes \mathbf{b}), \quad (4)$$

where

$$\Phi := \begin{bmatrix} \Phi_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Phi_M \end{bmatrix}$$

and

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ \vdots \\ a_M \mathbf{b} \end{bmatrix}.$$

¹We are using circular convolution in our model problem for the ease of analysis.

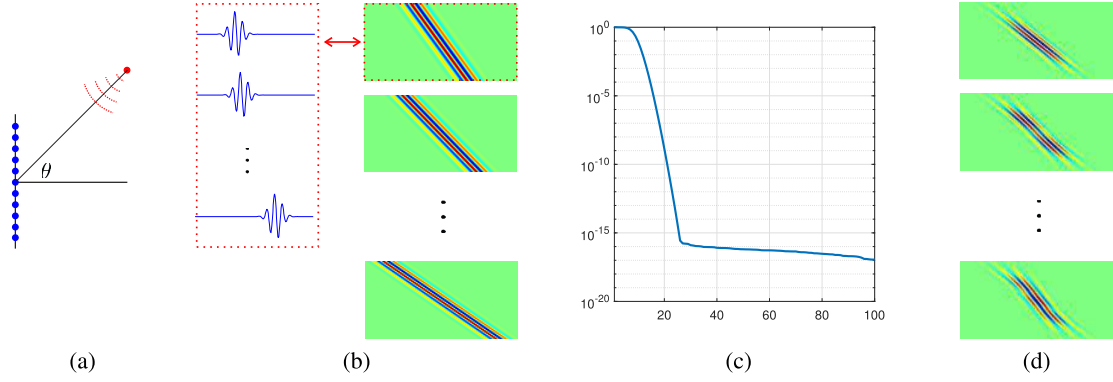


Fig. 1. Illustration of construction of a joint linear subspace model from a parametric model. (a) A uniform array of M sensors. (b) Examples of $\{g_{\vec{r}}\}$ rearranged as M -by- K matrices. (c) Sorted eigenvalues of H_R in a logarithmic scale. (d) First few dominant eigenvectors of H_R rearranged as M -by- K matrices.

This alternative expression can be interpreted as a linear subspace model with respect to the basis $\Phi \in \mathbb{C}^{MK \times MD}$ with a separability (rank-1) prior on the coefficient vector.

For an example of how a model like this might arise, we consider the following stylized problem for array processing illustrated in Figure 1. Figure 1(a) shows a linear array. Suppose we know that if a source is at location \vec{r} then the concatenation of the channel responses between the source location and the array elements is $g_{\vec{r}} \in \mathbb{C}^{MK}$. In simple environments, these channel responses might look very similar to one another in that they are all (sampled) versions of the same shifted function (see Figure 1(b)). The delays are induced by the differences in sensor locations relative to the source, while the shape of the response might be determined by the instrumentation used to take the measurements (e.g. the frequency response of the sensors) — there could even be small differences in this shape from element to element.

Suppose now that there is uncertainty in the source location that we model as $\vec{r} \in \mathcal{R}$, where \mathcal{R} is some region in space. As we vary \vec{r} over the set \mathcal{R} , the responses $g_{\vec{r}}$ trace out a portion of a manifold in \mathbb{C}^{MK} . We can (approximately) embed this manifold in a linear subspace of dimension D by looking at the D principal eigenvectors of the matrix

$$H_R = \int_{\vec{r} \in \mathcal{R}} g_{\vec{r}} g_{\vec{r}}^* d\vec{r}.$$

The dimension D that allows an accurate embedding will depend on the size of \mathcal{R} and smoothness properties of the mapping from \vec{r} to $g_{\vec{r}}$. In this case, we are building Φ above by taking the $MK \times D$ matrix that has the principal eigenvectors as columns and apportioning the first K rows to Φ_1 , the next K rows to Φ_2 , etc.

This technique of embedding a parametric model into a linear space has been explored for source localization and channel estimation in underwater acoustics in [44] and [45], and some analysis in the context of compressed sensing is provided in [46]. However, it is not robust in one important way. In practice, the gains (the amplitude of the channel response) can vary between elements in the array, and this variation is enough to compromise the subspace embedding

described above. The bilinear model (4) explicitly accounts for these channel-to-channel variations.

In this paper, we are interested in when equations of the form (1) can be solved for h_m with the structural constraint (4); we present two different algorithms for doing so in the sections below. The effectiveness of these algorithms will of course be affected by properties of Φ (including the number of channels M and embedding dimension D) as well as the number of samples L . While empirical models like the one described above are used in practice (see in particular [45]), we will analyze generic instances of this problem, where the linear model is drawn at random.

III. SPECTRAL METHODS FOR MULTICHANNEL BLIND DECONVOLUTION

A classical method for treating the multichannel blind deconvolution problem is to recast it as an eigenvalue problem: we create a correlation matrix using the measured data $\{y_m\}$, and estimate the channels from the smallest eigenvector² of this matrix. These methods were pioneered in the mid-1990s in [19]–[21], and we briefly review the central ideas in this section. The methods we present in the next section operate on the same basic principles, but explicitly enforce structural constraints on the solution.

The cross-convolution method for multichannel blind deconvolution [19] follows directly from the commutativity of the convolution operator. If there is no noise in the observations (1), then it must be the case that

$$y_{m_1} \circledast h_{m_2} - y_{m_2} \circledast h_{m_1} = 0$$

for all $m_1, m_2 = 1, \dots, M$. Using T_{y_m} as the matrix whose action is convolution with y_m with a signal of length K , we see that the channel responses h_{m_1} and h_{m_2} must obey the linear constraints $T_{y_{m_1}} h_{m_2} - T_{y_{m_2}} h_{m_1} = 0$. We can collect all pairs of these linear constraints into a large system, expressed as

$$Y \underline{h} = 0_{M(M-1)L/2, 1} \quad (5)$$

²By which mean the eigenvector corresponding to the smallest eigenvalue.

with $\mathbf{Y} \in \mathbb{C}^{M(M-1)L/2 \times MK}$ defined by

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \\ \vdots \\ \mathbf{Y}^{(M-1)} \end{bmatrix}, \quad (6)$$

where

$$\mathbf{Y}^{(i)} = \begin{bmatrix} \mathbf{0}_{L,K} & \dots & \mathbf{0}_{L,K} & \mathbf{T}_{y_{i+1}} & -\mathbf{T}_{y_i} & & \\ \vdots & & \vdots & \vdots & & \ddots & \\ \mathbf{0}_{L,K} & \dots & \mathbf{0}_{L,K} & \mathbf{T}_{y_M} & & & -\mathbf{T}_{y_i} \end{bmatrix}$$

(i-1) blocks
(M-i+1) blocks

for $i = 1, \dots, M-1$.

It is shown in [19] and [21] that $\underline{\mathbf{h}}$ is uniquely determined up to a scaling by (5) (i.e. \mathbf{Y} has a null space that is exactly 1-dimensional) under the mild algebraic condition that the polynomials generated by the $(\underline{\mathbf{h}}_m)_{m=1}^M$ have no common zeros. In the presence of noise, $\underline{\mathbf{h}}$ is estimated as the minimum eigenvector of $\mathbf{Y}^* \mathbf{Y}$:

$$\hat{\underline{\mathbf{h}}} = \underset{\|\mathbf{g}\|_2=1}{\operatorname{argmin}} \mathbf{g}^* \mathbf{Y}^* \mathbf{Y} \mathbf{g}. \quad (7)$$

Note that $\mathbf{Y}^* \mathbf{Y}$ is computed by cross-correlating the outputs. Therefore, $\mathbf{Y}^* \mathbf{Y}$ is computed at a low computational cost using the fast Fourier transform. Furthermore, the size of $\mathbf{Y}^* \mathbf{Y}$, which is $MK \times MK$, does not grow as the length L of the observations increases. When there is additive white noise, this cross-correlation matrix will in expectation be the noise-free version plus a scaled identity. This means that as the sample size gets large, the noise and noise-free cross-correlation matrices will have the same eigenvectors, and so the estimate (7) is consistent.

A similar technique can be used if we have a linear model for the channel responses, $\underline{\mathbf{h}} = \Phi \mathbf{u}$. We can estimate the expansion coefficients \mathbf{u} by solving

$$\begin{aligned} & \underset{\mathbf{v}}{\operatorname{minimize}} \quad \mathbf{v}^* \Phi^* (\mathbf{Y}^* \mathbf{Y} - \varrho \mathbf{I}) \Phi \mathbf{v} \\ & \text{subject to } \|\mathbf{v}\|_2 = 1, \end{aligned} \quad (8)$$

where ϱ is a scalar that depends on the variance of the additive noise (this correction is made so that eigen-structure more closely matches that of $\Phi^* \mathbf{Y}^* \mathbf{Y} \Phi$ for noise-free \mathbf{Y}). In [22], it was shown that a linear model can significantly improve the stability of the estimate of $\underline{\mathbf{h}}$ in the presence of noise, and gave a rigorous non-asymptotic analysis of the estimation error for generic bases Φ .

IV. NON-CONVEX OPTIMIZATION ALGORITHMS

Our proposed framework is to solve an optimization program similar to (7) and (8) above, but with the additional constraint that $\underline{\mathbf{h}}$ obey the bilinear form (4).

Given the noisy measurements $\{\mathbf{y}_m\}$ in (1), we create the matrix

$$\mathbf{A} = \Phi^* (\mathbf{Y}^* \mathbf{Y} - \hat{\sigma}_w^2 (M-1) L \mathbf{I}_{MK}) \Phi,$$

where $\hat{\sigma}_w^2$ is an estimate of the noise variance σ_w^2 (we will briefly discuss how to estimate the noise variance later in

this section), and \mathbf{Y} is formed as in (6). We then solve a program that is similar to the eigenvalue problems above, but with a Kronecker product constraint on the expansion coefficients:

$$\begin{aligned} & \underset{\mathbf{v}, \mathbf{c}, \mathbf{d}}{\operatorname{minimize}} \quad \mathbf{v}^* \mathbf{A} \mathbf{v} \\ & \text{subject to } \|\mathbf{v}\|_2 = 1, \quad \mathbf{v} = \mathbf{c} \otimes \mathbf{d}. \end{aligned} \quad (9)$$

The norm and bilinear constraints make this a non-convex optimization program, and unlike the spectral methods discussed in the last section, there is no (known) computationally efficient algorithm to compute its solution.

We propose and analyze two non-convex optimization algorithms below for solving (9). The first is an alternating eigenvalue method, which iterates between minimizing for \mathbf{c} in (9) with \mathbf{d} fixed, then minimizing for \mathbf{d} with \mathbf{c} fixed. The second is a variation on the truncated power method [47], whose iterations consist of applications of the matrix \mathbf{A} followed by a projection to enforce the structural constraints.

The performance of both of these methods relies critically on constructing a suitable starting point. We discuss one method for doing so below, then establish its efficacy in Proposition 7 in Section V-B below.

A. Alternating Eigenvectors

While program (9) is non-convex, it becomes tractable if one of the terms in the tensor constraint is held constant. If we have an estimate $\hat{\mathbf{b}}$ for \mathbf{b} , and fix $\mathbf{d} = \hat{\mathbf{b}}$, then we can solve for \mathbf{c} using

$$\underset{\mathbf{c}}{\operatorname{minimize}} \quad \mathbf{c}^* \mathbf{A}_{\hat{\mathbf{b}}} \mathbf{c} \quad \text{subject to } \|\mathbf{c}\|_2 = 1,$$

where

$$\mathbf{A}_{\hat{\mathbf{b}}} = (\mathbf{I}_M \otimes \hat{\mathbf{b}})^* \mathbf{A} (\mathbf{I}_M \otimes \hat{\mathbf{b}})$$

and

$$\mathbf{I}_M \otimes \hat{\mathbf{b}} = \begin{bmatrix} \hat{\mathbf{b}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{b}} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \hat{\mathbf{b}} \end{bmatrix}.$$

The solution is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}_{\hat{\mathbf{b}}}$. Similarly, with an estimate $\hat{\mathbf{a}} = [\hat{a}_1, \dots, \hat{a}_M]^T$ plugged in for \mathbf{c} , we solve

$$\underset{\mathbf{d}}{\operatorname{minimize}} \quad \mathbf{d}^* \mathbf{A}_{\hat{\mathbf{a}}} \mathbf{d} \quad \text{subject to } \|\mathbf{d}\|_2 = 1,$$

where

$$\mathbf{A}_{\hat{\mathbf{a}}} = (\hat{\mathbf{a}} \otimes \mathbf{I}_D)^* \mathbf{A} (\hat{\mathbf{a}} \otimes \mathbf{I}_D), \quad \hat{\mathbf{a}} \otimes \mathbf{I}_D = \begin{bmatrix} \hat{a}_1 \mathbf{I} \\ \hat{a}_2 \mathbf{I} \\ \vdots \\ \hat{a}_M \mathbf{I} \end{bmatrix},$$

which is again given by the smallest eigenvector of $\mathbf{A}_{\hat{\mathbf{a}}}$.

We summarize this method of “alternating eigenvectors” in Algorithm 1. The function `MinEigVec` returns the eigenvector of the input matrix corresponding to its smallest eigenvalue.

Algorithm 1 Alternating Eigenvectors

input : A, b_0
output: \hat{h}
1 $\hat{b} \leftarrow b_0$
2 **while** *stop condition not satisfied* **do**
3 $\hat{a} \leftarrow \text{MinEigVec}((\mathbf{I}_M \otimes \hat{b}^*)A(\mathbf{I}_M \otimes \hat{b}))$
4 $\hat{b} \leftarrow \text{MinEigVec}((\hat{a}^* \otimes \mathbf{I}_D)A(\hat{a} \otimes \mathbf{I}_D))$
5 **end**
6 $\hat{h} \leftarrow \Phi(\hat{a} \otimes \hat{b})$

B. Rank-1 Truncated Power Method

A standard tool from numerical linear algebra to compute the largest eigenvector of a symmetric matrix is the *power method*, where the matrix is iteratively applied to a starting vector, with a renormalization at each step. (The same method can be used to compute the smallest eigenvector simply by subtracting the matrix from an appropriate scalar multiple of the identity.) In [47], a variation on this algorithm was introduced that forced the iterates to be sparse. This was done simply by hard thresholding after each application of the matrix.

Our rank-1 truncated power method follows the same template. We create a matrix B by subtracting A above from a multiple of the identity,

$$B = \gamma \mathbf{I}_{MD} - A,$$

then iteratively apply B starting with an initial vector u_0 . After each application of B , we project the result onto the set of rank-1 matrices by computing the singular vector corresponding to the largest singular value, and then renormalize.

We summarize the *rank-1 truncated power method* in Algorithm 2. Some care must be taken in choosing the value of γ . We want to ensure that the smallest eigenvalue of A gets mapped to the largest (in magnitude) eigenvalue of B , but we also want the relative gap between the largest and second largest eigenvalues of B to be as large as possible. In our analysis below, we use the conservative value of $\gamma = E[\|A\|]$. We also used this in the numerical results in Section VI. Alternatively, one could estimate the largest rank-1 constrained “eigenvalue” by applying Algorithm 2 to A itself, which may accelerate the convergence.

Algorithm 2 Rank-1 Truncated Power Method

input : B, v_0
output: v_t , a vectorized rank-1 matrix whose factors are the estimates \hat{a}, \hat{b}
1 $t \leftarrow 1$
2 **while** *stop condition not satisfied* **do**
3 $\tilde{v}_t \leftarrow B v_{t-1}$
4 $\hat{V}_t \leftarrow \text{Rank1Approx}(\text{mat}(\tilde{v}_t))$
5 $v_t \leftarrow \text{vec}(\hat{V}_t) / \|\text{vec}(\hat{V}_t)\|_2$
6 $t \leftarrow t + 1$
7 **end**

C. Spectral Initialization

Both the alternating eigenvectors and rank-1 truncated power method require an initial estimate of the channel gains a and the basis coefficients b . Because the program they are trying to solve is non-convex, this starting point must be chosen carefully.

Our spectral initialization is inspired from the lifting reformulation (e.g., see [10] for the lifting in blind deconvolution). The observation equations (1) can be recast as a linear operator acting on the Kronecker product of the unknowns x, b, a . Let $\mathcal{A} : \mathbb{C}^{LDM} \rightarrow \mathbb{C}^{ML}$ be a linear map such that³

$$\mathcal{A}(x \otimes b \otimes a) = \begin{bmatrix} x \otimes a_1 S^* \Phi_1 b \\ \vdots \\ x \otimes a_M S^* \Phi_M b \end{bmatrix}. \quad (10)$$

Concatenating the $\{y_m\}$ and $\{w_m\}$ into vectors y and w of length ML , we can rewrite (1) as

$$y = \mathcal{A}(\mathcal{X}) + w,$$

where $\mathcal{X} = x \otimes b \otimes a$.

A natural initialization scheme is to apply the adjoint of \mathcal{A} to y , then project the result onto the feasible set of vectors that can be arranged as rank-1 tensors (this technique is often used to initialize non-convex programs for recovering rank-1 matrices from linear measurements [48], [49]). However, there is no known algorithm for computing the projection onto the set of rank-1 tensors that has strong optimality guarantees.

We avoid this by exploiting the positivity of the multipliers, $a_m \geq 0$. The action of the operator $(\mathbf{I}_{LD} \otimes \mathbf{1}_{1,M})$ has the effect of summing down the third mode of the tensor:

$$(\mathbf{I}_{LD} \otimes \mathbf{1}_{1,M})(x \otimes b \otimes a) = \left(\sum_{m=1}^M a_m \right) (x \otimes b).$$

Since they are positive, a_1, \dots, a_M sum constructively above, and we can get an estimate of $x \otimes b$ by applying this operator to $\mathcal{A}^* y$. The positivity constraint on a can be weakened if estimates of the phases of a_1, \dots, a_M are available as prior information. In this scenario, the known phase information is absorbed into the basis Φ and one can focus on estimating only the gains.

The first step of our initialization, then, is to compute

$$\Gamma = \text{mat}((\mathbf{I}_{LD} \otimes \mathbf{1}_{1,M})\mathcal{A}^* y), \quad (11)$$

where the operator $\text{mat}(\cdot)$ takes a vector in \mathbb{C}^{LD} and produces a $D \times L$ matrix by column-major ordering.

Once corrected for noise, the leading eigenvector of $\Gamma \Gamma^*$ gives us a rough estimate of the channel coefficients b . In Section VII, we show that the random matrix $\Gamma \Gamma^* - \sigma_w^2 L \sum_{m=1}^M \Phi_m^* \Phi_m$ concentrates around a scalar multiple of bb^* .

Finally, we note that there is a closed-form expression for computing Γ from the measurements $\{y_m\}$. This is given in the following lemma that is proved in Appendix X.

³We have defined how \mathcal{A} operates on length LDM vectors that can be arranged as rank-1 tensors. Its action on a general vector in \mathbb{C}^{LDM} can be derived by applying the expression in (10) to a series of LDM vectors that form a separable basis for tensors in $\mathbb{C}^L \times \mathbb{C}^D \times \mathbb{C}^M$.

Lemma 1: The matrix Γ in (11) can be written as

$$\Gamma = \sum_{m=1}^M \Phi_m^* S C_{y_m} J, \quad (12)$$

where $C_{y_m} \in \mathbb{C}^{L \times L}$ is the matrix whose action is the circular convolution with $y_m \in \mathbb{C}^L$, J is the “flip operator” modulo L :

$$J := [e_1 \quad e_L \quad e_{L-1} \quad \cdots \quad e_2], \quad (13)$$

and e_1, \dots, e_L are the standard basis vectors for \mathbb{R}^L .

We summarize our spectral initialization technique in Algorithm 3.

Algorithm 3 Spectral Initialization

input : $\{y_m\}_{m=1}^M$, $\{\Phi_m\}_{m=1}^M$, L , $\hat{\sigma}_w^2$

output: b_0

1 $\Gamma \leftarrow \sum_{m=1}^M \Phi_m^* S C_{y_m} J$

2 $b_0 \leftarrow \text{MaxEigVec}(\Gamma \Gamma^* - \hat{\sigma}_w^2 L \sum_{m=1}^M \Phi_m^* \Phi_m)$

In our analysis of the initialization, we assume that we know the noise variance σ_w^2 . In practice, having an accurate estimate can indeed make a difference in terms of numerical performance. In the numerical experiments in Section VI, we include simulations where we assume we know the noise variance exactly, and where we take the crude guess $\hat{\sigma}_w^2 = 0$. The latter of course does not perform as well as the former, but it still offers significant gains over disregarding the bilinear structure all together.

It is also possible to get an estimate of the noise variance through the low-rank matrix denoising technique described in [50], where we solve the convex program

$$\underset{X, \alpha}{\text{minimize}} \quad \|\Gamma \Gamma^* - \alpha \sum_{m=1}^M \Phi_m^* \Phi_m - X\|_F^2 + \lambda \|X\|_*,$$

and take $\hat{\sigma}_w^2 = \hat{\alpha}/L$. The theory developed in [50] for this procedure relies on the perturbation to the low-rank matrix being subgaussian, which unfortunately does not apply here, as the perturbation involves both intra- and inter-channel convolutions of the noise processes $\{w_m\}$.

V. MAIN RESULTS

A. Non-Asymptotic Analysis

Our main results give non-asymptotic performance guarantees for both Algorithm 1 and Algorithm 2 when their iterations start from the initial estimate by Algorithm 3 under the following two assumptions⁴:

- (A1) **Generic subspaces.** The random matrices Φ_1, \dots, Φ_M are independent copies of a K -by- D complex Gaussian matrix whose entries are independent and identically

distributed (iid) as $\mathcal{CN}(0, 1)$. Our theorems below hold with high probability with respect to $(\Phi_m)_{m=1}^M$.

- (A2) **Random noise.** The perturbations to the measurements $w_1, \dots, w_M \in \mathbb{C}^L$ are independent subgaussian vectors with $\mathbb{E}[w_m] = \mathbf{0}$ and $\mathbb{E}[w_m w_m^*] = \sigma_w^2 \mathbf{I}_L$, and are independent of the bases $(\Phi_m)_{m=1}^M$.

We present main theorems in two different scenarios. In the first, we assume that the input source is a white subgaussian random process. In the second scenario, we assume that the input source satisfies a kind of incoherence condition that essentially ensures that it is not too concentrated in the frequency domain (a characteristic that a random source has with high probability). The error bound for the deterministic model is more general but is also slightly weaker than that for the random model.

The theorems provide sufficient conditions on the observation length L that guarantee that the estimation error will fall below a certain threshold. The number of samples in these sufficient conditions depends on the length of the impulse responses K , their intrinsic dimensions D , the number of channels M , and the signal-to-noise-ratio (SNR) defined as

$$\eta := \frac{\mathbb{E}_\phi[\sum_{m=1}^M \|h_m \otimes x\|_2^2]}{\mathbb{E}_w[\sum_{m=1}^M \|w_m\|_2^2]}. \quad (14)$$

Under (A1) and (A2), it follows from the commutativity of convolution and Lemma 35 that η simplifies to

$$\eta = \frac{K \|x\|_2^2 \|u\|_2^2}{M L \sigma_w^2}. \quad (15)$$

In addition, the bounds will depend on the spread of the channel gains. We measure this disparity using the two flatness parameters

$$\mu := \max_{1 \leq m \leq M} \frac{\sqrt{M} a_m}{\|a\|_2} \quad (16)$$

and

$$\nu := \min_{1 \leq m \leq M} \frac{\sqrt{M} a_m}{\|a\|_2}. \quad (17)$$

Our results are most interesting when there are not too many weak channels, meaning $\mu = O(1)$ and $\nu = \Omega(1)$. To simplify the theorem statements below, we will assume these conditions on μ and ν . It is possible, however, to re-work their statements to make the dependence on μ, ν explicit.

We now present our first main result. Theorem 2 below assumes a *random common source* signal x . We present guarantees for Algorithms 1 and 2 simultaneously, with $\underline{h}_t = \Phi v_t$ as the channel estimate after iteration t (for the alternating eigenvectors method, take $v_t = \hat{a}_t \otimes \hat{b}_t$).

Theorem 2 (Random Source): We observe noisy channel outputs $\{y_m\}$ as in (1), with SNR η as in (14), and form a sequence of estimates $(\underline{h}_t)_{t \in \mathbb{N}}$ of the channel responses by either Algorithm 1 or Algorithm 2 from the initial estimate by Algorithm 3. Suppose assumptions (A1) and (A2) above hold. Let x be a sequence of zero-mean iid subgaussian random variables with variance σ_x^2 , $\eta \geq 1$, $\mu = O(1)$,

⁴The same assumptions were used for the analysis of the spectral method (8) in [22].

and $L \geq 3K$.⁵ Then for any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0, \alpha \in \mathbb{N}$ and constants $C_1(\beta), C_2(\beta)$ such that if there are a sufficient number of channels,

$$M \geq C_1(\beta) \log^\alpha(MKL), \quad (18)$$

that are sufficiently long (relative to the dimension D of the subspace prior),

$$K \geq C_1(\beta) D \log^\alpha(MKL), \quad (19)$$

and we have observed a sufficient number of samples at the output of each channel,

$$L \geq \frac{C_1(\beta) \log^\alpha(MKL)}{\eta} \left(\frac{K}{M^2} + \frac{D}{D \wedge M} \right), \quad (20)$$

then with probability exceeding $1 - CK^{-\beta}$, we can bound the approximation error by

$$\begin{aligned} \sin \angle(\underline{\mathbf{h}}_t, \underline{\mathbf{h}}) &\leq 2^{-t} \angle(\underline{\mathbf{h}}_0, \underline{\mathbf{h}}) + C_2(\beta) \log^\alpha(MKL) \\ &\quad \cdot \left[\frac{1}{\sqrt{\eta L}} \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{D \wedge M}} \right) + \frac{\sqrt{D}}{\eta \sqrt{ML}} \right] \end{aligned} \quad (21)$$

for all $t \in \mathbb{N}$.

The SNR requirement $\eta \geq 1$ was introduced to simplify the expressions in Theorem 2. The conditions in the low SNR regime $\eta < 1$ can be easily extracted from the proof of the theorem and Proposition 4 below.

We make the following remarks about the assumptions (18)–(20) in Theorem 2. The lower bound on the number of channels in (18) is very mild, M has to be only a logarithmic factor of the number of parameters involved in the problem. The condition (19) allows a low-dimensional subspace whose dimension scales (up to a logarithmic factor) linearly with K . For a fixed SNR and a large number of channels ($M = \Omega(\sqrt{K/D})$), the condition in (20) says that the length of observation can grow proportional to \sqrt{KD} . This is suboptimal when compared to the degrees of freedom $(M + D)/(M - 1)$ per channel. (The total number of unknowns is $L + M + D$ and we have ML equations.) In fact, if $L < K$, then the circular convolution modulo L of two vectors respectively of length K and L introduces aliasing due to the wrapping around of the vector of length K . This turns the deconvolution problem into the demixing problem of separating a mixture of convolutions. While it might be still possible to uniquely identify a solution in this blind demixing problem, the deconvolution approach in this paper does not apply. In other words, the requirement $L \geq K$ is the fundamental limitation of any approach that linearizes the problem using cross-convolution. However, this still marks a significant improvement over an earlier analysis of this problem [51], which depended on the concentration of subgaussian polynomial [52] and union bound arguments. The scaling laws of parameters have been sharpened significantly, and as we will see in the next section, its prediction

⁵Without the subspace prior, $L > K$ is necessary to claim that $\mathbf{Y}^* \mathbf{Y}$ has nullity 1 in the noiseless case. We used $L \geq 3K$ in the proof in order to use the identity that the circular convolutions of three vectors of length K modulo L indeed coincide with their linear convolution.

is consistent with the empirical results by Monte Carlo simulations in Section VI. Compared to the analysis for the other spectral method under the linear subspace model [22], Theorem 2 shows that the estimation error becomes smaller by factor \sqrt{D} .

To prove Theorem 2, we establish an intermediate result for the case where the input signal \mathbf{x} is deterministic. In this case, our bounds depend on the spectral norm $\rho_{\mathbf{x}}$ of the (appropriately restricted) autocorrelation matrix of \mathbf{x} ,

$$\rho_{\mathbf{x}} := \|\tilde{\mathbf{S}} \mathbf{C}_{\mathbf{x}}^* \mathbf{C}_{\mathbf{x}} \tilde{\mathbf{S}}^*\|,$$

where

$$\tilde{\mathbf{S}} = \begin{bmatrix} [\mathbf{0}_{K-1, L-K+1} & \mathbf{I}_{K-1}] \\ [\mathbf{I}_{2K-1} & \mathbf{0}_{2K-1, L-2K+1}] \end{bmatrix}. \quad (22)$$

Then the deterministic version of our recovery result is:

Theorem 3 (Deterministic Source): Suppose that the same assumptions hold as in Theorem 2, only with \mathbf{x} as a fixed sequence of numbers obeying

$$\rho_{\mathbf{x}} \leq C_3 \|\mathbf{x}\|_2^2. \quad (23)$$

If (18) and (19) hold, and

$$L \geq \frac{C_1(\beta) \log^\alpha(MKL)}{\eta} \left(\frac{K^2}{M^2} + \frac{KD}{D \wedge M} \right), \quad (24)$$

then with probability exceeding $1 - CK^{-\beta}$, we can bound the approximation error by

$$\begin{aligned} \sin \angle(\underline{\mathbf{h}}_t, \underline{\mathbf{h}}) &\leq 2^{-t} \angle(\underline{\mathbf{h}}_0, \underline{\mathbf{h}}) + \frac{C_2(\beta) \log^\alpha(MKL)}{\sqrt{\eta L}} \\ &\quad \cdot \left(\frac{K}{M} + \sqrt{\frac{KD}{D \wedge M}} \right) \end{aligned} \quad (25)$$

for all $t \in \mathbb{N}$.

The condition (23) can be interpreted as a kind of incoherence condition on the input signal \mathbf{x} . Since

$$\rho_{\mathbf{x}} \leq \|\mathbf{C}_{\mathbf{x}}\|^2 = L \|\hat{\mathbf{x}}\|_\infty^2,$$

where $\hat{\mathbf{x}} \in \mathbb{C}^L$ is the normalized discrete Fourier transform of \mathbf{x} , it is sufficient that $\hat{\mathbf{x}}$ is approximately flat for (23) to hold. This is a milder assumption than imposing an explicit stochastic model on \mathbf{x} as in Theorem 2. For the price of this relaxed condition, the requirement on L in (24) that activates Theorem 3 is more stringent compared to the analogous condition (20) in Theorem 2.

B. Proof of Main Results

The main results in Theorems 2 and 3 follow from the following proposition, the proof of which is deferred to Section V-C.

Proposition 4: Suppose the assumptions in (A1) and (A2) hold, $\rho_{\mathbf{x}}$ satisfies (23), $L \geq 3K$, $\mu = O(1)$, and $\nu = \Omega(1)$. For any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0, \alpha \in \mathbb{N}$ and constants C_1, C_2 that only depend on β , for which the following holds: If

$$K \geq C_1 D \log^\alpha(MKL), \quad (26)$$

$$M \geq C_1 \log^\alpha(MKL), \quad (27)$$

and

$$L \geq C_1 \log^\alpha(MKL) \cdot \left[\frac{\rho_{x,w}^2}{\eta K \sigma_w^2 \|\mathbf{x}\|_2^2} \left(\frac{D}{K \wedge M} + \frac{K}{M^2} + 1 \right) + \frac{D}{\eta^2} \right], \quad (28)$$

then

$$\sin \angle(\underline{\mathbf{h}}, \underline{\mathbf{h}}) \leq 2^{-t} \sin \angle(\underline{\mathbf{h}}_0, \underline{\mathbf{h}}) + \kappa, \quad \forall t \in \mathbb{N} \quad (29)$$

with probability $1 - CK^{-\beta}$, where κ satisfies

$$\kappa \leq C_2 \log^\alpha(MKL) \cdot \left(\frac{\sqrt{D}}{\eta \sqrt{ML}} + \frac{\rho_{x,w}}{\sqrt{\eta K L} \sigma_w \|\mathbf{x}\|_2} \cdot \left[\mu \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{M}} + \sqrt{\frac{D}{K}} \right) + 1 \right] \right). \quad (30)$$

The proofs of Theorems 2 and 3 are given by combining Proposition 4 with the following lemmas, taken from [22], which provide tail estimates on the signal autocorrelation and the signal-noise cross correlation.

Lemma 5 [22, Lemma 3.9]: Suppose (A2) holds and let \mathbf{x} be a fixed sequence of numbers obeying (23). For any $\beta \in \mathbb{N}$, there exists an absolute constant C such that

$$\rho_{x,w} \leq CK \sigma_w \sqrt{\rho_x} \sqrt{1 + \log M + \beta \log K}$$

holds with probability $1 - K^{-\beta}$.

Lemma 6 [22, Lemma 3.10]: Suppose (A2) holds and let \mathbf{x} be a sequence of zero-mean iid subgaussian random variables with variance σ_x^2 . Then

$$\frac{\rho_x}{\|\mathbf{x}\|_2^2} \leq \frac{L + C_\beta K \log^{3/2} L \sqrt{\log K}}{L - \sqrt{2L\beta \log K}}$$

and

$$\frac{\rho_{x,w}}{\sigma_w \|\mathbf{x}\|_2} \leq \frac{C_\beta K \log^5(MKL)}{\sqrt{L - \sqrt{2L\beta \log K}}}$$

hold with probability $1 - 3K^{-\beta}$.

C. Proof of Proposition 4

The proof of Proposition 4 is given by a set of propositions, which provide guarantees for Algorithm 1, Algorithm 2, and Algorithm 3. The first proposition provides a performance guarantee for the initialization by Algorithm 3. The proof of Proposition 7 is given in Section VII.

Proposition 7 (Initialization): Suppose the assumptions in (A1) and (A2) hold, ρ_x satisfies (23), and $L \geq 3K$. Let η, μ, ν be defined in (15), (16), (17), respectively. For any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0, \alpha \in \mathbb{N}$ and constants C_1, C_2 that only depend on β , for which the following holds: If

$$M \geq C_1 \log^\alpha(MKL) \cdot \left(\frac{\mu}{\nu} \right)^2 \quad (31)$$

and

$$L \geq C_1 \log^\alpha(MKL) \cdot \left[\frac{\rho_{x,w}^2}{\eta K \sigma_w^2 \|\mathbf{x}\|_2^2} \cdot \left(\frac{\mu^2 K}{\nu^4 M^2} + \frac{D}{\nu^2 M} \right) + \frac{D}{\eta^2 \nu^4 M} \right], \quad (32)$$

then the estimate $\hat{\mathbf{b}}$ by Algorithm 3 satisfies

$$\sin \angle(\hat{\mathbf{b}}, \mathbf{b}) \leq C_2 \log^\alpha(MKL) \left[\frac{\mu}{\nu \sqrt{M}} + \frac{\sqrt{D}}{\eta \nu^2 \sqrt{ML}} + \frac{\rho_{x,w}}{\sqrt{\eta K L} \sigma_w \|\mathbf{x}\|_2} \cdot \left(\frac{\mu \sqrt{K}}{\nu^2 M} + \frac{\sqrt{D}}{\nu \sqrt{M}} \right) \right] \quad (33)$$

with probability $1 - CK^{-\beta}$.

The second proposition, proved in Section VIII-B, provides a performance guarantee for the update of $\hat{\mathbf{a}}$ by Step 3 of Algorithm 1.

Proposition 8 (Update of Channel Gains): Suppose the assumptions in (A1) and (A2) hold, ρ_x satisfies (23), $L \geq 3K$, and the previous estimate $\hat{\mathbf{b}}$ satisfies

$$\angle(\mathbf{b}, \hat{\mathbf{b}}) \leq \frac{\pi}{4}. \quad (34)$$

For any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0, \alpha \in \mathbb{N}$ and constants C_1, C_2 that only depend on β , for which the following holds: If

$$K \geq C_1 \mu^4 D \log^\alpha(MKL), \quad (35)$$

$$M \geq C_1 \mu^4 \log^\alpha(MKL), \quad (36)$$

and

$$L \geq C_1 \log^\alpha(MKL) \cdot \left[\frac{\rho_{x,w}^2}{\eta K \sigma_w^2 \|\mathbf{x}\|_2^2} \left(\mu^2 \left(\frac{D}{K \wedge M} + \frac{K}{M^2} \right) + 1 \right) + \frac{D}{\eta^2} \right], \quad (37)$$

then the updated $\hat{\mathbf{a}}$ by Step 3 of Algorithm 1 satisfies

$$\sin \angle(\mathbf{a}, \hat{\mathbf{a}}) \leq \frac{1}{2} \sin \angle(\mathbf{b}, \hat{\mathbf{b}}) + \kappa \quad (38)$$

with probability $1 - CK^{-\beta}$, where κ satisfies (30).

We have a similar result for the update of $\hat{\mathbf{b}}$ by Step 4 of Algorithm 1, which is stated in the following proposition. The proof of Proposition 9 is provided in Section VIII-C.

Proposition 9 (Update of Subspace Coefficients): Suppose the assumptions in (A1) and (A2) hold, ρ_x satisfies (23), $L \geq 3K$, and the previous estimate $\hat{\mathbf{a}}$ satisfies

$$\angle(\mathbf{a}, \hat{\mathbf{a}}) \leq \frac{\pi}{4}. \quad (39)$$

For any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0, \alpha \in \mathbb{N}$ and constants C_1, C_2 that depend on β , for which the following holds: If (35), (36), and (37) are satisfied, then the updated $\hat{\mathbf{b}}$ by Step 4 of Algorithm 1 satisfies

$$\sin \angle(\mathbf{b}, \hat{\mathbf{b}}) \leq \frac{1}{2} \sin \angle(\mathbf{a}, \hat{\mathbf{a}}) + \kappa$$

with probability $1 - CK^{-\beta}$, where κ satisfies (30).

The next proposition shows the convergence of the rank-1 truncated power method from a provably accurate initialization. See Section IX for the proof.

Proposition 10 (Local Convergence of Rank-1 Truncated Power Method): Suppose the assumptions in (A1) and (A2) hold, ρ_x satisfies (23), and $L \geq 3K$. Let $0 < \mu < 1$, $0 < \tau < \frac{1}{3\sqrt{2}}$, and

$$c(\mu, \tau) = \min \left(\frac{1}{\mu \sqrt{1 - \tau^2}}, \frac{(1 + \mu)\tau}{1 - \mu} \right).$$

For any $\beta \in \mathbb{N}$, there exist absolute constants $C > 0$, $\alpha \in \mathbb{N}$, constants C'_1, C'_2 that only depend on β , for which the following holds: If (35), (36), and (37) are satisfied for $C_1 = c(\mu, \tau)C'_1, C_2 = c(\mu, \tau)C'_2$ and \mathbf{u}_0 satisfies

$$\sin \angle(\mathbf{u}_0, \mathbf{u}) \leq \tau,$$

then $(\mathbf{u}_t)_{t \in \mathbb{N}}$ produced by Algorithm 2 for $\mathbf{B} = \|\mathbb{E}[\mathbf{A}]\| \mathbf{I}_{MD} - \mathbf{A}$ with \mathbf{u}_0 satisfies

$$\sin \angle(\mathbf{u}_t, \mathbf{u}) \leq \mu^t \sin \angle(\mathbf{u}_0, \mathbf{u}) + \frac{(1 + \mu)\kappa}{1 - \mu}, \quad \forall t \in \mathbb{N} \quad (40)$$

with probability $1 - CK^{-\beta}$, where κ satisfies (30).

Finally, we derive the proof of Proposition 4 by combining the above propositions.

Proof of Proposition 4: Similar to the proof of [22, Proposition 3.3], we show that

$$\sin \angle(\mathbf{h}_t, \mathbf{h}) \leq \frac{\sigma_{\max}(\Phi)}{\sigma_{\min}(\Phi)} \cdot \sqrt{2} \sin \angle(\mathbf{u}_t, \mathbf{u}) \quad (41)$$

and

$$\sin \angle(\mathbf{u}_t, \mathbf{u}) \leq \max[\sin \angle(\mathbf{a}_t, \mathbf{a}), \sin \angle(\mathbf{b}_t, \mathbf{b})]. \quad (42)$$

Furthermore, as we choose C_1 in (26) sufficiently large, we can upper bound the condition number of Φ by a constant (e.g., 3) with high probability. We continue the proof conditioned on this event. Then the convergence results in Propositions 8, 9, and 10 imply (29).

Since $\mu = O(1)$, the conditions in (35), (36), (37) respectively reduce (26), (27), (28). Furthermore, since $\nu = \Omega(1)$, (32) is implied by (28). By choosing C_1 large enough, we can make the initial error bound in (33) small so that the conditions for previous estimates in Propositions 8, 9, 10 are satisfied and the assertion is obtained by these propositions. \square

VI. NUMERICAL RESULTS

In this section, we provide observation on empirical performance of the alternating eigenvectors method (AltEig) in Algorithm 1 and the rank-1 truncated power method (RTPM) in Algorithm 2, using the spectral initialization in Algorithm 3. We compare the two iterative algorithms to the classical cross-convolution method (CC) by Xu *et al.* [19], which only imposes the time-limited model on impulse responses, and to the subspace-constrained cross-convolution method (SCCC) [22], which imposes a linear subspace model on the impulse responses. This comparison will demonstrate how the estimation error improves progressively as we impose a stronger prior model on the impulse responses.

In our first experiment, we tested the algorithms on generic data where the basis Φ is an i.i.d. Gaussian matrix. The input source signal \mathbf{x} , subspace coefficient vector \mathbf{b} , and additive noise are i.i.d. Gaussian as well. The channel gain vector is generated by adding random perturbation to all-one vector so that $\mathbf{a} = \mathbf{1}_{M,1} + \alpha \xi / \|\xi\|_\infty$, where $\xi = [\xi_1, \dots, \xi_M]^\top$ and ξ_1, \dots, ξ_M are independent copies of a uniform random variable on $[-1, 1]$. We use a performance metric given as the 95th percentile of the estimation error in the sine of the principal angle between the estimate and the ground truth out of 1,000 trials. This amounts to the worst-case error after

excepting 5% of the instances. In other words, the estimation error is less than this threshold with high probability no less than 0.95.

Figure 2 compares the estimation error by the four algorithms as we vary the problem parameters. Figure 2(a) shows that the error as a function of the oversampling factor L/K , which is the ratio of the length of observation L to the number of nonzero coefficients in each impulse response. SCCC provides smaller estimation error than CC in order of magnitude by exploiting the additional linear subspace prior. Then AltEig and RTPM provide further reduced estimation error again in order of magnitude compared to SCCC by exploiting the bilinear prior that imposes the separability structure in addition to the linear subspace prior. As expected, longer observation provides a smaller estimation error for all methods. Furthermore, as shown in Figure 2(b), the estimation error increases as a function of the ratio D/K , which accounts for the relative dimension of the subspace. More interestingly, as our main theorems imply, the performance difference between SCCC and AltEig/RTPM becomes more significant as we add more channels (Figure 2(c)). The estimation error by these method scales proportionally as a function of SNR (Figure 2(d)). Similarity of channel gains, as captured by the parameter α , did not affect the estimation error significantly (Figure 2(e)). Moreover, when the two iterative algorithms (AltEig and RTPM) provide stable estimate, they converge very quickly. Figure 3 illustrates the convergence of the two algorithms for a random instance. The estimation error decays progressively for RTPM, whereas AltEig converges in less than 5 iterations.

To better visualize the overall trend, we performed a Monte Carlo simulation for the empirical phase transition. This is illustrated in Figure 4 with a color coding that uses a logarithmic scale with blue denoting the smallest and red the largest error within the regime of $(D/K, L/K)$. The error in the estimate by CC is large (≥ 0.1) regardless of D/K for the entire regime (Figure 4(a)). SCCC provides accurate estimates for small D/K and for large enough L/K (Figure 4(b)). On the other hand, it totally fails unless the dimension D of subspace is less than a certain threshold. Finally, AltEig and RTPM show almost the same empirical phase transitions, which imply robust recovery for larger D/K and for smaller L/K (Figures 4(c) and 4(d)).

The above illustrates the performance of SCCC, AltEig, and RTPM for $\hat{\sigma}_w^2 = \sigma_w^2$, i.e. in the scenario when the true noise variance is given. Figures 4(e) and 4(f) illustrate the empirical phase transitions for AltEig and RTPM when a crude estimate of σ_w^2 given as $\hat{\sigma}_w^2 = 0$ is used instead. These figures show that there is a nontrivial difference in the regime for accurate estimation depending on the quality of the estimate $\hat{\sigma}_w^2$. This opens up an interesting question of how to show a guarantee for the noise variance estimation. Nonetheless, even with $\hat{\sigma}_w^2 = 0$, both AltEig and RTPM show improvements in their empirical performances due to the extra structural constraint on the impulse responses over CC and SCCC, which are (partially) blind to the bilinear prior model.

In our second experiment, we tested the algorithms on synthesized data with a realistic underwater acoustic channel

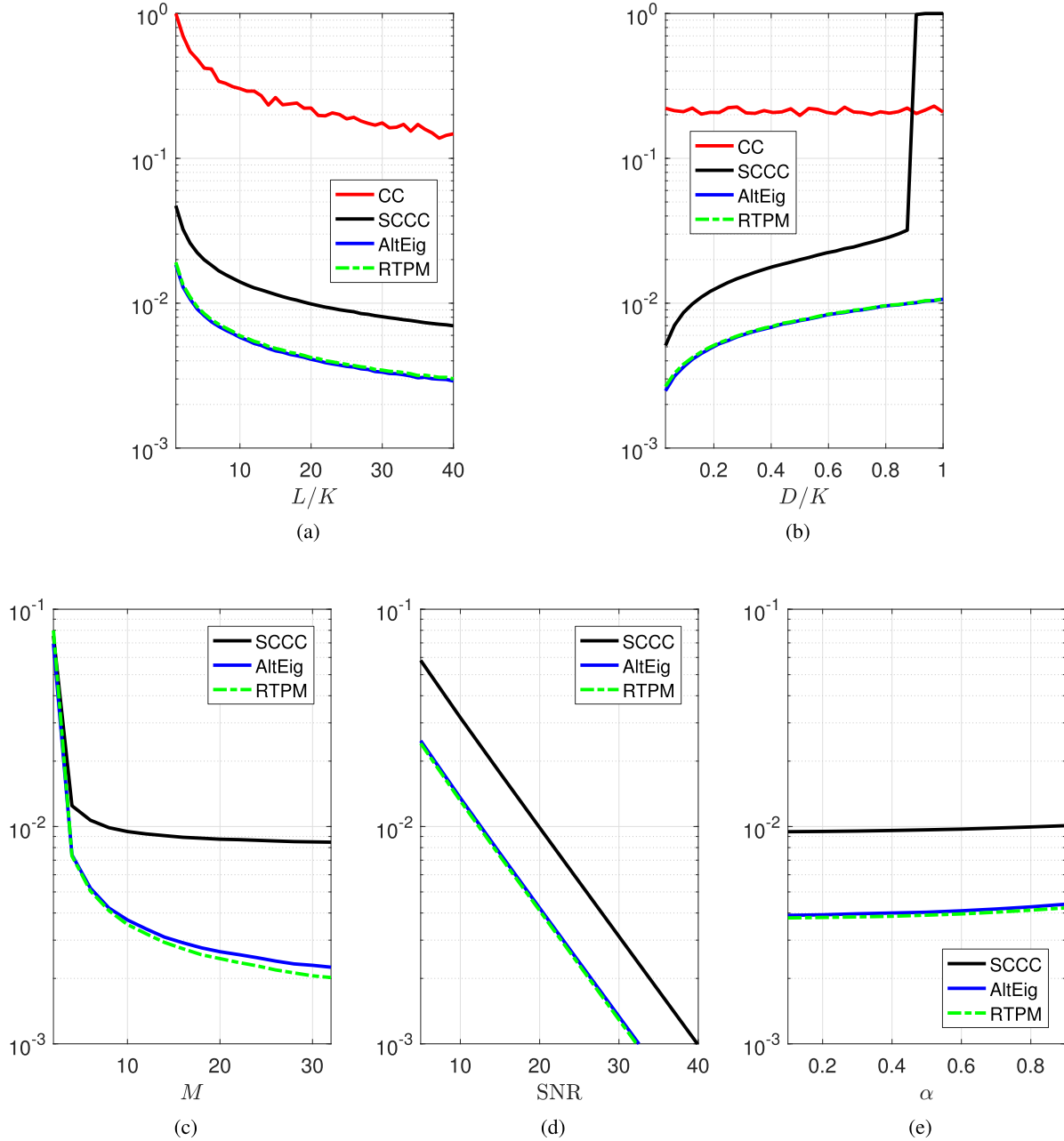


Fig. 2. Comparison of cross-convolution (CC), subspace-constrained cross-convolution (SCCC), alternating eigenvectors method (AltEig), and rank-1 truncated power method (RTPM). Default parameter setting: $M = 8$, $K = 256$, $D = 32$, $L = 20K$, SNR = 20 dB. The 95th percentile estimation error is plotted in a logarithmic scale as we vary each parameter as follows: (a) L . (b) D . (c) M . (d) SNR. (e) α .

model, where the impulse responses are approximated by a bilinear channel model. In an ocean acoustic array sensing scenario, receivers of the vertical line array (VLA) with equal spacing listen to the same source near the ocean surface at a distance. In a simple channel, each receiver will observe essentially the same signal, only at different delays that depend on the orientation of the source relative to the array. This geometry makes the shape of CIRs closely linked to one another (and hence amenable to a joint linear model), while the relative gains of the receivers are independent. A detailed description on how to form the basis Φ for a particular underwater environment can be found in [53].

We performed Monte-Carlo simulation to demonstrate the robustness of our method on realistic acoustic channels which represent an at-sea experiment carried out in the Santa Barbara Channel. In the simulation, the common driving source signal, $\mathbf{x} \in \mathbb{R}^L$, is white Gaussian noise filtered in an arbitrary bandwidth representative of shipping noise spectra (400–600 Hz) for $L = 2000$. Each CIR is of length $K = 500$ and represents a Gaussian-windowed pulse in the band of 400–600 Hz. The number of channels M is 31. The basis $\Phi \in \mathbb{R}^{K \times D}$ is of dimension $D = 8$. The number of trials in the Monte-Carlo simulation is 100. In this experiment, unlike the previous experiments with Gaussian bases,

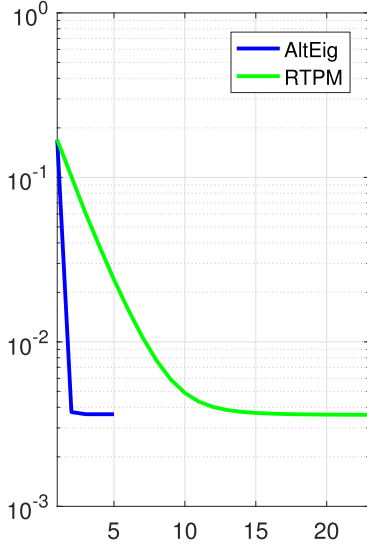


Fig. 3. Convergence of alternating eigenvectors method (AltEig) and rank-1 truncated power method (RTPM) for a random instance. x -axis: iteration index, y -axis: log of the estimation error. $M = 8$, $K = 256$, $D = 32$, $L = 20$ K , SNR = 20 dB.

AltEig does not provide a stable recovery. Therefore, we report the results only for RTPM. As for the estimation of the noise variance, since the basis matrices are unitary, there is no need to subtract the expectation of the noise auto-correlation term. Figure 5 shows order statistics of the estimation error in a log scale. The empirical success rate of the recovery in this experiment is lower than that in the first experiment with generic data. The median of the estimation error approaches to the modeling error due to approximation with a bilinear model as we increase the SNR.

VII. ANALYSIS OF SPECTRAL INITIALIZATION

We prove Proposition 7 in this section. Recall that Algorithm 3 computes an initial estimate $\hat{\mathbf{b}}$ of the true parameter vector \mathbf{b} as an eigenvector of $\mathbf{\Gamma}\mathbf{\Gamma}^* - \sum_{m=1}^M \sigma_w^2 \mathbf{L}\mathbf{\Phi}_m^* \mathbf{\Phi}_m$ corresponding to the largest eigenvalue in magnitude. Let us decompose the matrix $\mathbf{\Gamma}$ in (11) as $\mathbf{\Gamma} = \mathbf{\Gamma}_s + \mathbf{\Gamma}_n$, where $\mathbf{\Gamma}_s$ and $\mathbf{\Gamma}_n$ respectively correspond to the noise-free portion and noise portion of $\mathbf{\Gamma}$. In other words, $\mathbf{\Gamma}_s$ is obtained as we replace $\mathbf{y}_m = \mathbf{h}_m \otimes \mathbf{x} + \mathbf{w}_m$ in the expression of $\mathbf{\Gamma}$ in (12) by its first summand $\mathbf{h}_m \otimes \mathbf{x}$. Similarly, $\mathbf{\Gamma}_n$ is obtained as we replace \mathbf{y}_m by \mathbf{w}_m . Then it follows that

$$\mathbb{E}_w[\mathbf{\Gamma}_n \mathbf{\Gamma}_n^*] = \sum_{m=1}^M \sigma_w^2 \mathbf{L} \mathbf{\Phi}_m^* \mathbf{\Phi}_m.$$

By direct calculation, we obtain that the expectation of $\mathbf{\Gamma}_s$ is written as

$$\mathbb{E}[\mathbf{\Gamma}_s] = \sum_{m=1}^M K a_m \mathbf{b} \mathbf{x}^\top = K \|\mathbf{a}\|_1 \mathbf{b} \mathbf{x}^\top. \quad (43)$$

Therefore,

$$\mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^* = K^2 \|\mathbf{x}\|_2^2 \|\mathbf{a}\|_1^2 \mathbf{b} \mathbf{b}^*. \quad (44)$$

It is straightforward to check that the rank-1 matrix $\mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^*$ has an eigenvector, which is collinear with \mathbf{b} . Thus as we interpret $\mathbf{\Gamma}\mathbf{\Gamma}^* - \sum_{m=1}^M \sigma_w^2 \mathbf{L} \mathbf{\Phi}_m^* \mathbf{\Phi}_m$ as a perturbed version of $\mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^*$, the error in $\hat{\mathbf{b}}$ is upper bounded by the classical result in linear algebra known as the Davis-Kahan theorem [54]. Among numerous variations of the original Davis-Kahan theorem available in the literature, we will use a consequence of a particular version [55, Th. 8.1.12], which is stated as the following lemma.

Lemma 11 (A Special Case of the Davis-Kahan Theorem): Let $\mathbf{M}, \mathbf{\underline{M}} \in \mathbb{C}^{n \times n}$ be symmetric matrices and λ denote the largest eigenvalue of $\mathbf{\underline{M}}$ in magnitude. Suppose that $\lambda > 0$ and has multiplicity 1. Let $\mathbf{Q} = [\mathbf{q}_1, \mathbf{Q}_2] \in \mathbb{C}^{n \times n}$ be a unitary matrix such that \mathbf{q}_1 is an eigenvector of $\mathbf{\underline{M}}$ corresponding to λ . Partition the matrix $\mathbf{Q}^* \mathbf{M} \mathbf{Q}$ as follows:

$$\mathbf{Q}^* \mathbf{M} \mathbf{Q} = \begin{bmatrix} \lambda & \mathbf{0}_{1, n-1} \\ \mathbf{0}_{n-1, 1} & \mathbf{D} \end{bmatrix}.$$

If

$$\|\mathbf{D}\| + \|\mathbf{M} - \mathbf{\underline{M}}\| \leq \frac{\lambda}{5}, \quad (45)$$

then the largest eigenvalue of \mathbf{M} in magnitude has multiplicity 1 and the corresponding eigenvector $\tilde{\mathbf{q}}$ satisfies

$$\sin \angle(\tilde{\mathbf{q}}, \mathbf{q}_1) \leq \frac{4\|(\mathbf{M} - \mathbf{\underline{M}})\mathbf{q}_1\|_2}{\lambda}. \quad (46)$$

Remark 12: In Lemma 11, the rank-1 matrix $\lambda \mathbf{q}_1 \mathbf{q}_1^*$ is considered as the ground truth matrix. Then $\mathbf{M} - \mathbf{\underline{M}} + \mathbf{Q}_2 \mathbf{D} \mathbf{Q}_2^*$ corresponds to perturbation in \mathbf{M} relative to the ground truth matrix $\mathbf{\underline{M}}$. Also note that $\mathbf{Q}_2 \mathbf{D} \mathbf{Q}_2^* \mathbf{q}_1 = \mathbf{0}$.

In the remainder of this section, we obtain an upper bound on the error in $\hat{\mathbf{b}}$ by applying Lemma 11 to $\mathbf{\underline{M}} = \mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^*$, $\mathbf{M} = \mathbf{\Gamma}\mathbf{\Gamma}^* - \sum_{m=1}^M \sigma_w^2 \mathbf{L} \mathbf{\Phi}_m^* \mathbf{\Phi}_m$, $\mathbf{q}_1 = \mathbf{b}$, and $\hat{\mathbf{q}} = \hat{\mathbf{b}}$.

By (44), we have $\mathbf{D} = \mathbf{0}$ and $\lambda = K^2 \|\mathbf{x}\|_2^2 \|\mathbf{a}\|_1^2 \|\mathbf{b}\|_2^2$. Then we show that the spectral norm of the perturbation term, which is rewritten as

$$\mathbf{\Gamma}\mathbf{\Gamma}^* - \mathbb{E}_w[\mathbf{\Gamma}_n \mathbf{\Gamma}_n^*] - \mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^* = \mathbf{\Gamma}_s \mathbf{\Gamma}_s^* - \mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^*, \quad (47a)$$

$$+ \mathbf{\Gamma}_s \mathbf{\Gamma}_n^* + \mathbf{\Gamma}_n \mathbf{\Gamma}_s^*, \quad (47b)$$

$$+ \mathbf{\Gamma}_n \mathbf{\Gamma}_n^* - \mathbb{E}_w[\mathbf{\Gamma}_n \mathbf{\Gamma}_n^*], \quad (47c)$$

satisfies (45). We will compute an upper estimate of the spectral norm of each summand, divided by λ , separately. Then we combine these estimates using the triangle inequality.

A. Perturbation Due to Signal Term

Note that the first summand $\mathbf{\Gamma}_s \mathbf{\Gamma}_s^* - \mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^*$ in (47a) has entries, which are fourth-order Gaussian random variables. We decompose it using second-order random variables as

$$\begin{aligned} \mathbf{\Gamma}_s \mathbf{\Gamma}_s^* - \mathbb{E}[\mathbf{\Gamma}_s] \mathbb{E}[\mathbf{\Gamma}_s]^* &= (\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s])(\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s])^* \\ &\quad + \mathbb{E}[\mathbf{\Gamma}_s](\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s])^* \\ &\quad + (\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s])\mathbb{E}[\mathbf{\Gamma}_s]^*. \end{aligned} \quad (48)$$

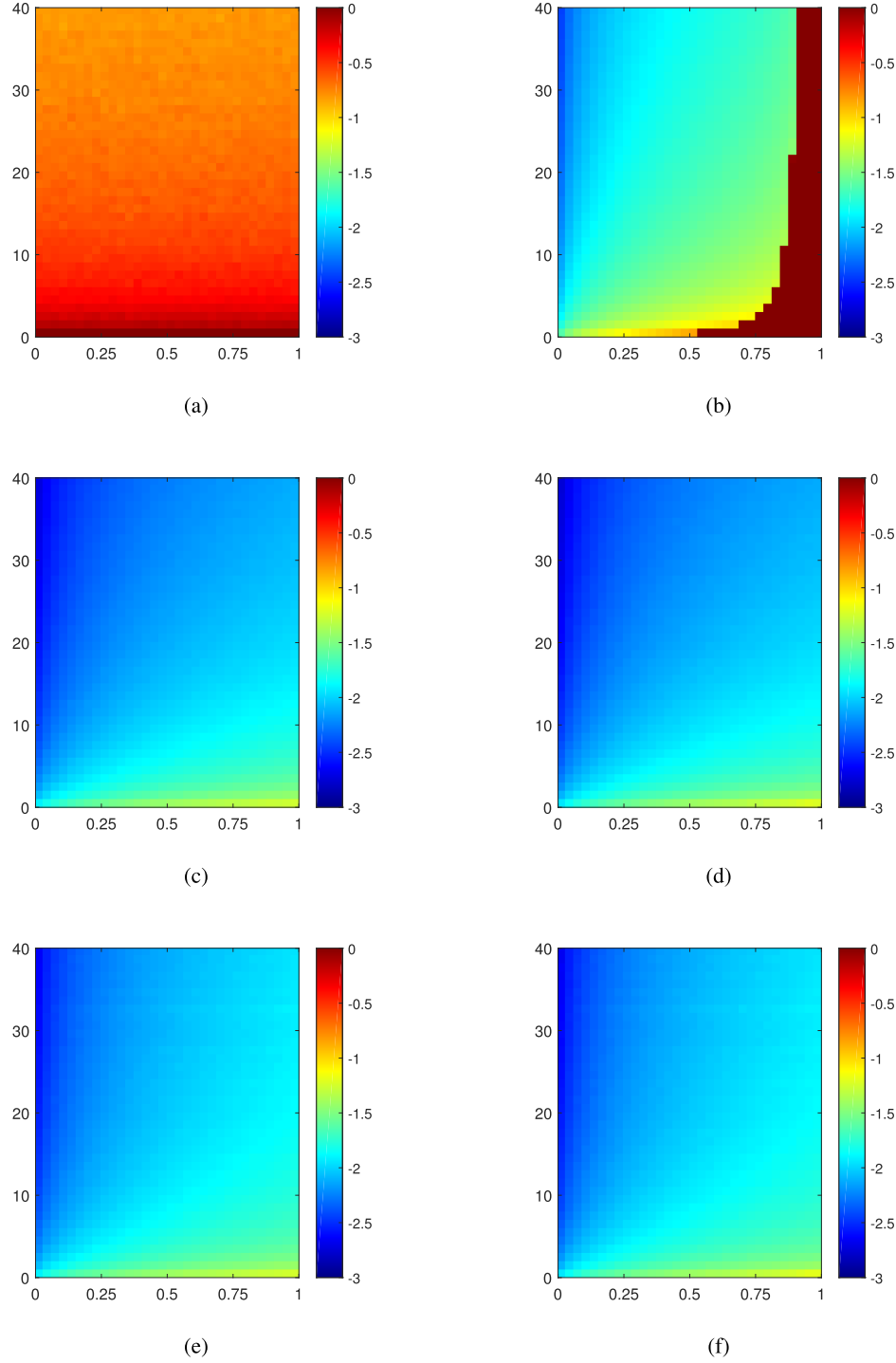


Fig. 4. Empirical phase transition in the 95th percentile of the log estimation error. x -axis: D/K . y -axis: L/K . $K = 256$, $M = 8$, $\text{SNR} = 20$ dB. (a) cross-convolution method [19]. (b) subspace-constrained cross-convolution method [22]. (c) alternating eigenvectors method ($\hat{\sigma}_w^2 = \sigma_w^2$). (d) rank-1 truncated power method ($\hat{\sigma}_w^2 = \sigma_w^2$). (e) alternating eigenvectors method ($\hat{\sigma}_w^2 = 0$). (f) rank-1 truncated power method ($\hat{\sigma}_w^2 = 0$).

We have already computed $\mathbb{E}[\mathbf{\Gamma}_s]$ in (43). It remains to upper bound the spectral norm of $\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s]$. By the definitions of $\mathbf{\Gamma}_s$ and ρ_x , we obtain

$$\begin{aligned} \|\mathbf{\Gamma}_s - \mathbb{E}[\mathbf{\Gamma}_s]\| &\leq \left\| \sum_{m=1}^M a_m (\Phi_m^* S C S^* \Phi_m b - \mathbb{E}_\phi[\Phi_m^* S C S^* \Phi_m b]) \check{S}^* \check{S} C x \right\| \\ &\leq \left\| \sum_{m=1}^M a_m (\Phi_m^* S C S^* \Phi_m b - \mathbb{E}_\phi[\Phi_m^* S C S^* \Phi_m b]) \check{S}^* \right\| \|\check{S} C x\| \\ &\leq \sqrt{\rho_x} \left\| \sum_{m=1}^M a_m (\Phi_m^* S C S^* \Phi_m b - \mathbb{E}_\phi[\Phi_m^* S C S^* \Phi_m b]) \check{S}^* \right\|, \end{aligned} \quad (49)$$

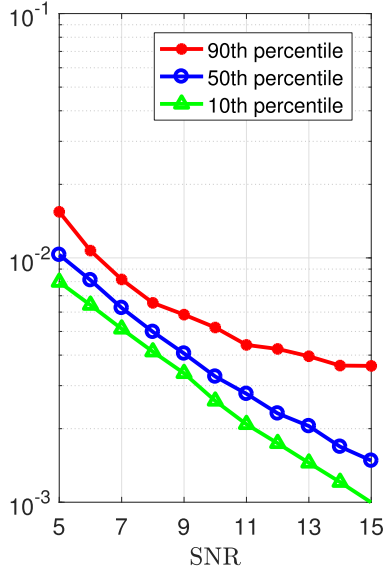


Fig. 5. Order statistics for the log estimation error in varying SNR for the underwater channel model.

where $\check{\mathbf{S}} \in \mathbb{R}^{(2K-1) \times L}$ is defined by

$$\check{\mathbf{S}} = \begin{bmatrix} [\mathbf{0}_{K-1, L-K+1} \quad \mathbf{I}_{K-1}] \\ [\mathbf{I}_K \quad \mathbf{0}_{K, L-K}] \end{bmatrix}.$$

The right-hand side of (49) except the constant factor $\sqrt{\rho_x}$ is upper bounded by the following lemma, which is proved in Appendix X.

Lemma 13: Suppose that (A2) holds. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\left\| \sum_{m=1}^M a_m (\Phi_m^* \mathbf{S} \mathbf{C} \mathbf{S}^* \Phi_m \mathbf{b} - \mathbb{E}_\phi [\Phi_m^* \mathbf{S} \mathbf{C} \mathbf{S}^* \Phi_m \mathbf{b}]) \check{\mathbf{S}}^* \right\| \leq C(\beta) K \sqrt{M} \|\mathbf{a}\|_\infty \|\mathbf{b}\|_2 \log^\alpha(MKL) \quad (50)$$

holds with probability $1 - K^{-\beta}$.

By applying (43), (49), Lemma 13 together with the fact $\sqrt{\rho_x} \leq C_0 \|\mathbf{x}\|_2$ to (48), we obtain that

$$\begin{aligned} \frac{\|\Gamma_s \Gamma_s^* - \mathbb{E}[\Gamma_s] \mathbb{E}[\Gamma_s]^*\|}{\lambda} &\leq \frac{C(\beta) \sqrt{M} \|\mathbf{a}\|_\infty \log^\alpha(MKL)}{\|\mathbf{a}\|_1} \\ &\leq \frac{C(\beta) \mu \log^\alpha(MKL)}{v \sqrt{M}} \end{aligned} \quad (51)$$

holds with probability $1 - K^{-\beta}$.

B. Perturbation Due to Signal-Noise Cross Term

Next we consider the second term in (47b). By the triangle inequality, we have

$$\|\Gamma_s \Gamma_n^* + \Gamma_n \Gamma_s^*\| \leq \|\Gamma_s \Gamma_n^*\| + \|\Gamma_n \Gamma_s^*\| \leq 2 \|\Gamma_s \Gamma_n^*\|.$$

Therefore, it suffices to upper estimate $\|\Gamma_s \Gamma_n^*\|$. To this end, we decompose $\Gamma_s \Gamma_n^*$ as

$$\Gamma_s \Gamma_n^* = (\Gamma_s - \mathbb{E}[\Gamma_s]) \Gamma_n^* + \mathbb{E}[\Gamma_s] \Gamma_n^*. \quad (52)$$

Note that the first summand in the right-hand side of (52) is written as

$$\begin{aligned} &(\Gamma_s - \mathbb{E}[\Gamma_s]) \Gamma_n^* \\ &= \left(\sum_{m=1}^M a_m \Phi_m^* \mathbf{S} \mathbf{C} \mathbf{S}^* \Phi_m \mathbf{b} \check{\mathbf{S}}^* - \mathbb{E}_\phi [a_m \Phi_m^* \mathbf{S} \mathbf{C} \mathbf{S}^* \Phi_m \mathbf{b} \check{\mathbf{S}}^*] \right) \\ &\quad \cdot \left(\sum_{m'=1}^M \check{\mathbf{S}} \mathbf{C} \mathbf{x} \mathbf{C}_{\mathbf{w}_{m'}}^* \mathbf{S}^* \Phi_{m'} \right), \end{aligned} \quad (53)$$

where the first and second factors of the right-hand side of (53) are upper bounded in the spectral norm respectively by Lemma 13 and by the following lemma. (See Appendix X for the proof.)

Lemma 14: Suppose that (A1) and (A2) hold. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\left\| \sum_{m=1}^M \check{\mathbf{S}} \mathbf{C} \mathbf{x} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{S}^* \Phi_m \right\| \leq C(\beta) \rho_{x,w} \sqrt{MK} \log^\alpha(MKL) \quad (54)$$

holds with probability $1 - K^{-\beta}$.

By applying Lemmas 13 and (54), we obtain that

$$\|(\Gamma_s - \mathbb{E}[\Gamma_s]) \Gamma_n^*\| \leq C(\beta) \rho_{x,w} M K^{3/2} \|\mathbf{a}\|_\infty \|\mathbf{b}\|_2 \log^\alpha(MKL) \quad (55)$$

holds with probability $1 - K^{-\beta}$.

Next, the second summand in the right-hand side of (52) is written as

$$\mathbb{E}[\Gamma_s] \Gamma_n^* = K \|\mathbf{a}\|_1 \mathbf{b} \left(\sum_{m'=1}^M \mathbf{e}_1^* \mathbf{C} \mathbf{x} \mathbf{C}_{\mathbf{w}_{m'}}^* \mathbf{S}^* \Phi_{m'} \right), \quad (56)$$

whose spectral norm is upper bounded by using the following lemma.

Lemma 15: Suppose that (A1) and (A2) hold. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\left\| \sum_{m'=1}^M \mathbf{e}_1^* \mathbf{C} \mathbf{x} \mathbf{C}_{\mathbf{w}_{m'}}^* \mathbf{S}^* \Phi_{m'} \right\| \leq C(\beta) \rho_{x,w} \sqrt{MD} \log^\alpha(MKL)$$

holds with probability $1 - K^{-\beta}$.

The proof of Lemma 15 is very similar to that of Lemma 14. The proof of Lemma 14 involves the following optimization formulation:

$$\max_{\mathbf{z} \in B_2^{2K-1}} \max_{\mathbf{q} \in B_2^D} \sum_{m=1}^M \mathbf{z}^* \check{\mathbf{S}} \mathbf{C} \mathbf{x} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{S}^* \Phi_m \mathbf{q}.$$

Instead of maximizing over $\mathbf{z} \in B_2^{2K-1}$, we fix \mathbf{z} to $\check{\mathbf{S}} \mathbf{e}_1$. Equivalently, we replace the unit ball B_2^{2K-1} by the singleton set $\{\check{\mathbf{S}} \mathbf{e}_1\}$. This replacement simply removes the entropy integral corresponding to B_2^{2K-1} . Except this point, the proofs for the two lemmas are identical. Thus we omit further details.

Applying Lemma 15 to (56) implies that

$$\|\mathbb{E}[\Gamma_s] \Gamma_n^*\| \leq C(\beta) \rho_{x,w} \sqrt{MK} \sqrt{D} \|\mathbf{a}\|_1 \|\mathbf{b}\|_2 \log^\alpha(MKL) \quad (57)$$

holds with probability $1 - K^{-\beta}$.

By combining (55) and (57), after plugging in the definitions of η , μ , and ν , we obtain that

$$\frac{\|\Gamma_s \Gamma_n^* + \Gamma_n \Gamma_s^*\|}{\lambda} \leq \frac{C(\beta) \log^\alpha(MKL)}{\sqrt{\eta}} \cdot \frac{\rho_{x,w}}{\|\mathbf{x}\|_2 \sigma_w \sqrt{L}} \cdot \left(\frac{\mu}{\nu^2 M} + \frac{\sqrt{D}}{\nu \sqrt{MK}} \right) \quad (58)$$

holds with probability $1 - 2K^{-\beta}$.

C. Perturbation Due to Noise Term

Finally, we derive an upper bound on the spectral norm of the last term in (47c) using the following lemma, which is proved in Appendix X.

Lemma 16: Suppose that (A2) holds. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\|\Gamma_n \Gamma_n^* - \mathbb{E}_w[\Gamma_n \Gamma_n^*]\| \leq C(\beta) \rho_w M^{3/2} \sqrt{KL} \log^\alpha(MKL)$$

holds with probability $1 - K^{-\beta}$.

We also use a tail bound on ρ_w given by the following lemma from [22].

Lemma 17 ([22, Lemma 5.9]): Suppose that (A2) holds. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\rho_w \leq C(\beta) \sigma_w^2 \sqrt{KL} \log^\alpha(MKL) \quad (59)$$

holds with probability $1 - K^{-\beta}$.

By Lemma 17 and (15), the corresponding relative perturbation is upper bounded by

$$\frac{\|\Gamma_n \Gamma_n^* - \mathbb{E}_w[\Gamma_n \Gamma_n^*]\|}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{a}\|_1^2 \|\mathbf{b}\|_2^2} \leq \frac{C(\beta) \log^\alpha(MKL)}{\eta} \cdot \frac{\sqrt{D}}{\nu^2 \sqrt{ML}} \quad (60)$$

with probability $1 - K^{-\beta}$.

Then it follows from (51), (58), and (60) that the condition in (45) is satisfied by the assumptions in (32) and (31). Therefore, Lemma 11 provides the upper bound on the estimation error in (33), which is obtained by plugging (51), (58), and (60) to (46). This completes the proof.

VIII. CONVERGENCE OF ALTERNATING EIGEN METHOD

Algorithm 1 iteratively updates the estimates of \mathbf{a} , \mathbf{b} from a function of the matrix $\mathbf{A} = \Phi^*(\mathbf{Y}^* \mathbf{Y} - \sigma_w^2 (M-1) \mathbf{L} \mathbf{I}_{MK}) \Phi$ and previous estimates. Propositions 8 and 9 show the convergence of the iterations in Algorithm 1 that alternately update the estimates $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ under the randomness assumptions in (A1) and (A2). Similarly to the analysis of the spectral initialization in Section VII, we prove Propositions 8 and 9 by using the Davis-Kahan Theorem in Lemma 11. To this end, we first compute tail estimates of norms of the deviation of the random matrix \mathbf{A} from its expectation $\underline{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$ below.

A. Tail Estimates of Deviations

Algorithm 1 updates the estimates $\hat{\mathbf{a}}$ as the least dominant eigenvector of $(\mathbf{I}_M \otimes \hat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \hat{\mathbf{b}})$ where $\hat{\mathbf{b}}$ denotes the estimate in the previous step. The other estimate $\hat{\mathbf{b}}$ is updated similarly from $(\hat{\mathbf{a}}^* \otimes \mathbf{I}_D) \mathbf{A} (\hat{\mathbf{a}} \otimes \mathbf{I}_D)$. The matrices involved

in these updates are restricted version of \mathbf{A} with separable projection operators.

In order to get a tightened perturbation bound for the estimates, we introduce a new matrix norm with this separability structure. To define the new norm, we need operators that rearrange an M -by- D matrix into a column vector of length MD and vice versa. For $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_M] \in \mathbb{C}^{M \times D}$, define

$$\text{vec}(\mathbf{V}) = [\mathbf{v}_1^\top, \dots, \mathbf{v}_M^\top]^\top.$$

Let $\text{mat}(\cdot)$ denote the inverse of $\text{vec}(\cdot)$ so that

$$\text{mat}(\text{vec}(\mathbf{V})) = \mathbf{V}, \quad \forall \mathbf{V} \in \mathbb{C}^{M \times D}$$

and

$$\text{vec}(\text{mat}(\mathbf{v})) = \mathbf{v}, \quad \forall \mathbf{v} \in \mathbb{C}^{MD}.$$

With these vectorization and matricization operators, we define the matricized S_p -norm of $\mathbf{v} \in \mathbb{C}^{MD}$ by

$$||| \mathbf{v} |||_{S_p} = \|\text{mat}(\mathbf{v})\|_{S_p}.$$

Then the matricized operator norm of $\mathbf{M} \in \mathbb{C}^{MD \times MD}$ is defined by

$$||| \mathbf{M} |||_{S_p \rightarrow S_q} := \max_{||| \mathbf{v} |||_{S_p} \leq 1} ||| \mathbf{M} \mathbf{v} |||_{S_q}.$$

For $p = 1$ and $q = \infty$, by the Courant-Fischer minimax principle, the matricized operator norm is written as a variational form, i.e. $||| \mathbf{M} |||_{S_1 \rightarrow S_\infty}$ is a solution to

$$\begin{aligned} & \underset{\mathbf{Y}, \mathbf{Y}' \in \mathbb{C}^{M \times D}}{\text{maximize}} \quad |\langle \text{vec}(\mathbf{Y}'), \mathbf{M} \text{vec}(\mathbf{Y}) \rangle| \\ & \text{subject to} \quad \|\mathbf{Y}\|_{S_1} \leq 1, \quad \|\mathbf{Y}'\|_{S_1} \leq 1. \end{aligned}$$

Since the unit ball with respect to the S_1 -norm is given as the convex hull of all unit- S_2 -norm matrices of rank-1, $||| \mathbf{M} |||_{S_1 \rightarrow S_\infty}$ is a solution to

$$\begin{aligned} & \underset{\mathbf{Y}, \mathbf{Y}' \in \mathbb{C}^{M \times D}}{\text{maximize}} \quad |\langle \text{vec}(\mathbf{Y}'), \mathbf{M} \text{vec}(\mathbf{Y}) \rangle| \\ & \text{subject to} \quad \|\mathbf{Y}\|_{S_2} \leq 1, \quad \|\mathbf{Y}'\|_{S_2} \leq 1 \\ & \quad \text{rank}(\mathbf{Y}) = \text{rank}(\mathbf{Y}') = 1. \end{aligned} \quad (61)$$

Therefore, by dropping the rank-1 constraints in (61), we obtain

$$||| \mathbf{M} |||_{S_1 \rightarrow S_\infty} \leq \|\mathbf{M}\|, \quad \forall \mathbf{M} \in \mathbb{C}^{MD \times MD}. \quad (62)$$

The following lemma provides a tail estimate of $||| \mathbf{E} |||_{S_1 \rightarrow S_\infty}$ divided by $K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2$, which amounts to the spectral gap between the two smallest eigenvalues of $\underline{\mathbf{A}}$. Compared to the analogous tail estimate for its spectral norm, derived in [22, Sec. 3.2], the tail estimate for $||| \mathbf{E} |||_{S_1 \rightarrow S_\infty}$ is smaller in order. This is the reason why we obtain a better sample complexity by introducing the extra rank-1 structure to the prior model on impulse responses.

Lemma 18: Let $\mathbf{E} = \mathbf{A} - \underline{\mathbf{A}}$. For any $\beta \in \mathbb{N}$, there exist a numerical constant C and a constant $C(\beta)$ that depends only

on β such that

$$\frac{||| \mathbf{E} |||_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2} \leq C(\beta) \log^\alpha(MKL) \cdot \left(\left(\frac{1}{\sqrt{M}} + \sqrt{\frac{D}{K}} \right) \mu^2 + \frac{\sqrt{D}}{\eta \sqrt{L}} + \frac{\rho_{x,w}}{\sqrt{\eta K L} \sigma_w \|\mathbf{x}\|_2} \left[\mu \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{M}} + \sqrt{\frac{D}{K}} \right) + 1 \right] \right) \quad (63)$$

holds with probability $1 - CK^{-\beta}$.

Proof of Lemma 18: The derivation of (63) is similar to that for the analogous tail estimate for $\|\mathbf{E}\|$ in [22, Sec. 3.2]. We use the same decomposition of \mathbf{E} , which is briefly summarized below.

We decompose \mathbf{Y} as $\mathbf{Y} = \mathbf{Y}_s + \mathbf{Y}_n$, where the noise-free portion \mathbf{Y}_s (resp. the noise portion \mathbf{Y}_n) is obtained as we replace $\mathbf{y}_m = \mathbf{h}_m \otimes \mathbf{x} + \mathbf{w}_m$ in \mathbf{Y} by its first summand $\mathbf{h}_m \otimes \mathbf{x}$ (resp. by its second summand \mathbf{w}_m) for all $m = 1, \dots, M$. Then \mathbf{E} is written as the sum of three matrices whose entries are given as polynomials of subgaussian random variables of different order as follows.

$$\mathbf{E} = \Phi^* \mathbf{Y}_s^* \mathbf{Y}_s \Phi - \mathbb{E}[\Phi^* \mathbf{Y}_s^* \mathbf{Y}_s \Phi] \quad (64)$$

$$+ \Phi^* \mathbf{Y}_s^* \mathbf{Y}_n \Phi + \Phi^* \mathbf{Y}_n^* \mathbf{Y}_s \Phi \quad (65)$$

$$+ \Phi^* (\mathbf{Y}_n^* \mathbf{Y}_n - \sigma_w^2 (M-1) \mathbf{L} \mathbf{I}_{MK}) \Phi. \quad (66)$$

We first compute tail estimates of the components; the tail estimate in (63) is then obtained by combining these results via the triangle inequality.

For the first summand in (64) and the last summand in (66), we were not able to reduce their tail estimates in order compared to the spectral norms. Thus we use their tail estimates on the spectral norms derived in [22, Sec. 3.2], which are also valid tail estimates by (62). For the completeness, we provide the corresponding lemmas below.

Lemma 19 [22, Lemma 3.5]: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there exist a numerical constant $\alpha \in \mathbb{N}$ and a constant $C(\beta)$ that depends only on β such that

$$\frac{||| \Phi^* \mathbf{Y}_s^* \mathbf{Y}_s \Phi - \mathbb{E}[\Phi^* \mathbf{Y}_s^* \mathbf{Y}_s \Phi] |||_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2} \leq C(\beta) \log^\alpha(MKL) \left(\sqrt{\frac{1}{M}} + \sqrt{\frac{D}{K}} \right) \mu^2 \quad (67)$$

holds with probability $1 - CK^{-\beta}$.

Lemma 20 [22, Lemma 3.7]: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that

$$\frac{||| \Phi^* (\mathbf{Y}_n^* \mathbf{Y}_n - \sigma_w^2 (M-1) \mathbf{L} \mathbf{I}_{MK}) \Phi |||_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2} \leq \frac{C(\beta) \log^\alpha(MKL)}{\eta} \cdot \sqrt{\frac{D}{L}} \quad (68)$$

with probability $1 - CK^{-\beta}$.

For the second and third terms in (65), we use their tail estimates given in the following lemma, the proof of which is provided in Appendix X.

Lemma 21: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there exists a constant $C(\beta)$ that depends only on β such that, conditional on the noise vector \mathbf{w} ,

$$\frac{||| \Phi^* \mathbf{Y}_s^* \mathbf{Y}_n \Phi |||_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2} \leq \frac{C(\beta) \rho_{x,w}}{\sqrt{\eta K L} \sigma_w \|\mathbf{x}\|_2} \cdot \left[\mu \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{M}} + \sqrt{\frac{D}{K}} \right) + 1 \right] \quad (69)$$

holds with probability $1 - CK^{-\beta}$.

Finally, the tail estimate in (63) is obtained by combining (67), (67), and (69) via the triangle inequality. This completes the proof. \square

We will also make use of a tail estimate of $||| \mathbf{E} \mathbf{u} |||_{S_\infty} / \|\mathbf{u}\|_2$, again normalized by factor $K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2$. The following lemma, which provides a relevant tail estimate, is a direct consequence of Lemma 21 and [22, Lemma 3.8].

Lemma 22: Let $\mathbf{E} = \mathbf{A} - \underline{\mathbf{A}}$. For any $\beta \in \mathbb{N}$, there exist a numerical constant C and a constant $C(\beta)$ that depends only on β such that

$$\frac{||| \mathbf{E} \mathbf{u} |||_{S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^3} \leq C(\beta) \log^\alpha(MKL) \left(\frac{\rho_{x,w}}{\sqrt{\eta K L} \sigma_w \|\mathbf{x}\|_2} \cdot \left[\mu \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{M}} + \sqrt{\frac{D}{K}} \right) + 1 \right] + \frac{\sqrt{D}}{\eta \sqrt{ML}} \right) \quad (70)$$

holds with probability $1 - CK^{-\beta}$.

B. Proof of Proposition 8

To simplify notations, let $\theta = \angle(\mathbf{b}, \hat{\mathbf{b}})$ denote the principal angle between the two subspaces spanned respectively by \mathbf{b} and $\hat{\mathbf{b}}$, i.e., $\theta \in [0, \pi/2]$ satisfies

$$\sin \theta = \|\mathbf{P}_{\mathbf{b}^\perp} \hat{\mathbf{b}}\|_2, \quad \cos \theta = \|\mathbf{P}_{\mathbf{b}} \hat{\mathbf{b}}\|_2,$$

where $\mathbf{P}_{\mathbf{b}}$ denotes the orthogonal projection onto the span of \mathbf{b} . The assumption in (34) implies $\theta \leq \pi/4$.

Recall that Algorithm 1 updates $\hat{\mathbf{a}}$ from a given estimate $\hat{\mathbf{b}}$ in the previous step as the eigenvector of the matrix $(\mathbf{I}_M \otimes \hat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \hat{\mathbf{b}})$ corresponding to the smallest eigenvalue. Without loss of generality, we may assume that $\|\hat{\mathbf{b}}\|_2 = 1$.

By direct calculation, we obtain that $\underline{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$ is rewritten as

$$\underline{\mathbf{A}} = K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 (\|\mathbf{a}\|_2^2 \mathbf{I}_M - \text{diag}(|\mathbf{a}|^2)) \otimes \mathbf{P}_{\mathbf{b}^\perp} + K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 (\|\mathbf{a}\|_2^2 \mathbf{I}_M - \mathbf{a} \mathbf{a}^*) \otimes \mathbf{P}_{\mathbf{b}}. \quad (71)$$

Then

$$\begin{aligned} & (\mathbf{I}_M \otimes \hat{\mathbf{b}}^*) \underline{\mathbf{A}} (\mathbf{I}_M \otimes \hat{\mathbf{b}}) \\ &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 (\|\mathbf{a}\|_2^2 \mathbf{I}_M - \cos^2 \theta \mathbf{a} \mathbf{a}^*) \\ & \quad - K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \sin^2 \theta \text{diag}(|\mathbf{a}|^2). \end{aligned} \quad (72)$$

Here $|\mathbf{a}|^2$ denotes the vector whose k th entry is the squared magnitude of the k th entry of \mathbf{a} and $\text{diag}(|\mathbf{a}|^2)$ is a diagonal matrix whose diagonal entries are given by $|\mathbf{a}|^2$.

We verify that the matrix $\|\mathbf{a}\|_2^2 \mathbf{I}_M - \cos^2 \theta \mathbf{a} \mathbf{a}^*$ is positive definite and its smallest eigenvalue, which has multiplicity 1, is smaller than the next smallest eigenvalue by $\|\mathbf{a}\|_2^2 \cos^2 \theta$. Furthermore, \mathbf{a} is collinear with the eigenvector corresponding to the smallest eigenvalue.

Let us consider the following matrix:

$$\begin{aligned} & K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \mathbf{I}_M - (\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \widehat{\mathbf{b}}) \\ &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \cos^2 \theta \mathbf{a} \mathbf{a}^* \\ &+ K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \sin^2 \theta \text{diag}(|\mathbf{a}|^2) \\ &- (\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{E} (\mathbf{I}_M \otimes \widehat{\mathbf{b}}), \end{aligned}$$

which we considered as a perturbed version of $K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \cos^2 \theta \mathbf{a} \mathbf{a}^*$. Then the perturbation, that is the difference of the two matrices, satisfies

$$\begin{aligned} & \left\| K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \mathbf{I}_M - (\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \widehat{\mathbf{b}}) \right. \\ & \quad \left. - K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \cos^2 \theta \mathbf{a} \mathbf{a}^* \right\| \\ & \leq \left\| K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \sin^2 \theta \text{diag}(|\mathbf{a}|^2) \right\| \\ & \quad + \|(\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{E} (\mathbf{I}_M \otimes \widehat{\mathbf{b}})\| \\ & \leq K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_\infty^2 \sin^2 \theta + \|\mathbf{E}\|_{S_1 \rightarrow S_\infty}. \quad (73) \end{aligned}$$

For sufficiently large $C_1(\beta)$, the conditions in (34), (35), (36), and (37) imply

$$\begin{aligned} & K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \cos^2 \theta \\ & > 2(K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_\infty^2 \sin^2 \theta + \|\mathbf{E}\|_{S_1 \rightarrow S_\infty}). \end{aligned}$$

Therefore, $\widehat{\mathbf{a}}$ is a unique dominant eigenvector of $K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \mathbf{I}_M - (\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \widehat{\mathbf{b}})$.

Next we apply Lemma 11 for

$$\begin{aligned} \underline{\mathbf{M}} &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \cos^2 \theta \mathbf{a} \mathbf{a}^*, \\ \mathbf{M} &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \mathbf{I}_M \\ &\quad - (\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{A} (\mathbf{I}_M \otimes \widehat{\mathbf{b}}), \\ \mathbf{q}_1 &= \frac{\mathbf{a}}{\|\mathbf{a}\|_2}, \\ \tilde{\mathbf{q}} &= \widehat{\mathbf{a}}. \end{aligned}$$

Then λ and \mathbf{D} in Lemma 11 are given as $\lambda = K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \cos^2 \theta$ and $\mathbf{D} = \mathbf{0}$.

By (73), we have

$$\begin{aligned} \frac{\|\mathbf{M} - \underline{\mathbf{M}}\|}{\lambda} &\leq \frac{\|\mathbf{a}\|_\infty^2 \sin^2 \theta}{\|\mathbf{a}\|_2^2 \cos^2 \theta} + \frac{\|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \cos^2 \theta} \\ &\leq \frac{\mu^2}{M} + \frac{2\|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2}, \end{aligned}$$

where the last step follows from (34). Therefore, for sufficiently large $C_1(\beta)$, the conditions in (35), (36), (37) combined with Lemma 18 satisfy (45) in Lemma 11 and we obtain the error bound in (46).

It remains to compute $\|(\mathbf{M} - \underline{\mathbf{M}})\mathbf{q}_1\|_2/\lambda$. The ℓ_2 -norm of $(\mathbf{M} - \underline{\mathbf{M}})\mathbf{q}_1$ satisfies

$$\begin{aligned} \|(\mathbf{M} - \underline{\mathbf{M}})\mathbf{q}_1\|_2 &\leq \frac{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \sin^2 \theta \|\text{diag}(|\mathbf{a}|^2)\mathbf{a}\|_2}{\|\mathbf{a}\|_2} \\ &\quad + \frac{\|(\mathbf{I}_M \otimes \widehat{\mathbf{b}}^*) \mathbf{E} (\mathbf{I}_M \otimes \widehat{\mathbf{b}})\|_2}{\|\mathbf{a}\|_2} \\ &\leq K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_\infty^2 \sin^2 \theta \\ &\quad + 3 \sin \theta \|\mathbf{E}\|_{S_1 \rightarrow S_\infty} \\ &\quad + \frac{\cos^2 \theta \|\mathbf{E} (\mathbf{I}_M \otimes \widehat{\mathbf{b}})\|_{S_\infty}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}, \quad (74) \end{aligned}$$

where the second step follow from the decomposition of $\widehat{\mathbf{b}}$ given by

$$\widehat{\mathbf{b}} = \mathbf{P}_b \widehat{\mathbf{b}} + \mathbf{P}_{b^\perp} \widehat{\mathbf{b}}.$$

which satisfies $\|\mathbf{P}_b \widehat{\mathbf{b}}\|_2 = \cos \theta$ and $\|\mathbf{P}_{b^\perp} \widehat{\mathbf{b}}\|_2 = \sin \theta$. By dividing the right-hand side of (74) by λ , we obtain

$$\begin{aligned} & \frac{4\|(\mathbf{M} - \underline{\mathbf{M}})\mathbf{q}_1\|_2}{\lambda} \\ & \leq \frac{4\|\mathbf{a}\|_\infty^2 \sin^2 \theta}{\|\mathbf{a}\|_2^2 \cos^2 \theta} + \frac{12 \sin^2 \theta \|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \cos^2 \theta} \\ & \quad + \frac{4\|\mathbf{E} \mathbf{u}\|_{S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^3} \\ & \leq \left(\frac{8\mu^2}{M} + \frac{24\|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2} \right) \sin \theta + \frac{4\|\mathbf{E} \mathbf{u}\|_{S_\infty}}{K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^3}, \quad (75) \end{aligned}$$

where the second step follows from (34).

By Lemma 18, the constant factor for $\sin \theta$ in (75) becomes less than 1/2 as we choose $C_1(\beta)$ in (35), (36), (37) sufficiently large. This gives (38), where the expression for κ follows from Lemma 22. This completes the proof.

C. Proof of Proposition 9

The proof of Proposition 9 is similar to that of Proposition 8. Thus we will only highlight the differences between the two proofs.

Without loss of generality, we assume that $\|\widehat{\mathbf{a}}\|_2 = 1$. Let $\check{\theta} = \angle(\mathbf{a}, \widehat{\mathbf{a}})$. The assumption in (39) implies $\check{\theta} \leq \pi/4$. This time, we compute the least dominant eigenvector of $(\widehat{\mathbf{a}}^* \otimes \mathbf{I}_D) \mathbf{A} (\widehat{\mathbf{a}} \otimes \mathbf{I}_D)$. From (71), we obtain

$$\begin{aligned} & (\widehat{\mathbf{a}}^* \otimes \mathbf{I}_D) \mathbf{A} (\widehat{\mathbf{a}} \otimes \mathbf{I}_D) \\ &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{a}\|_2^2 (\|\mathbf{b}\|_2^2 \mathbf{I}_D - \cos^2 \check{\theta} \mathbf{b} \mathbf{b}^*) \\ &\quad - K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\| \odot \widehat{\mathbf{a}} \otimes \mathbf{P}_{b^\perp}. \quad (76) \end{aligned}$$

We consider the matrix

$$\begin{aligned} & K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2 \mathbf{I}_D - (\widehat{\mathbf{a}}^* \otimes \mathbf{I}_D) \mathbf{A} (\widehat{\mathbf{a}} \otimes \mathbf{I}_D) \\ &= K^2 \|\mathbf{x}\|_2^2 \|\mathbf{a}\|_2^2 \cos^2 \check{\theta} \mathbf{b} \mathbf{b}^* \\ &\quad + K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\| \odot \widehat{\mathbf{a}} \otimes \mathbf{P}_{b^\perp} \\ &\quad - (\widehat{\mathbf{a}}^* \otimes \mathbf{I}_D) \mathbf{E} (\widehat{\mathbf{a}} \otimes \mathbf{I}_D) \end{aligned}$$

as a perturbed version of $K^2\|\mathbf{x}\|_2^2\|\mathbf{a}\|_2^2\cos^2\check{\theta}\mathbf{b}\mathbf{b}^*$. The difference of the two matrices satisfies

$$\begin{aligned} & \left\| K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\mathbf{I}_D - (\hat{\mathbf{a}}^* \otimes \mathbf{I}_D)\mathbf{A}(\hat{\mathbf{a}} \otimes \mathbf{I}_D) \right. \\ & \quad \left. - K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\cos^2\check{\theta}\mathbf{P}_b \right\| \\ & \leq \left\| K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\odot\hat{\mathbf{a}}\|\mathbf{P}_{b^\perp}\| \right. \\ & \quad \left. + \|(\hat{\mathbf{a}}^* \otimes \mathbf{I}_D)\mathbf{E}(\hat{\mathbf{a}} \otimes \mathbf{I}_D)\| \right\| \\ & \leq K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_\infty^2 + \|\mathbf{E}\|_{S_1 \rightarrow S_\infty}. \end{aligned}$$

For sufficiently large $C_1(\beta)$, the conditions in (34), (35), (36), and (37) imply

$$\begin{aligned} & K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\cos^2\theta \\ & > 2(K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_\infty^2\sin^2\theta + \|\mathbf{E}\|_{S_1 \rightarrow S_\infty}). \end{aligned}$$

Therefore, $\hat{\mathbf{b}}$ is also a unique dominant eigenvector of $K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\mathbf{I}_D - (\hat{\mathbf{a}}^* \otimes \mathbf{I}_D)\mathbf{A}(\hat{\mathbf{a}} \otimes \mathbf{I}_D)$.

Next we apply Lemma 11 for

$$\begin{aligned} \underline{\mathbf{M}} &= K^2\|\mathbf{x}\|_2^2\|\mathbf{a}\|_2^2\cos^2\check{\theta}\mathbf{b}\mathbf{b}^*, \\ \mathbf{M} &= K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\mathbf{I}_D \\ & \quad - (\hat{\mathbf{a}}^* \otimes \mathbf{I}_D)\mathbf{A}(\hat{\mathbf{a}} \otimes \mathbf{I}_D), \\ \mathbf{q}_1 &= \frac{\mathbf{b}}{\|\mathbf{b}\|_2}, \\ \tilde{\mathbf{q}} &= \hat{\mathbf{b}}. \end{aligned}$$

Then λ and \mathbf{D} in Lemma 11 are given as $\lambda = K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2\cos^2\check{\theta}$ and $\mathbf{D} = \mathbf{0}$.

Similarly to the proof of Proposition 8, we show

$$\frac{\|\mathbf{M} - \underline{\mathbf{M}}\|}{\lambda} \leq \frac{2\mu^2}{M} + \frac{2\|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2\|\mathbf{x}\|_2^2\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2}$$

and

$$\frac{4\|(\mathbf{M} - \underline{\mathbf{M}})\mathbf{q}_1\|_2}{\lambda} \leq \frac{24\|\mathbf{E}\|_{S_1 \rightarrow S_\infty}}{K^2\|\mathbf{x}\|_2^2\|\mathbf{u}\|_2^2}\sin\theta + \frac{4\|\mathbf{E}\mathbf{u}\|_{S_\infty}}{K^2\|\mathbf{x}\|_2^2\|\mathbf{u}\|_2^3}.$$

Here we used the decomposition of $\hat{\mathbf{a}}$ given by

$$\hat{\mathbf{a}} = \mathbf{P}_a\hat{\mathbf{a}} + \mathbf{P}_{a^\perp}\hat{\mathbf{a}},$$

which satisfies $\|\mathbf{P}_a\hat{\mathbf{a}}\|_2 = \cos\check{\theta}$ and $\|\mathbf{P}_{a^\perp}\hat{\mathbf{a}}\|_2 = \sin\check{\theta}$.

The remaining steps are identical to those in the proof of Proposition 8 and we omit further details.

IX. CONVERGENCE OF RANK-1 TRUNCATED POWER METHOD

In this section, we prove Proposition 10. First we present a theorem that shows local convergence of the rank-1 truncated

power method for general matrix input \mathbf{B} . Then we will show the proof of Proposition 10 as its corollary.

The separability structure in (9) corresponds to the rank-1 structure when the eigenvector is rearranged as a matrix. We introduce a collection of structured subspaces, where their Minkowski sum is analogous to the support in the sparsity model. For $(\mathbf{a}, \mathbf{b}) \in \mathbb{C}^M \times \mathbb{C}^D$, we define

$$T(\mathbf{a}, \mathbf{b}) := \{\mathbf{a} \otimes \xi + \mathbf{q} \otimes \mathbf{b} \mid \xi \in \mathbb{C}^D, \mathbf{q} \in \mathbb{C}^M\}.$$

Then

$$\text{mat}(T(\mathbf{a}, \mathbf{b})) = \{\text{mat}(\mathbf{v}) \mid \mathbf{v} \in T(\mathbf{a}, \mathbf{b})\}$$

is equivalent to the tangent space of the rank-1 matrix $\mathbf{U} = \mathbf{a}\mathbf{b}^\top$.

Now we state a local convergence result for the rank-1 truncated power method in the following theorem, the proof of which is postponed to Section IX-A.

Theorem 23: Let $\mathbf{u} = \mathbf{a} \otimes \mathbf{b}$ be a unique dominant eigenvector of $\underline{\mathbf{B}}$. Let $\tilde{\lambda}_2(\underline{\mathbf{B}})$ be defined in (77), as shown at the bottom of this page. Suppose that there exist $0 < \mu < 1$ and $0 < \tau < \frac{1}{3\sqrt{2}}$, for which (78), as shown at the bottom of this page

$$\frac{4\sqrt{6}\|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty}}{\lambda_1(\underline{\mathbf{B}})} \leq \min\left[\frac{1}{3\sqrt{2}}, \frac{(1-\mu)\tau}{1+\mu}\right], \quad (79)$$

and

$$\tilde{\lambda}_2(\underline{\mathbf{B}}) + 6\|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty} \leq \frac{\lambda_1(\underline{\mathbf{B}})}{5} \quad (80)$$

hold. If $\sin\angle(\mathbf{u}_0, \mathbf{u}) \leq \tau$, then $(\mathbf{u}_t)_{t \in \mathbb{N}}$ produced by Algorithm 2 satisfies

$$\begin{aligned} \sin\angle(\mathbf{u}_t, \mathbf{u}) &\leq \mu \sin\angle(\mathbf{u}_{t-1}, \mathbf{u}) \\ &\quad + \frac{(1+\mu)4\sqrt{6}\|(\mathbf{B} - \underline{\mathbf{B}})\mathbf{u}\|_{S_\infty}}{\lambda_1(\underline{\mathbf{B}})} \end{aligned} \quad (81)$$

for all $t \in \mathbb{N}$.

Proposition 10 is a direct consequence of Theorem 23 for the case where the input matrix \mathbf{B} is given as $\mathbf{B} = \|\mathbb{E}[\mathbf{A}]\| \mathbf{I}_{MD} - \mathbf{A}$ and its proof is presented below.

Proof of Proposition 10: We apply Theorem 23 for

$$\mathbf{B} = \|\mathbb{E}[\mathbf{A}]\| \mathbf{I}_{MD} - \mathbf{A}$$

and

$$\underline{\mathbf{B}} = K^2\|\mathbf{x}\|_2^2\mathbf{u}\mathbf{u}^*.$$

Then the difference between \mathbf{B} and $\underline{\mathbf{B}}$ is written as

$$\mathbf{B} - \underline{\mathbf{B}} = (\|\mathbb{E}[\mathbf{A}]\| - K^2\|\mathbf{x}\|_2^2\|\mathbf{u}\|_2^2)\mathbf{I}_{MD} + K^2\|\mathbf{x}\|_2^2\mathbf{Y} - \mathbf{E}. \quad (82)$$

$$\tilde{\lambda}_2(\underline{\mathbf{B}}) := \sup_{\mathbf{v}, (\hat{\mathbf{a}}, \hat{\mathbf{b}}), (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})} \left\{ \mathbf{v}^* \underline{\mathbf{B}} \mathbf{v} \mid \|\mathbf{v}\|_2 \leq 1, \mathbf{v} \in \mathbf{u}^\perp \cap [T(\mathbf{a}, \mathbf{b}) + T(\hat{\mathbf{a}}, \hat{\mathbf{b}}) + T(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})] \right\}. \quad (77)$$

$$\frac{\sqrt{5}(\tilde{\lambda}_2(\underline{\mathbf{B}}) + 6\|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty})}{\sqrt{1-\tau^2}\lambda_1(\underline{\mathbf{B}}) - \tau\tilde{\lambda}_2(\underline{\mathbf{B}}) - 6(\sqrt{1-\tau^2} + \tau)\|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty}} < \mu, \quad (78)$$

In Section VIII-B, we have computed $\underline{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$ in (71), which is rewritten as

$$\underline{\mathbf{A}} = K^2 \|\mathbf{x}\|_2^2 (\|\mathbf{u}\|_2^2 \mathbf{P}_{\mathbf{u}^\perp} - \mathbf{\Upsilon}) \quad (83)$$

with

$$\mathbf{\Upsilon} = \text{diag}(|\mathbf{a}|^2) \otimes \|\mathbf{b}\|_2^2 \mathbf{P}_{\mathbf{b}^\perp},$$

where $\mathbf{u} = \mathbf{a} \otimes \mathbf{b}$.

Therefore, it follows from (83) that

$$\begin{aligned} |\|\underline{\mathbf{A}}\| - K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}\|_2^2| &\leq K^2 \|\mathbf{x}\|_2^2 \|\mathbf{\Upsilon}\| \\ &\leq K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_\infty^2. \end{aligned} \quad (84)$$

Then by plugging in (84) to (82), we obtain

$$|\|\mathbf{B} - \underline{\mathbf{B}}\||_{S_1 \rightarrow S_\infty} \leq 2K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_\infty^2 + |\|\mathbf{E}\||_{S_1 \rightarrow S_\infty} \quad (85)$$

On the other hand, $\underline{\mathbf{B}}$ is a rank-1 matrix whose eigenvector is collinear with \mathbf{u} and the largest eigenvalue is given by

$$\lambda_1(\underline{\mathbf{B}}) = K^2 \|\mathbf{x}\|_2^2 \|\mathbf{b}\|_2^2 \|\mathbf{a}\|_2^2. \quad (86)$$

Therefore, $\underline{\mathbf{B}}$ also satisfies

$$\tilde{\lambda}_2(\underline{\mathbf{B}}) = 0.$$

Since $\tilde{\lambda}_2(\underline{\mathbf{B}}) = 0$, (78) and (79) are implied by

$$\frac{|\|\mathbf{B} - \underline{\mathbf{B}}\||_{S_1 \rightarrow S_\infty}}{\lambda_1(\underline{\mathbf{B}})} \leq C_0 \min \left[\mu \sqrt{1 - \tau^2}, \frac{(1 - \mu)\tau}{1 + \mu} \right] \quad (87)$$

for a numerical constant C_0 .

By applying (86) and the tail estimate of $|\|\mathbf{E}\||_{S_1 \rightarrow S_\infty}$ given in Lemma 18 to (85), we verify that the sufficient condition in (87) is implied by (35), (36), and (37) for $C_1 = c(\mu, \tau)C'_1$, $C_2 = c(\mu, \tau)C'_2$ where C'_1 and C'_2 are constants that only depend on β .

Since the conditions in (78) and (79) are satisfied, Theorem 23 provides the error bound in (40). This completes the proof. \square

A. Proof of Theorem 23

In order to prove Theorem 23, we first provide lemmas, which show upper bounds on the estimation error, given in terms of the principal angle, in the corresponding steps of Algorithm 2.

The first lemma provides upper bounds on norms of a matrix and a vector when they are restricted with a projection operator onto a subspace with the separability structure.

Lemma 24: Let

$$\check{T} = \sum_{k=1}^r T(\mathbf{a}_k, \mathbf{b}_k)$$

for $\{(\mathbf{a}_k, \mathbf{b}_k)\}_{k=1}^r \subset \mathbb{C}^M \times \mathbb{C}^D$, $\mathbf{M} \in \mathbb{C}^{MD \times MD}$, and $\mathbf{u} \in \mathbb{C}^{MD}$. Then

$$\|\mathbf{P}_{\check{T}} \mathbf{M} \mathbf{P}_{\check{T}}\| \leq 2r |\|\mathbf{M}\||_{S_1 \rightarrow S_\infty}$$

and

$$\|\mathbf{P}_{\check{T}} \mathbf{M} \mathbf{u}\|_2 \leq \sqrt{2r} |\|\mathbf{M} \mathbf{u}\||_{S_\infty}.$$

Proof: Let $\mathbf{v} \in \check{T}$. Then $\text{rank}(\text{mat}(\mathbf{v})) \leq 2r$. Let

$$\text{mat}(\mathbf{v}) = \sum_{l=1}^{2r} \sigma_l \mathbf{q}_l \xi_l^\top$$

denotes the singular value decomposition of $\text{mat}(\mathbf{v})$, where $\|\mathbf{q}_l\|_2 = \|\xi_l\|_2 = 1$ and $\sigma_l \geq 0$ for $k = 1, \dots, 2r$. Then

$$\mathbf{v} = \sum_{l=1}^{2r} \sigma_l \mathbf{q}_l \otimes \xi_l.$$

Similarly, we can represent $\mathbf{v}' \in \check{T}$ as

$$\mathbf{v}' = \sum_{j=1}^{2r} \sigma'_j \mathbf{q}'_j \otimes \xi'_j.$$

Then

$$\begin{aligned} |\langle \mathbf{v}', \mathbf{M} \mathbf{v} \rangle| &\leq \sum_{j,l=1}^{2r} \sigma_l \sigma'_j |\langle \mathbf{q}'_j \otimes \xi'_j, \mathbf{M}(\mathbf{q}_l \otimes \xi_l) \rangle| \\ &\leq \sum_{l=1}^{2r} \sigma_l \sum_{j=1}^{2r} \sigma'_j |\|\mathbf{M}\||_{S_1 \rightarrow S_\infty} \\ &\leq 2r \|\mathbf{v}\|_2 \|\mathbf{v}'\|_2 |\|\mathbf{M}\||_{S_1 \rightarrow S_\infty}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|\mathbf{P}_{\check{T}} \mathbf{M} \mathbf{P}_{\check{T}}\| &= \sup_{\mathbf{v}, \mathbf{v}' \in \check{T}} \{|\langle \mathbf{v}', \mathbf{M} \mathbf{v} \rangle| \mid \|\mathbf{v}\|_2 = \|\mathbf{v}'\|_2 = 1\} \\ &\leq 2r |\|\mathbf{M}\||_{S_1 \rightarrow S_\infty}. \end{aligned}$$

This proves the first assertion. The second assertion is obtained in a similar way by fixing $\mathbf{v} = \mathbf{u}$. \square

The following lemma is a direct consequence of the Davis-Kahan Theorem together with Lemma 24.

Lemma 25 (Perturbation): Let $\{(\mathbf{a}_k, \mathbf{b}_k)\}_{k=1}^r \subset \mathbb{C}^M \times \mathbb{C}^D$ satisfy

$$T(\mathbf{a}, \mathbf{b}) \subset \sum_{k=1}^r T(\mathbf{a}_k, \mathbf{b}_k) =: \check{T}.$$

Let \mathbf{v} (resp. \mathbf{u}) be a unique most dominant eigenvector of $\mathbf{P}_{\check{T}} \mathbf{M}_1 \mathbf{P}_{\check{T}}$ (resp. $\mathbf{P}_{\check{T}} \mathbf{M}_2 \mathbf{P}_{\check{T}}$). If

$$\begin{aligned} \lambda_2(\mathbf{P}_{\check{T}} \mathbf{M}_2 \mathbf{P}_{\check{T}}) + 2r |\|\mathbf{M}_1 - \mathbf{M}_2\||_{S_1 \rightarrow S_\infty} \\ \leq \frac{\lambda_1(\mathbf{P}_{\check{T}} \mathbf{M}_2 \mathbf{P}_{\check{T}})}{5}, \end{aligned} \quad (88)$$

then

$$\sin \angle(\mathbf{v}, \mathbf{u}) \leq \frac{4\sqrt{2r} |\|\mathbf{M}_1 - \mathbf{M}_2\||_{S_\infty}}{\lambda_1(\mathbf{P}_{\check{T}} \mathbf{M}_2 \mathbf{P}_{\check{T}})}.$$

The following lemma shows how the conventional power method converges depending on the largest and second largest eigenvalues.

Lemma 26 (A Single Iteration of Power Method [56, Th. 1.1]): Let \mathbf{M} have a unique dominant eigenvector \mathbf{v} . Then

$$\sin \angle(\mathbf{M} \hat{\mathbf{v}}, \mathbf{v}) \leq \frac{\lambda_2(\mathbf{M}) \sin \angle(\hat{\mathbf{v}}, \mathbf{v})}{\lambda_1(\mathbf{M}) \cos \angle(\hat{\mathbf{v}}, \mathbf{v}) - \lambda_2(\mathbf{M}) \sin \angle(\hat{\mathbf{v}}, \mathbf{v})}$$

for any $\hat{\mathbf{v}}$ such that $\langle \hat{\mathbf{v}}, \mathbf{v} \rangle \neq 0$.

The following lemma is a modification of [47, Lemma 12] and shows that the correlation is partially preserved after the rank-1 truncation. Unlike the canonical sparsity model, where the atoms are mutually orthogonal, in the low-rank atomic model, atoms in an atomic decomposition may have correlation. Our proof addresses this general case and the argument here also applies to an abstract atomic model.

Lemma 27 (Correlation after the Rank-1 Truncation): Let $\check{\mathbf{v}} \in \mathbb{C}^{MD}$ satisfy $\|\check{\mathbf{v}}\|_2 = 1$ and $\text{rank}(\text{mat}(\check{\mathbf{v}})) = 1$. For $\mathbf{v} \in \mathbb{C}^{MD}$ such that $\|\mathbf{v}\|_2 = 1$, let $\widehat{\mathbf{V}} \in \mathbb{C}^{M \times D}$ denote the best rank-1 approximation of $\mathbf{V} = \text{mat}(\mathbf{v})$ and $\widehat{\mathbf{v}} = \text{vec}(\widehat{\mathbf{V}})$. Then

$$|\langle \widehat{\mathbf{v}}, \check{\mathbf{v}} \rangle| \geq |\langle \mathbf{v}, \check{\mathbf{v}} \rangle| - \min \left(\sqrt{1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2}, 2(1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2) \right). \quad (89)$$

Proof of Lemma 27: There exist $\check{\mathbf{a}} \in \mathbb{C}^M$ and $\check{\mathbf{b}} \in \mathbb{C}^D$ such that

$$\mathbf{U} = \text{mat}(\check{\mathbf{v}}) = \check{\mathbf{a}}\check{\mathbf{b}}^\top.$$

Let $\widehat{\mathbf{a}} \in \mathbb{C}^D$ and $\widehat{\mathbf{b}} \in \mathbb{C}^D$ respectively denote the left and right singular vectors of the rank-1 matrix $\widehat{\mathbf{V}}$. Define $T_1 = T(\{\check{\mathbf{a}}, \check{\mathbf{b}}\})$, $T_2 = T(\{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}\})$, and $T_3 = T_1 \cap T_2$. Then $T_1 + T_2$ is rewritten as

$$T_1 + T_2 = \mathbf{P}_{T_2^\perp} T_1 \oplus T_2 = \mathbf{P}_{T_2^\perp} T_1 \oplus T_3 \oplus \mathbf{P}_{T_3^\perp} T_2. \quad (90)$$

Similarly, we also have

$$T_1 + T_2 = T_1 \oplus \mathbf{P}_{T_1^\perp} T_2 = \mathbf{P}_{T_3^\perp} T_1 \oplus T_3 \oplus \mathbf{P}_{T_1^\perp} T_2. \quad (91)$$

By the definition of T_2 , we have

$$\|\mathbf{P}_{T_2} \mathbf{v}\|_2 \geq \|\mathbf{P}_{T_1} \mathbf{v}\|_2.$$

Therefore,

$$\|\mathbf{P}_{\mathbf{P}_{T_3^\perp} T_2} \mathbf{v}\|_2 \geq \|\mathbf{P}_{\mathbf{P}_{T_3^\perp} T_1} \mathbf{v}\|_2.$$

Then by (90) and (91) it follows that

$$\|\mathbf{P}_{\mathbf{P}_{T_1^\perp} T_2} \mathbf{v}\|_2 \geq \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2. \quad (92)$$

By the Cauchy-Schwartz inequality and the Pythagorean identity, we have

$$\begin{aligned} |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2 &= |\langle \mathbf{P}_{T_1} \mathbf{v}, \check{\mathbf{v}} \rangle|^2 \\ &\leq \|\mathbf{P}_{T_1} \mathbf{v}\|_2^2 \\ &\leq 1 - \|\mathbf{P}_{T_1^\perp} \mathbf{v}\|_2^2 \\ &\leq 1 - \|\mathbf{P}_{\mathbf{P}_{T_1^\perp} T_2} \mathbf{v}\|_2^2 \\ &\leq 1 - \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2^2, \end{aligned}$$

where the last step follow from (92). The above inequality is rearranged as

$$\|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 \leq \sqrt{1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2}. \quad (93)$$

We may assume that $|\langle \mathbf{v}, \check{\mathbf{v}} \rangle| > 2^{-1/2}$. Otherwise, the right-hand side of (89) becomes negative and the inequality holds trivially. Then by (93) we have

$$\|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 < |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|,$$

which also implies

$$\|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2 < |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|. \quad (94)$$

Since $\mathbf{P}_{T_1+T_2} \check{\mathbf{v}} = \check{\mathbf{v}}$, we have

$$\begin{aligned} |\langle \mathbf{v}, \check{\mathbf{v}} \rangle| &= |\langle \mathbf{P}_{T_1+T_2} \mathbf{v}, \check{\mathbf{v}} \rangle| \\ &= |\langle (\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} + \mathbf{P}_{T_2}) \mathbf{v}, \check{\mathbf{v}} \rangle| \\ &= |\langle \mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}, \check{\mathbf{v}} \rangle| + |\langle \mathbf{P}_{T_2} \mathbf{v}, \check{\mathbf{v}} \rangle| \\ &\leq \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2 \\ &\quad + \|\mathbf{P}_{T_2} \mathbf{v}\|_2 \|\mathbf{P}_{T_2} \check{\mathbf{v}}\|_2 \\ &\leq \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2 \\ &\quad + \sqrt{1 - \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2^2} \sqrt{1 - \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2^2}. \end{aligned}$$

By solving the above inequality for $\|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2$ under the condition in (94), we obtain

$$\begin{aligned} \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2 &\leq \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 |\langle \mathbf{v}, \check{\mathbf{v}} \rangle| \\ &\quad + \sqrt{1 - \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2^2} \sqrt{1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2} \\ &\leq \min(1, 2\sqrt{1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2}). \end{aligned} \quad (95)$$

Since $\mathbf{P}_{T_2}(\mathbf{v} - \widehat{\mathbf{v}}) = \mathbf{0}$, we have

$$\begin{aligned} |\langle \mathbf{v}, \check{\mathbf{v}} \rangle| - |\langle \widehat{\mathbf{v}}, \check{\mathbf{v}} \rangle| &\leq |\langle \mathbf{v} - \widehat{\mathbf{v}}, \check{\mathbf{v}} \rangle| \\ &= |\langle \mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1}(\mathbf{v} - \widehat{\mathbf{v}}), \check{\mathbf{v}} \rangle| \\ &= |\langle \mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}, \check{\mathbf{v}} \rangle| \\ &\leq \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \mathbf{v}\|_2 \|\mathbf{P}_{\mathbf{P}_{T_2^\perp} T_1} \check{\mathbf{v}}\|_2 \\ &\leq \min \left(\sqrt{1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2}, 2(1 - |\langle \mathbf{v}, \check{\mathbf{v}} \rangle|^2) \right), \end{aligned}$$

where the last step follows from (93) and (95). The assertion is obtained by a rearrangement. \square

Proof of Theorem 23: We use the mathematical induction and it suffices to show $\sin \angle(\mathbf{v}_t, \mathbf{u}) \leq \tau$ and (81) hold provided that $\sin \angle(\mathbf{v}_{t-1}, \mathbf{u}) \leq \tau$ for fixed t .

Since $\text{rank}(\text{mat}(\mathbf{v}_t)) = 1$, there exist $\mathbf{a}_t \in \mathbb{C}^M$ and $\mathbf{b}_t \in \mathbb{C}^D$ such that $\mathbf{v}_t = \mathbf{a}_t \otimes \mathbf{b}_t$. Similarly, there exist $\mathbf{a}_{t-1} \in \mathbb{C}^M$ and $\mathbf{b}_{t-1} \in \mathbb{C}^D$ that satisfy $\mathbf{v}_{t-1} = \mathbf{a}_{t-1} \otimes \mathbf{b}_{t-1}$. Let

$$\check{\mathbf{T}} = T(\mathbf{a}_{t-1}, \mathbf{b}_{t-1}) + T(\mathbf{a}_t, \mathbf{b}_t) + T(\mathbf{a}, \mathbf{b}).$$

Then define

$$\check{\mathbf{v}}'_t = \frac{\mathbf{P}_{\check{\mathbf{T}}} \mathbf{B} \mathbf{P}_{\check{\mathbf{T}}} \mathbf{v}_{t-1}}{\|\mathbf{P}_{\check{\mathbf{T}}} \mathbf{B} \mathbf{P}_{\check{\mathbf{T}}} \mathbf{v}_{t-1}\|_2}.$$

Note that Algorithm 2 produces the same result even when $\check{\mathbf{v}}_t$ is replaced by $\check{\mathbf{v}}'_t$. Indeed, since $\mathbf{P}_{\check{\mathbf{T}}} \mathbf{v}_{t-1} = \mathbf{v}_{t-1}$, it follows that $\text{mat}(\mathbf{B} \mathbf{v}_{t-1})$ and $\text{mat}(\mathbf{B} \mathbf{P}_{\check{\mathbf{T}}} \mathbf{v}_{t-1})$ are collinear, so are their rank-1 approximations. Moreover, by \mathbf{v}_t is obtained normalizing as the normalized rearrangement of the rank-1 approximation of $\text{mat}(\mathbf{B} \mathbf{v}_{t-1})$, by the construction of $\check{\mathbf{T}}$, it follows that $\text{mat}(\mathbf{P}_{\check{\mathbf{T}}} \mathbf{B} \mathbf{P}_{\check{\mathbf{T}}} \mathbf{v}_{t-1})$ is also collinear with $\text{mat}(\mathbf{B} \mathbf{v}_{t-1})$.

Let $\widehat{\mathbf{V}}'_t$ denote the rank-1 approximation of $\text{mat}(\check{\mathbf{v}}'_t)$ and $\widehat{\mathbf{u}}'_t = \text{vec}(\widehat{\mathbf{V}}'_t)$. Then we have

$$\mathbf{v}_t = \widehat{\mathbf{u}}'_t / \|\widehat{\mathbf{u}}'_t\|_2.$$

Let $\mathbf{v}(\check{T})$ denote a unique most dominant eigenvector of $\mathbf{P}_{\check{T}} \mathbf{B} \mathbf{P}_{\check{T}}$. Since $\|\check{\mathbf{v}}_t\|_2 = 1$, we have $\|\check{\mathbf{u}}'_t\|_2 \leq 1$. Therefore,

$$\begin{aligned} \sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) &= \sqrt{1 - |\langle \mathbf{v}_t, \mathbf{v}(\check{T}) \rangle|^2} \\ &\leq \sqrt{1 - |\langle \check{\mathbf{u}}'_t, \mathbf{v}(\check{T}) \rangle|^2}. \end{aligned}$$

We apply Lemma 27 with $\check{\mathbf{v}} = \mathbf{v}(\check{T})$ and $\mathbf{v} = \check{\mathbf{v}}'_t$. By Lemma 27, we have

$$\begin{aligned} &|\langle \check{\mathbf{u}}'_t, \mathbf{v}(\check{T}) \rangle| \\ &\geq |\langle \check{\mathbf{v}}'_t, \mathbf{v}(\check{T}) \rangle| \\ &\quad - \min \left(\sqrt{1 - |\langle \check{\mathbf{v}}'_t, \mathbf{v}(\check{T}) \rangle|^2}, 2(1 - |\langle \check{\mathbf{v}}'_t, \mathbf{v}(\check{T}) \rangle|^2) \right), \end{aligned}$$

which implies

$$\begin{aligned} \sqrt{1 - |\langle \check{\mathbf{u}}'_t, \mathbf{v}(\check{T}) \rangle|^2} &\leq \sqrt{5} \sqrt{1 - |\langle \check{\mathbf{v}}'_t, \mathbf{u}(\check{T}) \rangle|^2} \\ &= \sqrt{5} \sin \angle(\check{\mathbf{v}}'_t, \mathbf{v}(\check{T})). \end{aligned}$$

We apply Lemma 26 with $\mathbf{M} = \mathbf{P}_{\check{T}} \mathbf{B} \mathbf{P}_{\check{T}}$, $\mathbf{v} = \mathbf{v}(\check{T})$, and $\hat{\mathbf{v}} = \mathbf{v}_{t-1}$. Then

$$\begin{aligned} &\sin \angle(\check{\mathbf{v}}'_t, \mathbf{v}(\check{T})) \\ &\leq \frac{\lambda_2(\mathbf{M}) \sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T}))}{\lambda_1(\mathbf{M}) \cos \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})) - \lambda_2(\mathbf{M}) \sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T}))} \\ &\leq \frac{\lambda_2(\mathbf{M})}{\lambda_1(\mathbf{M}) \sqrt{1 - \tau^2} - \lambda_2(\mathbf{M}) \tau} \cdot \sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})), \quad (96) \end{aligned}$$

where the last step follows from $\sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})) \leq \tau'$.

Next we compute the two largest eigenvalues of $\mathbf{P}_{\check{T}} \mathbf{B} \mathbf{P}_{\check{T}}$. Since \mathbf{u} is a unique dominant eigenvector of $\underline{\mathbf{B}}$ and $\mathbf{P}_{\check{T}} \mathbf{u} = \mathbf{u}$, we have $\lambda_1(\mathbf{P}_{\check{T}} \underline{\mathbf{B}} \mathbf{P}_{\check{T}}) = \lambda_1(\underline{\mathbf{B}})$. Therefore, by the triangle inequality,

$$\begin{aligned} \lambda_1(\mathbf{P}_{\check{T}} \mathbf{B} \mathbf{P}_{\check{T}}) &\geq \lambda_1(\mathbf{P}_{\check{T}} \underline{\mathbf{B}} \mathbf{P}_{\check{T}}) - \|\mathbf{P}_{\check{T}}(\mathbf{B} - \underline{\mathbf{B}})\mathbf{P}_{\check{T}}\| \\ &\geq \lambda_1(\underline{\mathbf{B}}) - 6 \|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty}. \quad (97) \end{aligned}$$

By the variational characterization of eigenvalues, we have

$$\lambda_2(\mathbf{P}_{\check{T}} \underline{\mathbf{B}} \mathbf{P}_{\check{T}}) = \sup_{\mathbf{v}} \{\mathbf{v}^* \underline{\mathbf{B}} \mathbf{v} \mid \|\mathbf{v}\|_2 \leq 1, \mathbf{v} \in \mathbf{u}^\perp \cap \check{T}\} \leq \tilde{\lambda}_2(\underline{\mathbf{B}}).$$

Therefore,

$$\begin{aligned} \lambda_2(\mathbf{P}_{\check{T}} \mathbf{B} \mathbf{P}_{\check{T}}) &\leq \lambda_2(\mathbf{P}_{\check{T}} \underline{\mathbf{B}} \mathbf{P}_{\check{T}}) + \|\mathbf{P}_{\check{T}}(\mathbf{B} - \underline{\mathbf{B}})\mathbf{P}_{\check{T}}\| \\ &\leq \tilde{\lambda}_2(\underline{\mathbf{B}}) + 6 \|\mathbf{B} - \underline{\mathbf{B}}\|_{S_1 \rightarrow S_\infty}. \quad (98) \end{aligned}$$

By plugging in (97) and (98) into (96), we obtain that (78) implies

$$\sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) \leq \mu \sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})). \quad (99)$$

Moreover, by the transitivity of the angle function [57], we also have

$$\angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})) \leq \angle(\mathbf{v}_{t-1}, \mathbf{u}) + \angle(\mathbf{u}, \mathbf{v}(\check{T})). \quad (100)$$

Next we apply Lemma 25 for $\mathbf{M}_1 = \mathbf{B}$, $\mathbf{M}_2 = \underline{\mathbf{B}}$, and $\mathbf{v} = \mathbf{v}(\check{T})$. Since (80) implies (88), it follows from Lemma 25 that

$$\sin \angle(\mathbf{u}, \mathbf{v}(\check{T})) \leq \frac{4\sqrt{6} \|\mathbf{B} - \underline{\mathbf{B}}\|_{S_\infty}}{\lambda_1(\underline{\mathbf{B}})}.$$

Then (79) implies

$$\sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) < \frac{1}{3\sqrt{2}}.$$

Since $\sin \angle(\mathbf{v}_{t-1}, \mathbf{u}) \leq \tau < \frac{1}{3\sqrt{2}}$, it follows that (100) implies

$$\sin \angle(\mathbf{v}_{t-1}, \mathbf{v}(\check{T})) \leq \sin \angle(\mathbf{v}_{t-1}, \mathbf{u}) + \sin \angle(\mathbf{u}, \mathbf{v}(\check{T})). \quad (101)$$

By (79), (99), and (101),

$$\sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) < \frac{1}{3\sqrt{2}}.$$

Similarly to the previous case, the transitivity of the angle function implies

$$\angle(\mathbf{v}_t, \mathbf{u}) \leq \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) + \angle(\mathbf{v}(\check{T}), \mathbf{u}).$$

Then it follows that

$$\sin \angle(\mathbf{v}_t, \mathbf{u}) \leq \sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) + \sin \angle(\mathbf{v}(\check{T}), \mathbf{u}).$$

By collecting the above inequalities, we obtain

$$\sin \angle(\mathbf{v}_t, \mathbf{u}) \leq \mu \sin \angle(\mathbf{v}_t, \mathbf{v}(\check{T})) + (1 + \mu) \sin \angle(\mathbf{v}(\check{T}), \mathbf{u}). \quad (102)$$

Finally, we verify that (102) and (79) imply $\sin \angle(\mathbf{v}_t, \mathbf{u}) \leq \tau$. This completes the proof. \square

X. CONCLUSION

We studied two iterative algorithms and their performance guarantees for a multichannel blind deconvolution that imposes a bilinear model on channel impulse responses. Such a bilinear model is obtained, for example, by embedding a parametric model for the shapes of the impulse responses into a low-dimensional subspace through manifold embedding, while the channel gains are treated as independent variables. Unlike recent theoretical results on blind deconvolution in the literature, we do not impose a strong geometric constraint on the input source signal. Under the bilinear model, we modified classical cross-convolution method based on the commutativity of the convolution to overcome its critical weakness of sensitivity to noise. The bilinear system model imposes a strong prior on the unknown channel impulse responses, which enables us to recover the system with short observation. The constraint by the bilinear model, on the other hand, makes the recovery no longer a simple eigenvalue decomposition problem. Therefore, standard algorithms in numerical linear algebra do not apply to this non-convex optimization problem. We propose two iterative algorithms along with a simple spectral initialization. When the basis in the bilinear model is generic, we have shown that the proposed algorithms converge linearly to a stable estimate of the unknown channel parameters with provable non-asymptotic performance guarantees.

Mathematically, our analysis involves tail estimates of norms of several structured random matrices, which are written as suprema of coupled high-order subgaussian processes. In an earlier version of our approach [51], we used the concentration of a polynomial in subgaussian random vector [52] together with the union bound through the ϵ -net argument. In this revised analysis, we factorized high-order random processes

using Gaussian processes of the first or second order and computed the supremum using sharp tail estimates in the literature (e.g., [58]). This change has already provided a significant improvement in scaling laws of key parameters in the main results but the sharpened scaling law is still suboptimal compared to the degrees of freedom in the underlying model. Because we formulated the observations with the circular convolution modulo L , it is necessary to assume that L is no smaller than the length K of the impulse responses. Otherwise, the problem becomes a demixing problem that separates mixture of convolutions. The original formulation given in terms of the linear convolution in the literature (e.g., [19] [20]) also requires that the length of observation exceeds the length of the impulse responses. It has not been explored yet whether one can further reduce the sample complexity in solving the resulting blind demixing problem for $L < K$ with the priors considered in this paper. On the other hand, we expect that it would be possible to further sharpen the estimates for structured random matrices. It remains as an interesting open question how to extend the sharp estimates on suprema of second-order chaos processes [58] to higher orders similarly to the extension of the Hanson-Wright inequality [59] for concentration of subgaussian quadratic forms to higher-order polynomials [52].

APPENDIX A TOOLBOX

In this section, we provide a collection of lemmas, which serve as mathematical tools to derive estimates of structured random matrices.

Lemma 28 (Complexification of Hanson-Wright Inequality [60, Th. 1.1]): Let $\mathbf{A} \in \mathbb{C}^{m \times n}$. Let $\mathbf{g} \in \mathbb{C}^n$ be a standard complex Gaussian vector. For any $0 < \zeta < 1$, there exists an absolute constant C such that

$$\begin{aligned} & \| \mathbf{A} \mathbf{g} \|_2^2 - \mathbb{E} [\| \mathbf{A} \mathbf{g} \|_2^2] \\ & \leq C [\| \mathbf{A}^* \mathbf{A} \|_F \sqrt{\log(2\zeta^{-1})} \vee \| \mathbf{A} \|^2 \log(2\zeta^{-1})] \end{aligned}$$

holds with probability $1 - \zeta$.

Lemma 29 (Complexification of Hanson-Wright Inequality [60, Th. 2.1]): Let $\mathbf{A} \in \mathbb{C}^{m \times n}$. Let $\mathbf{g} \in \mathbb{C}^n$ be a standard complex Gaussian vector. For any $0 < \zeta < 1$, there exists an absolute constant C such that

$$\| \mathbf{A} \mathbf{g} \|_2 - \| \mathbf{A} \|_F \leq C \| \mathbf{A} \| \sqrt{\log(2\zeta^{-1})}$$

holds with probability $1 - \zeta$.

The following lemma is a direct consequence of Maurey's empirical method [61].

Lemma 30 (Maurey's empirical method [62, Lemma 3.1]): Let $k, m, n \in \mathbb{N}$ and $T : \ell_1^k(\mathbb{R}) \rightarrow \ell_\infty^m(\ell_2^d(\mathbb{R}))$ be a linear operator. Then

$$\begin{aligned} & \int_0^\infty \sqrt{\log N(T(B_{\ell_1^k(\mathbb{R})}), \| \cdot \|_{\ell_\infty^m(\ell_2^d(\mathbb{R}))}, t) dt} \\ & \leq C \sqrt{(1 + \log k)(1 + \log m)(1 + \log m + \log d) \| T \|}. \end{aligned}$$

Lemma 30 extends to the complex field case, which is shown in the following corollary.

Corollary 31: Let $k, m, n \in \mathbb{N}$ and $T : \ell_1^k(\mathbb{C}) \rightarrow \ell_\infty^m(\ell_2^d(\mathbb{C}))$ be a linear operator. Then

$$\begin{aligned} & \int_0^\infty \sqrt{\log N(T(B_{\ell_1^k(\mathbb{C})}), \| \cdot \|_{\ell_\infty^m(\ell_2^d(\mathbb{C}))}, t) dt} \\ & \leq C \sqrt{(1 + \log k)(1 + \log m)(1 + \log m + \log d) \| T \|}. \end{aligned}$$

The following lemma provides tail estimates of suprema of subgaussian processes.

Lemma 32: Let $\xi \in \mathbb{C}^n$ be a standard Gaussian vector with $\mathbb{E} \xi \xi^* = \mathbf{I}_n$, $\Delta \subset \mathbb{C}^n$, and $0 < \zeta < e^{-1/2}$. There is an absolute constants C such that

$$\sup_{f \in \Delta} |f^* \xi| \leq C \sqrt{\log(\zeta^{-1})} \int_0^\infty \sqrt{\log N(\Delta, \| \cdot \|_2, t) dt}$$

holds with probability $1 - \zeta$.

Theorem 33 ([58, Th. 3.1]): Let $\xi \in \mathbb{C}^n$ be an L -subgaussian vector with $\mathbb{E} \xi \xi^* = \mathbf{I}_n$, $\Delta \subset \mathbb{C}^{m \times n}$, and $0 < \zeta < 1$. There exists a constant $C(L)$ that only depends on L such that

$$\begin{aligned} & \sup_{\mathbf{M} \in \Delta} \| \mathbf{M} \xi \|_2^2 - \mathbb{E} [\| \mathbf{M} \xi \|_2^2] \\ & \leq C(L) \left[K_1 + K_2 \sqrt{\log(2\zeta^{-1})} + K_3 \log(2\zeta^{-1}) \right] \end{aligned}$$

holds with probability $1 - \zeta$, where K_1 , K_2 , and K_3 are given by

$$\begin{aligned} K_1 &:= \gamma_2(\Delta, \| \cdot \|) [\gamma_2(\Delta, \| \cdot \|) + d_F(\Delta)] \\ &\quad + d_F(\Delta) d_S(\Delta), \\ K_2 &:= d_S(\Delta) [\gamma_2(\Delta, \| \cdot \|) + d_F(\Delta)], \\ K_3 &:= d_S^2(\Delta). \end{aligned}$$

Using the polarization identity, this result on the suprema of second order chaos processes has been extended from a subgaussian quadratic form to a subgaussian bilinear form [63].

Theorem 34 (A corollary of [63, Th. 2.3]): Let $\xi \in \mathbb{C}^n$ be an L -subgaussian vector with $\mathbb{E} \xi \xi^* = \mathbf{I}_n$, $\Delta_1, \Delta_2 \subset \mathbb{C}^{m \times n}$, $0 < \zeta < 1$, and $a > 0$. There exists a constant $C(L)$ that only depends on L such that

$$\begin{aligned} & \sup_{\mathbf{M}_1 \in \Delta_1, \mathbf{M}_2 \in \Delta_2} | \langle \mathbf{M}_1 \xi, \mathbf{M}_2 \xi \rangle - \mathbb{E} [\langle \mathbf{M}_1 \xi, \mathbf{M}_2 \xi \rangle] | \\ & \leq C(L) \left[\tilde{K}_1 + \tilde{K}_2 \sqrt{\log(8\zeta^{-1})} + \tilde{K}_3 \log(8\zeta^{-1}) \right], \end{aligned}$$

holds with probability $1 - \zeta$, where \tilde{K}_1 , \tilde{K}_2 , and \tilde{K}_3 are given by

$$\begin{aligned} \tilde{K}_1 &:= [a\gamma_2(\Delta_1, \| \cdot \|) + a^{-1}\gamma_2(\Delta_2, \| \cdot \|)] \\ &\quad \cdot [a\gamma_2(\Delta_1, \| \cdot \|) + a^{-1}\gamma_2(\Delta_2, \| \cdot \|)] \\ &\quad + a d_F(\Delta_1) + a^{-1} d_F(\Delta_2) \\ &\quad + [a d_F(\Delta_1) + a^{-1} d_F(\Delta_2)] \\ &\quad \cdot [a d_S(\Delta_1) + a^{-1} d_S(\Delta_2)], \\ \tilde{K}_2 &:= [a d_S(\Delta_1) + a^{-1} d_S(\Delta_2)] \\ &\quad \cdot [a\gamma_2(\Delta_1, \| \cdot \|) + a^{-1}\gamma_2(\Delta_2, \| \cdot \|)] \\ &\quad + a d_F(\Delta_1) + a^{-1} d_F(\Delta_2), \\ \tilde{K}_3 &:= [a d_S(\Delta_1) + a^{-1} d_S(\Delta_2)]^2. \end{aligned}$$

A special case of Theorem 34 where $a = 1$ was shown in [63, Th. 2.3]. Note that the bilinear form satisfies

$$\langle \mathbf{M}_1 \xi, \mathbf{M}_2 \xi \rangle = \langle a \mathbf{M}_1 \xi, a^{-1} \mathbf{M}_2 \xi \rangle, \quad \forall a > 0.$$

Moreover, the γ_2 functional and the radii with respect to the Frobenius and spectral norms are all 1-homogeneous functions. Therefore, Theorem 34 is a direct consequence of [63, Th. 2.3].

Since $a > 0$ in Theorem 34 is arbitrary, one can minimize the tail estimate over $a > 0$.

APPENDIX B EXPECTATIONS

The following lemmas on the expectation of structured random matrices are derived in [22]. For the convenience of the readers, we include the lemmas. Here the matrix $\tilde{\Phi}_m \in \mathbb{C}^{L \times D}$ denotes the zero-padded matrix of Φ_m given by $\tilde{\Phi}_m \mathbf{S}^\top \Phi_m$ for $m = 1, \dots, M$, where $\mathbf{S} \in \mathbb{R}^{K \times L}$ is defined in (2).

Lemma 35 ([22, Lemma B.1]): Under the assumption in (A1),

$$\mathbb{E}[\mathbf{C}_{\tilde{\Phi}_m \mathbf{u}_m}^* \mathbf{C}_{\tilde{\Phi}_m \mathbf{u}_m}] = K \|\mathbf{u}_m\|_2^2 \mathbf{I}_L.$$

Lemma 36 ([22, Lemma B.2]): Under the assumption in (A1),

$$\mathbb{E}[\mathbf{C}_{\tilde{\Phi}_m \mathbf{u}_m}^* \tilde{\Phi}_m] = K \mathbf{e}_1 \mathbf{u}_m^*.$$

Lemma 37 ([22, Lemma B.3]): Under the assumption in (A1),

$$\begin{aligned} \mathbb{E}[\tilde{\Phi}_m^* \mathbf{C}_{\tilde{\Phi}_{m'} \mathbf{u}_{m'}}^* \mathbf{C}_x^* \mathbf{C}_x \mathbf{C}_{\tilde{\Phi}_{m'} \mathbf{u}_{m'}} \tilde{\Phi}_m] \\ = \begin{cases} K^2 \|\mathbf{x}\|_2^2 \|\mathbf{u}_{m'}\|_2^2 \mathbf{I}_D & m \neq m', \\ K^2 \|\mathbf{x}\|_2^2 (\|\mathbf{u}_{m'}\|_2^2 \mathbf{I}_D + \mathbf{u}_{m'} \mathbf{u}_{m'}^*) & m = m'. \end{cases} \end{aligned}$$

APPENDIX C PROOF OF LEMMA 1

Let $\mathbf{x}' \in \mathbb{C}^L$ and $\mathbf{b}' \in \mathbb{C}^D$. By the definition of an adjoint operator, we have

$$\langle \mathbf{x}' \otimes \mathbf{b}' \otimes \mathbf{1}_{M,1}, \mathcal{A}^*(\mathbf{y}) \rangle = \langle \mathcal{A}(\mathbf{x}' \otimes \mathbf{b}' \otimes \mathbf{1}_{M,1}), \mathbf{y} \rangle.$$

Then by the definition of \mathcal{A} , we continue as

$$\begin{aligned} \langle \mathcal{A}(\mathbf{x}' \otimes \mathbf{b}' \otimes \mathbf{1}_{M,1}), \mathbf{y} \rangle \\ &= \sum_{m=1}^M \langle \mathbf{C}_{\mathbf{S}^* \Phi_m \mathbf{b}'} \mathbf{x}', \mathbf{y}_m \rangle \\ &= \sum_{m=1}^M \mathbf{x}'^* \mathbf{C}_{\mathbf{S}^* \Phi_m \mathbf{b}'}^* \mathbf{y}_m \\ &= \sum_{m=1}^M \mathbf{x}'^* (\mathbf{J} \mathbf{S}^* \overline{\Phi_m \mathbf{b}'} \otimes \mathbf{y}_m) \\ &= \sum_{m=1}^M \mathbf{x}'^* \mathbf{J} (\mathbf{S}^* \overline{\Phi_m \mathbf{b}'} \otimes \mathbf{J} \mathbf{y}_m) \\ &= \sum_{m=1}^M \mathbf{x}'^* \mathbf{J} \mathbf{C}_{\mathbf{y}_m}^\top \mathbf{S}^* \overline{\Phi_m \mathbf{b}'}. \end{aligned}$$

Here we used the fact that the transpose of \mathbf{C}_h satisfies $\mathbf{C}_h^\top = \mathbf{C}_{\mathbf{J}h}$.

Finally, by tensorizing the last term, we obtain

$$\begin{aligned} \sum_{m=1}^M \mathbf{x}'^* \mathbf{J} \mathbf{C}_{\mathbf{y}_m}^\top \mathbf{S}^* \overline{\Phi_m \mathbf{b}'} \\ &= \sum_{m=1}^M \mathbf{x}'^* ((\mathbf{b}')^* \otimes \mathbf{I}_L) \text{vec}(\mathbf{J} \mathbf{C}_{\mathbf{y}_m}^\top \mathbf{S}^* \overline{\Phi_m}) \\ &= \sum_{m=1}^M (\mathbf{b}' \otimes \mathbf{x}')^* \text{vec}(\mathbf{J} \mathbf{C}_{\mathbf{y}_m}^\top \mathbf{S}^* \overline{\Phi_m}) \\ &= \sum_{m=1}^M (\mathbf{x}' \otimes \mathbf{b}')^* \text{vec}(\Phi_m^* \mathbf{S} \mathbf{C}_{\mathbf{y}_m} \mathbf{J}). \end{aligned}$$

Then the assertion follows since \mathbf{x}' and \mathbf{b}' were arbitrary.

APPENDIX D PROOF OF LEMMA 13

The left-hand side of (50) is rewritten as a variational form given by

$$\begin{aligned} \sup_{\substack{\mathbf{z} \in B_2^{2K-1} \\ \mathbf{q} \in B_2^D}} \left(\underbrace{\sum_{m=1}^M a_m \mathbf{q}^* \Phi_m^* \mathbf{S} \mathbf{C}_{\mathbf{S}^* \Phi_m \mathbf{b}} \check{\mathbf{S}}^* \mathbf{z}}_{(\#\#)} \right. \\ \left. - \mathbb{E}_\phi \left[\sum_{m=1}^M a_m \mathbf{q}^* \Phi_m^* \mathbf{S} \mathbf{C}_{\mathbf{S}^* \Phi_m \mathbf{b}} \check{\mathbf{S}}^* \mathbf{z} \right] \right). \quad (103) \end{aligned}$$

Since

$$\begin{aligned} \mathbf{q}^* \Phi_m^* \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^* \Phi_m \mathbf{b} &= \text{vec}(\Phi_m)^* (\bar{\mathbf{q}} \otimes \mathbf{I}_K) \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \\ &\quad \cdot \mathbf{S}^* (\mathbf{b}^\top \otimes \mathbf{I}_K) \text{vec}(\Phi_m) \\ &= \text{vec}(\Phi_m)^* (\bar{\mathbf{q}} \mathbf{b}^\top \otimes \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^*) \text{vec}(\Phi_m), \end{aligned}$$

$(\#\#)$ is rewritten as

$$\sum_{m=1}^M a_m \text{vec}(\Phi_m)^* (\bar{\mathbf{q}} \mathbf{b}^\top \otimes \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^*) \text{vec}(\Phi_m).$$

Let $\phi = [\text{vec}(\Phi_1)^\top, \dots, \text{vec}(\Phi_M)^\top]^\top$. Then

$$\begin{aligned} \sum_{m=1}^M a_m \text{vec}(\Phi_m)^* (\bar{\mathbf{q}} \mathbf{b}^\top \otimes \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^*) \text{vec}(\Phi_m) \\ = \phi^* \left(\sum_{m=1}^M a_m \mathbf{e}_m \mathbf{e}_m^* \otimes \bar{\mathbf{q}} \mathbf{b}^\top \otimes \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^* \right) \phi. \end{aligned}$$

Therefore, the objective function in the supremum in (103) becomes a second-order chaos process. We compute the tail estimate of the supremum by applying Theorem 34 with

$$\Delta_1 = \left\{ \sum_{m=1}^M a_m \mathbf{e}_m \mathbf{e}_m^* \otimes \mathbf{q}^\top \otimes \mathbf{I}_K \mid \mathbf{q} \in B_2^D \right\}$$

and

$$\Delta_2 = \left\{ \sum_{m=1}^M \mathbf{e}_m \mathbf{e}_m^* \otimes \mathbf{b}^\top \otimes \mathbf{S} \mathbf{C}_{\check{\mathbf{S}}^* \mathbf{z}}^* \mathbf{S}^* \mid \mathbf{q} \in B_2^{2K-1} \right\}.$$

By direct calculation, the radii of Δ_1 and Δ_2 are given as follows:

$$\begin{aligned} d_S(\Delta_1) &\leq \|\mathbf{a}\|_\infty, \\ d_F(\Delta_1) &\leq \|\mathbf{a}\|_2 \sqrt{K}, \\ d_S(\Delta_2) &\leq \|\mathbf{b}\|_2 \sqrt{K}, \\ d_F(\Delta_2) &\leq \|\mathbf{b}\|_2 \sqrt{MK}. \end{aligned}$$

Here, we used the fact that

$$\|SC_{\check{S}^*z}S^*\| \leq \|SC_{\check{S}^*z}S^*\|_F \leq \sqrt{K}\|z\|_2 \leq \sqrt{K}.$$

Moreover, since

$$d_S(\Delta_1) \leq \|\mathbf{a}\|_\infty \|\mathbf{q}\|_2,$$

by Dudley's inequality and a standard volume argument, we have

$$\begin{aligned} \gamma_2(\Delta_1) &\leq C_1 \|\mathbf{a}\|_\infty \int_0^\infty \sqrt{\log N(B_2^D, \|\cdot\|_2, t)} dt \\ &\leq C_2 \|\mathbf{a}\|_\infty \sqrt{D}. \end{aligned}$$

On the other hand, since

$$C_{\check{S}^*z} = \sqrt{L}F^* \text{diag}(F\check{S}^*z)F,$$

we have

$$d_S(\Delta_2) \leq \|\mathbf{b}\|_2 \sqrt{L} \|F\check{S}^*z\|_\infty,$$

which implies

$$\begin{aligned} \gamma_2(\Delta_2) &\leq C_1 \|\mathbf{b}\|_2 \sqrt{L} \\ &\quad \cdot \int_0^\infty \sqrt{\log N(F\check{S}^*B_2^{2K-1}, \|\cdot\|_\infty, t)} dt \\ &\leq C_3 \|\mathbf{b}\|_2 \sqrt{LK} \\ &\quad \cdot \int_0^\infty \sqrt{\log N(F\check{S}^*B_1^{2K-1}, \|\cdot\|_\infty, t)} dt \\ &\leq C_4 \|\mathbf{b}\|_2 \sqrt{K} \sqrt{\log(2K-1)} \log^{3/2} L \\ &\leq C_5 \|\mathbf{b}\|_2 \sqrt{K} \sqrt{\log K} \log^{3/2} L, \end{aligned}$$

where the third step follows from Corollary 31.

By applying these estimates to Theorem 34 with

$$a = \sqrt{\frac{\gamma_2(\Delta_2, \|\cdot\|) d_F(\Delta_2)}{\gamma_2(\Delta_1, \|\cdot\|) d_F(\Delta_1)}}, \quad (104)$$

we obtain that the supremum in (103) is less than

$$C'(\beta) \log^\alpha(MKL) \cdot (\sqrt{MK}^{3/4} D^{1/4} + \sqrt{MK} + \sqrt{MKD})$$

with probability $1 - K^{-\beta}$. By the arithmetic-geometric mean inequality,

$$\sqrt{MK}^{3/4} D^{1/4} \leq \frac{\sqrt{MK} + \sqrt{MKD}}{2}.$$

We also have $\sqrt{MK} \geq \sqrt{MKD}$ since $K \geq D$. This completes the proof.

APPENDIX E PROOF OF LEMMA 14

The spectral norm in the left-hand side of (54) is expressed as a variational form given by

$$\sup_{z \in B_2^{2K-1}} \sup_{q \in B_2^D} \sum_{m=1}^M z^* \check{S} C_x C_{w_m}^* S^* \Phi_m q. \quad (105)$$

The objective function in (105) is rewritten as

$$\begin{aligned} &\sum_{m=1}^M z^* \check{S} C_x C_{w_m}^* S^* \Phi_m q \\ &= \sum_{m=1}^M (q^\top \otimes z^* \check{S} C_x C_{w_m}^* S^*) \text{vec}(\Phi_m) \\ &= \left\langle f(q, z), \sum_{m=1}^M e_m \otimes \text{vec}(\Phi_m) \right\rangle, \end{aligned}$$

where

$$f(q, z) = \sum_{m=1}^M e_m \otimes \bar{q} \otimes SC_{w_m} C_x^* \check{S} z.$$

Note that

$$\sum_{m=1}^M e_m \otimes \text{vec}(\Phi_m) = \begin{bmatrix} \text{vec}(\Phi_1) \\ \vdots \\ \text{vec}(\Phi_M) \end{bmatrix}$$

is a standard complex Gaussian vector of length MKD . Therefore, we compute a tail estimate of the supremum in (105) by applying Lemma 32 with

$$\Delta = \{f(q, z) \mid q \in B_2^D, z \in B_2^{2K-1}\}.$$

Since

$$\begin{aligned} &\|f(q, z) - f(q', z')\|_2 \\ &\leq \|f(q, z) - f(q', z)\|_2 + \|f(q', z) - f(q', z')\|_2 \\ &\leq \sqrt{M} \|SC_{w_m} C_x^* \check{S}\| (\|q - q'\|_2 + \|z - z'\|_2) \\ &\leq \sqrt{M} \rho_{x,w} (\|q - q'\|_2 + \|z - z'\|_2), \end{aligned}$$

we have

$$\begin{aligned} &\int_0^\infty \sqrt{\log N(\Delta, \|\cdot\|_2, t)} dt \\ &\leq \sqrt{M} \rho_{x,w} \left(\int_0^\infty \sqrt{\log N(B_2^{2K-1}, \|\cdot\|_2, t)} dt \right. \\ &\quad \left. + \int_0^\infty \sqrt{\log N(B_2^D, \|\cdot\|_2, t)} dt \right) \\ &\leq C \rho_{x,w} \sqrt{MK}, \end{aligned}$$

where the last step follows from a standard volume argument and the fact that $K \geq D$. The assertion then follows by applying the above estimate to Lemma 32.

APPENDIX F
PROOF OF LEMMA 16

We use the following lemma from [22] to prove Lemma 16.

Lemma 38 ([22, Lemma 5.3]): Let $\Psi \in \mathbb{C}^{K \times D}$ satisfy that $\text{vec}(\Psi)$ follows $\mathcal{CN}(0, \mathbf{I}_{KD})$, $0 < \zeta < 1$, and $A \in \mathbb{C}^{K \times K}$. Then

$$\|\Psi^* A \Psi - \mathbb{E}_\phi[\Psi^* A \Psi]\| \leq C \|A\| \sqrt{KD} \log(8\zeta^{-1})$$

holds with probability $1 - \zeta$.

Note that $\Gamma_n \Gamma_n^*$ is expressed as

$$\Gamma_n \Gamma_n^* = \sum_{m, m'=1}^M \Phi_m^* S C_{w_m} C_{w_{m'}}^* S^* \Phi_{m'}.$$

We apply Lemma 38 with

$$\Psi = [\Phi_1^\top, \dots, \Phi_M^\top]^\top$$

and

$$A = \sum_{m, m'=1}^M e_m e_{m'}^* S (C_{w_m} C_{w_{m'}}^* - \mathbb{E}_w[C_{w_m} C_{w_{m'}}^*]) S^*.$$

By the block-Gershgorin-disk theorem [64], it follows that

$$\|A\| \leq \max_{1 \leq m \leq M} \sum_{m'=1}^M \|S(C_{w_m} C_{w_{m'}}^* - \mathbb{E}_w[C_{w_m} C_{w_{m'}}^*]) S^*\| \leq M \rho_w.$$

Then the assertion follows by Lemma 38.

APPENDIX G
PROOF OF LEMMA 21

We decompose $\Phi^* Y_s^* Y_n \Phi$ into two parts respectively corresponding to the diagonal block portion and the off-diagonal block portion of $Y_s^* Y_n$:

$$\Phi^* Y_s^* Y_n \Phi = (g) + (h),$$

where

$$(g) = \sum_{m=1}^M e_m e_m^* \otimes \left(\sum_{\substack{m'=1 \\ m' \neq m}}^M \bar{a}_{m'} \tilde{\Phi}_m^* C_{\tilde{\Phi}_{m'} b}^* C_x^* C_{w_{m'}} \tilde{\Phi}_m \right),$$

$$(h) = - \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \neq m}}^M e_m e_{m'}^* \otimes \bar{a}_{m'} \tilde{\Phi}_m^* C_{\tilde{\Phi}_{m'} b}^* C_x^* C_{w_m} \tilde{\Phi}_{m'}. \quad (106)$$

Since $\|\cdot\|_{S_1 \rightarrow S_\infty}$ is a valid norm, by the triangle inequality, we have

$$\|\Phi^* Y_s^* Y_n \Phi\|_{S_1 \rightarrow S_\infty} \leq \|(g)\|_{S_1 \rightarrow S_\infty} + \|(h)\|_{S_1 \rightarrow S_\infty}.$$

Furthermore, by (62), we also have

$$\|(g)\|_{S_1 \rightarrow S_\infty} \leq \|(g)\|.$$

We use a tail estimate of $\|(g)\|$ derived in the proof of [22, Lemma 3.6]. It has been shown that

$$\|(g)\| \leq C(\beta) \rho_{x,w} K \sqrt{D} \|a\|_2 \|b\|_2 \log^\alpha(MKL) \quad (107)$$

with probability $1 - CK^{-\beta}$ (See [22, Sec. 5.3]). We will show that the tail estimate of $\|(g)\|$ is dominated by that for $\|(h)\|_{S_1 \rightarrow S_\infty}$.

For the part corresponding to the off-diagonal portion of $Y_s^* Y_n$, we add and subtract the diagonal sum in (h) and obtain

$$(h) = (k) + (l)$$

for

$$(k) = \sum_{m=1}^M e_m e_m^* \otimes \Phi_m^* S C_x^* C_{w_m} \check{S}^* Z_m,$$

$$(l) = - \sum_{m, m'=1}^M e_m e_{m'}^* \otimes \Phi_m^* S C_x^* C_{w_m} \check{S}^* Z_{m'}, \quad (108)$$

where

$$Z_m := \bar{a}_m \check{S} C_{\tilde{\Phi}_m b}^* \tilde{\Phi}_m, \quad m = 1, \dots, M.$$

Again, since $\|(k)\|_{S_1 \rightarrow S_\infty} \leq \|(k)\|$, we can use a tail estimate of $\|(k)\|$ derived in the proof of [22, Lemma 3.6]. It has been shown that

$$\|(k)\| \leq \rho_{x,w} C(\beta) K^{3/2} \|a\|_\infty \|b\|_2 \log^\alpha(MKL)$$

holds with probability $1 - CK^{-\beta}$. We will show that the tail estimate of $\|(k)\|$ is dominated by that for $\|(l)\|_{S_1 \rightarrow S_\infty}$, which we derive below.

Through a factorization of the full 2D summation in (l), we obtain

$$\|(l)\|_{S_1 \rightarrow S_\infty} \leq \left\| \underbrace{\sum_{m=1}^M e_m^* \otimes \check{S} C_{w_m}^* C_x S^* \Phi_m}_{(o)} \right\|_{S_1 \rightarrow S_2} \cdot \left\| \underbrace{\sum_{m'=1}^M e_{m'}^* \otimes (Z_{m'} - \mathbb{E}[Z_{m'}])}_{(p)} \right\|_{S_1 \rightarrow S_2} + \left\| \underbrace{\sum_{m, m'=1}^M e_m e_{m'}^* \otimes \Phi_m^* S C_x^* C_{w_m} \check{S}^* \mathbb{E}[Z_{m'}]}_{(q)} \right\|_{S_1 \rightarrow S_\infty}.$$

Note that $\|(o)\|_{S_1 \rightarrow S_2}$ is written as the supremum of a Gaussian process and is bounded by the following lemma.

Lemma 39: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there exists a constant $C(\beta)$ that depends only on β such that, conditional on the noise vector w ,

$$\left\| \sum_{m=1}^M e_m^* \otimes \check{S} C_{w_m}^* C_x S^* \Phi_m \right\|_{S_1 \rightarrow S_2} \leq C \sqrt{1 + \beta} \rho_{x,w} \sqrt{M + D + K} \log K$$

holds with probability $1 - K^{-\beta}$.

Proof of Lemma 39: Let $\phi_m = \text{vec}(\Phi_m)$ for $m = 1, \dots, M$ and $\phi = [\phi_1^\top, \dots, \phi_M^\top]^\top$. Let $q = [q_1, \dots, q_M]^\top \in \mathbb{C}^M$.

Then it follows from (61) that

$$\begin{aligned} & \left\| \sum_{m=1}^M \mathbf{e}_m^* \otimes \check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^* \Phi_m \right\|_{S_1 \rightarrow S_2} \\ &= \sup_{\substack{\mathbf{z} \in B_2^{2K-1} \\ \xi \in B_2^D, \mathbf{q} \in B_2^M}} \underbrace{\left| \sum_{m=1}^M q_m \mathbf{z}^* \check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^* \Phi_m \xi \right|}_{(\S\S)} \end{aligned}$$

where $(\S\S)$ satisfies

$$\begin{aligned} (\S\S) &= \left| \sum_{m=1}^M (q_m \xi^\top \otimes \mathbf{z}^* \check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^*) \phi_m \right| \\ &= \left| \sum_{m=1}^M q_m \mathbf{e}_m^* \otimes (\xi^\top \otimes \mathbf{z}^* \check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^*) \phi \right|. \end{aligned}$$

Let

$$\mathbf{f}(\mathbf{z}, \xi, \mathbf{q}) = \sum_{m=1}^M \bar{q}_m \mathbf{e}_m \otimes (\bar{\xi} \otimes \mathbf{S} \mathbf{C}_x^* \mathbf{C}_{\mathbf{w}_m} \check{\mathbf{S}}^* \mathbf{z}).$$

Then we obtain

$$\begin{aligned} & \left\| \sum_{m=1}^M \mathbf{e}_m^* \otimes \check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^* \Phi_m \right\| \\ &= \sup_{\mathbf{z} \in B_2^{2K-1}} \sup_{\xi \in B_2^D} \sup_{\mathbf{q} \in B_2^M} |\mathbf{f}(\mathbf{z}, \xi, \mathbf{q})^* \phi|. \end{aligned}$$

Note that $\mathbf{f}(\mathbf{z}, \xi)^* \phi$, conditioned on \mathbf{w} , is a centered Gaussian process. We compute a tail estimate of this supremum by applying Lemma 32 with

$$\Delta = \{\mathbf{f}(\mathbf{z}, \xi, \mathbf{q}) \mid \mathbf{z} \in B_2^{2K-1}, \xi \in B_2^D, \mathbf{q} \in B_2^M\}.$$

Then we need to compute the entropy integral for Δ . Recall

$$\begin{aligned} \rho_{x,w} &= \max_{1 \leq m \leq M} \|\check{\mathbf{S}} \mathbf{C}_x^* \mathbf{C}_{\mathbf{w}_m} \check{\mathbf{S}}^*\| \\ &\geq \|\check{\mathbf{S}} \mathbf{C}_{\mathbf{w}_m}^* \mathbf{C}_x \mathbf{S}^*\|, \quad \forall m = 1, \dots, M. \end{aligned}$$

By the triangle inequality, we obtain

$$\begin{aligned} & \|\mathbf{f}(\mathbf{z}, \xi, \mathbf{q}) - \mathbf{f}(\mathbf{z}', \xi', \mathbf{q}')\|_2 \\ &\leq \|\mathbf{f}(\mathbf{z}, \xi, \mathbf{q}) - \mathbf{f}(\mathbf{z}, \xi, \mathbf{q}')\|_2 \\ &\quad + \|\mathbf{f}(\mathbf{z}, \xi, \mathbf{q}') - \mathbf{f}(\mathbf{z}, \xi', \mathbf{q}')\|_2 \\ &\quad + \|\mathbf{f}(\mathbf{z}, \xi', \mathbf{q}') - \mathbf{f}(\mathbf{z}', \xi', \mathbf{q}')\|_2 \\ &\leq \rho_{x,w} (\|\mathbf{z}\|_2 \|\xi\|_2 \|\mathbf{q} - \mathbf{q}'\|_2 \\ &\quad + \|\mathbf{z}\|_2 \|\xi - \xi'\|_2 \|\mathbf{q}'\|_2 \\ &\quad + \|\mathbf{z} - \mathbf{z}'\|_2 \|\xi'\|_2 \|\mathbf{q}'\|_2) \\ &\leq \rho_{x,w} (\|\mathbf{q} - \mathbf{q}'\|_2 + \|\xi - \xi'\|_2 + \|\mathbf{z} - \mathbf{z}'\|_2). \end{aligned}$$

The integral of the log-entropy number is computed as

$$\begin{aligned} & \sup_{\mathbf{z} \in B_2^{2K-1}} \sup_{\xi \in B_2^D} \sup_{\mathbf{q} \in B_2^M} |\mathbf{f}(\mathbf{z}, \xi, \mathbf{q})^* \phi| \\ &\leq C_1 \int_0^\infty \sqrt{\log N(\Delta, \|\cdot\|_2, t)} dt \\ &\leq C_1 \rho_{x,w} \left(\int_0^\infty \sqrt{\log N(B_2^M, \|\cdot\|_2, t)} dt \right. \\ &\quad + \int_0^\infty \sqrt{\log N(B_2^M, \|\cdot\|_2, t)} dt \\ &\quad \left. + \int_0^\infty \sqrt{\log N(B_2^{2K-1}, \|\cdot\|_2, t)} dt \right) \\ &\leq C_2 \rho_{x,w} \sqrt{M + D + K}, \end{aligned}$$

where the last step follows from a standard volume argument. Then the assertion follows from Lemma 32. \square

Next $\|(\mathbf{p})\|_{S_1 \rightarrow S_2}$ is written as the supremum of a second-order Gaussian chaos process and its tail estimate can be derived by Theorem 34. However, the rank-1 constraint on the domain does not provide any gain in reducing the tail estimate in this case. Therefore, we use a previous estimate on $\|(\mathbf{p})\|_{S_1 \rightarrow S_2}$ derived in [22, Lemma 5.6], which is stated in the following lemma.

Lemma 40: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there exist a numerical constant $\alpha \in \mathbb{N}$ and a constant $C(\beta)$ that depends only on β such that

$$\begin{aligned} & \left\| \sum_{m=1}^M \mathbf{e}_m^* \otimes (\mathbf{Z}_m - \mathbb{E}[\mathbf{Z}_m]) \right\| \\ &\leq C(\beta) \|\mathbf{a}\|_\infty \|\mathbf{b}\|_2 (K + \sqrt{MKD}) \log^\alpha(MKL) \end{aligned}$$

holds with probability $1 - K^{-\beta}$.

Similarly to $\|(\mathbf{o})\|_{S_1 \rightarrow S_2}$, one can rewrite $\|(\mathbf{q})\|_{S_1 \rightarrow S_\infty}$ as the supremum of a Gaussian process. The following lemma provides its tail estimate.

Lemma 41: Suppose that (A1) holds. For any $\beta \in \mathbb{N}$, there exists a constant $C(\beta)$ that depends only on β such that, conditional on the noise vector \mathbf{w} ,

$$\begin{aligned} & \left\| \sum_{m,m'=1}^M \mathbf{e}_m \mathbf{e}_{m'}^* \otimes \Phi_m^* \mathbf{S} \mathbf{C}_x^* \mathbf{C}_{\mathbf{w}_m} \check{\mathbf{S}}^* \mathbb{E}[\mathbf{Z}_{m'}] \right\|_{S_1 \rightarrow S_\infty} \\ &\leq C \sqrt{1 + \beta} \rho_{x,w} \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 K \sqrt{M + D} \log K \quad (109) \end{aligned}$$

holds with probability $1 - K^{-\beta}$.

Proof of Lemma 41: It follows from the variational form in (61) and Lemma 36 that the left-hand side of (109) is written as

$$\begin{aligned} & K \sup_{\substack{\mathbf{q}, \tilde{\mathbf{q}} \in B_2^M \\ \xi, \tilde{\xi} \in B_2^D}} \left| \sum_{m,m'=1}^M q_m q_{m'} a_{m'} \tilde{\xi}^* \Phi_m^* \mathbf{S} \mathbf{C}_x^* \mathbf{w}_m \mathbf{b}^* \xi \right| \\ &= K \sup_{\tilde{\mathbf{q}} \in B_2^M} \left| \sum_{m'=1}^M q_{m'} a_{m'} \right| \cdot \sup_{\xi \in B_2^D} |\mathbf{b}^* \xi| \\ &\quad \cdot \sup_{\substack{\mathbf{q} \in B_2^M \\ \tilde{\xi} \in B_2^D}} \left| \sum_{m=1}^M q_m \tilde{\xi}^* \Phi_m^* \mathbf{S} \mathbf{C}_x^* \mathbf{w}_m \right| \\ &\leq K \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \sup_{\substack{\mathbf{q} \in B_2^M \\ \tilde{\xi} \in B_2^D}} \left| \sum_{m=1}^M q_m \tilde{\xi}^* \Phi_m^* \mathbf{S} \mathbf{C}_x^* \mathbf{w}_m \right|. \quad (110) \end{aligned}$$

Note that the objective in the supremum in (110) is rewritten as

$$\begin{aligned} \left| \sum_{m=1}^M q_m \tilde{\xi}^* \Phi_m^* S C_x^* w_m \right| &= \left| \sum_{m=1}^M \bar{q}_m w_m^* C_x S^* \Phi_m \tilde{\xi} \right| \\ &= \left| \left(\sum_{m=1}^M e_m^* \otimes w_m^* C_x S^* \Phi_m \right) (\bar{q} \otimes \tilde{\xi}) \right|. \end{aligned}$$

Then it follows that

$$\begin{aligned} \left\| \sum_{m=1}^M e_m^* \otimes w_m^* C_x S^* \Phi_m \right\|_{S_1 \rightarrow S_2} &\leq C(\beta) \rho_{x,w} \sqrt{M+D} \log K \quad (111) \end{aligned}$$

holds with probability $1 - K^{-\beta}$. The proof of (111) is obtained as we replace $z \in B_2^{2K-1}$ in the proof of Lemma 39 by $[1, \mathbf{0}_{1,2K}]^T$.

The proof completes by plugging in the tail bound in (111) into (110). \square

By collecting these estimates, we obtain that

$$\begin{aligned} \|(I)\|_{S_1 \rightarrow S_\infty} &\leq C(\beta) \rho_{x,w} \|a\|_2 \|b\|_2 \log^\alpha(MKL) \\ &\quad \cdot \left(K \sqrt{M+D} \right. \\ &\quad \left. + \frac{\mu \sqrt{M+D+K} (K + \sqrt{MKD})}{\sqrt{M}} \right) \end{aligned}$$

holds with probability $1 - CK^{-\beta}$. Then the tail estimate of $\|(I)\|_{S_1 \rightarrow S_\infty}$ dominates those for $\|(g)\|$ and $\|(k)\|$. Therefore, we may ignore $\|(g)\|_{S_1 \rightarrow S_\infty}$ and $\|(k)\|_{S_1 \rightarrow S_\infty}$.

Therefore, by plugging in (15), we obtain that with probability $1 - CK^{-\beta}$, the relative perturbation due to $\Phi^* Y_s^* Y_n \Phi$ is upper bounded by

$$\begin{aligned} &\frac{\|\Phi^* Y_s^* Y_n \Phi\|_{S_1 \rightarrow S_\infty}}{K^2 \|x\|_2^2 \|a\|_2^2 \|b\|_2^2} \\ &\leq \frac{C(\beta) \rho_{x,w} \log^\alpha(MKL)}{K^2 \|x\|_2^2 \|a\|_2^2 \|b\|_2^2} \cdot \left(K \sqrt{M+D} \right. \\ &\quad \left. + \frac{\mu \sqrt{M+D+K} (K + \sqrt{MKD})}{\sqrt{M}} \right) \\ &\leq \frac{C'(\beta) \log^\alpha(MKL)}{\sqrt{\eta L}} \cdot \frac{\rho_{x,w}}{\sqrt{K} \sigma_w \|x\|_2} \\ &\quad \cdot \left(\mu \left(\frac{\sqrt{K}}{M} + \sqrt{\frac{D}{M}} + \sqrt{\frac{D}{K}} \right) + 1 \right). \end{aligned}$$

This completes the proof.

ACKNOWLEDGEMENT

K.L. thanks S. Bahmani, Y. Eldar, F. Krahmer, and T. Strohmer for discussions and feedback on an earlier version [51]. The authors appreciate anonymous referees for their constructive comments, which helped improve both the technical results and presentation.

REFERENCES

- [1] K. Lee, N. Tian, and J. Romberg, "Fast and guaranteed blind multichannel deconvolution under a bilinear system model," in *Proc. Inf. Theory Workshop*, Cambridge, U.K., Sep. 2016, pp. 54–56.
- [2] K. Lee, N. Tian, and J. Romberg, "Performance guarantees of spectral methods for passive sensing of multiple channels," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, Aug. 2017, pp. 1–5.
- [3] S. Choudhary and U. Mitra, "Sparse blind deconvolution: What cannot be done," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 3002–3006.
- [4] Y. Li, K. Lee, and Y. Bresler, "Identifiability in blind deconvolution with subspace or sparsity constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4266–4275, Jul. 2016.
- [5] E. Riegler, D. Stotz, and H. Bölcskei, "Information-theoretic limits of matrix completion," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 1836–1840.
- [6] Y. Li, K. Lee, and Y. Bresler, "Optimal sample complexity for blind gain and phase calibration," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5549–5556, Nov. 2016.
- [7] M. Kech and F. Krahmer, "Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems," *SIAM J. Appl. Algebra Geometry*, vol. 1, no. 1, pp. 20–37, 2017.
- [8] Y. Li, K. Lee, and Y. Bresler, "Identifiability in bilinear inverse problems with applications to subspace or sparsity-constrained blind gain and phase calibration," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 822–842, Feb. 2017.
- [9] Y. Li, K. Lee, and Y. Bresler, "Identifiability and stability in blind deconvolution under minimal assumptions," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4619–4633, Jul. 2017.
- [10] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, Mar. 2014.
- [11] G. Tang and B. Recht, "Convex blind deconvolution with random masks," in *OSA Tech. Dig.*, Jun. 2014, paper CW4C.1.
- [12] S. Bahmani and J. Romberg, "Lifting for blind deconvolution in random mask imaging: Identifiability and convex relaxation," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2203–2238, 2015.
- [13] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Blind recovery of sparse signals from subsampled convolution," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 802–821, Feb. 2017.
- [14] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *Appl. Comput. Harmonic Anal.*, to be published, doi: [10.1016/j.acha.2018.01.001](https://doi.org/10.1016/j.acha.2018.01.001).
- [15] Y. Chi, "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 782–794, Jun. 2016.
- [16] D. Yang, G. Tang, and M. B. Wakin, "Super-resolution of complex exponentials from modulations with unknown waveforms," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5809–5830, Oct. 2016.
- [17] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
- [18] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
- [19] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.
- [20] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
- [21] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 134–149, Jan. 1995.
- [22] K. Lee, F. Krahmer, and J. Romberg, (2017). "Spectral methods for passive imaging: Non-asymptotic performance and robustness." [Online]. Available: <https://arxiv.org/abs/1708.04343>
- [23] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.
- [24] V. Cambareri and L. Jacques, (2016). "Through the haze: A non-convex approach to blind gain calibration for linear random sensing models." [Online]. Available: <https://arxiv.org/abs/1610.09028>
- [25] L. Balzano and R. Nowak, "Blind calibration of sensor networks," in *Proc. 6th Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2007, pp. 79–88.

- [26] H. Q. Nguyen, S. Liu, and M. N. Do, "Subspace methods for computational relighting," *Comput. Imag.*, vol. 8657, p. 865703, Feb. 2013, doi: 10.1117/12.2011522.
- [27] R. L. Morrison, M. N. Do, and D. C. Munson, "MCA: A multichannel approach to SAR autofocus," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 840–853, Apr. 2009.
- [28] S. Ling and T. Strohmer. (2016). "Self-calibration and bilinear inverse problems via linear least squares." [Online]. Available: <https://arxiv.org/abs/1611.04196>
- [29] Y. Li, K. Lee, and Y. Bresler, "Blind gain and phase calibration for low-dimensional or sparse signal sensing via power iteration," in *Proc. Int. Conf. Sampling Theory Appl. (SampTA)*, Jul. 2017, pp. 119–123.
- [30] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.
- [31] R. H. Keshavan, "Efficient algorithms for collaborative filtering," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2012.
- [32] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory Comput.*, 2013, pp. 665–674.
- [33] M. Hardt, "Understanding alternating minimization for matrix completion," in *Proc. IEEE 55th Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2014, pp. 651–660.
- [34] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, Nov. 2016.
- [35] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.
- [36] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 739–747.
- [37] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, Feb. 2018.
- [38] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *Ann. Statist.*, vol. 44, no. 5, pp. 2221–2251, 2016.
- [39] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2796–2804.
- [40] I. Waldspurger, "Phase retrieval with random Gaussian sensing vectors by alternating projections," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3301–3312, May 2018.
- [41] R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam, "Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging," *SIAM J. Imag. Sci.*, vol. 8, no. 1, pp. 426–457, 2015.
- [42] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere I: Overview and the geometric picture," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 853–884, Feb. 2017.
- [43] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere II: Recovery by Riemannian trust-region method," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 885–914, Feb. 2017.
- [44] W. Mantzel, J. Romberg, and K. G. Sabra, "Round-robin multiple-source localization," *J. Acoust. Soc. Amer.*, vol. 135, no. 1, pp. 134–147, Jan. 2014.
- [45] N. Tian, S.-H. Byun, K. Sabra, and J. Romberg, "Multichannel myopic deconvolution in underwater acoustic channels via low-rank recovery," *J. Acoust. Soc. Amer.*, vol. 141, no. 5, pp. 3337–3348, 2017.
- [46] W. Mantzel and J. Romberg, "Compressed subspace matching on the continuum," *Inf. Inference*, vol. 4, no. 2, pp. 79–107, 2015.
- [47] X.-T. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 899–925, 2013.
- [48] K. Lee and Y. Bresler, "ADMIRA: Atomic decomposition for minimum rank approximation," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4402–4416, Sep. 2010.
- [49] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 937–945.
- [50] D. Donoho and M. Gavish, "Minimax risk of matrix denoising by singular value thresholding," *Ann. Statist.*, vol. 42, no. 6, pp. 2413–2440, 2014.
- [51] K. Lee, N. Tian, and J. Romberg. (2016). "Fast and guaranteed blind multichannel deconvolution under a bilinear system model." [Online]. Available: <https://arxiv.org/abs/1610.06469>
- [52] R. Adamczak and P. Wolff, "Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order," *Probab. Theory Rel. Fields*, vol. 162, nos. 3–4, pp. 531–586, 2015.
- [53] N. Tian, K. G. Sabra, and J. Romberg, "Blind deconvolution of sources of opportunity in ocean waveguides using bilinear channel models," *J. Acoust. Soc. Amer.*, vol. 142, no. 4, p. 2712, 2017.
- [54] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM J. Numer. Anal.*, vol. 7, no. 1, pp. 1–46, 1970.
- [55] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.
- [56] G. W. Stewart, *Matrix Algorithms: Eigensystems*, vol. 2. Philadelphia, PA, USA: SIAM, 2001.
- [57] P. Å. Wedin, "On angles between subspaces of a finite dimensional inner product space," in *Matrix Pencils*. Berlin, Germany: Springer, 1983, pp. 263–285.
- [58] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Commun. Pure Appl. Math.*, vol. 67, no. 11, pp. 1877–1904, Nov. 2014.
- [59] D. L. Hanson and F. T. Wright, "A bound on tail probabilities for quadratic forms in independent random variables," *Ann. Math. Statist.*, vol. 42, no. 3, pp. 1079–1083, 1971.
- [60] M. Rudelson and R. Vershynin, "Hanson–Wright inequality and sub-Gaussian concentration," *Electron. Commun. Probab.*, vol. 18, no. 82, pp. 1–9, 2013.
- [61] B. Carl, "Inequalities of Bernstein–Jackson-type and the degree of compactness of operators in Banach spaces," *Ann. L'Inst. Fourier*, vol. 35, no. 3, pp. 79–118, 1985.
- [62] M. Junge and K. Lee. (2017). "Generalized notions of sparsity and restricted isometry property. Part I: A unified framework." [Online]. Available: <https://arxiv.org/abs/1706.09410>
- [63] K. Lee and M. Junge. (2015). "RIP-like properties in subsampled blind deconvolution." [Online]. Available: <https://arxiv.org/abs/1511.06146>
- [64] D. G. Feingold and R. S. Varga, "Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem," *Pacific J. Math.*, vol. 12, no. 4, pp. 1241–1250, 1962.

Kiryung Lee received the B.S. and M.S. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2000 and 2002, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign, in 2012. He is currently a Research Scientist II in the School of Electrical and Computer Engineering at Georgia Institute of Technology. He was a Visiting Assistant Professor in the Department of Statistics at the University of Illinois at Urbana-Champaign in 2014–2015. In his earlier career, he worked as a research engineer at the Electronic Telecommunications Research Institute, Daejeon, Korea, and the LG Electronics, Seoul, Korea, in 2002–2004 and 2004–2006, respectively. His research interests are in mathematical theory and computational methods for inverse problems and optimization in imaging, signal processing, statistics, and machine learning.

Ning Tian is currently a Ph.D. candidate and research assistant in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. Ning Tian received the B.S.E.E. (2006), M.S. (2008) degree from Tsinghua University in Beijing, China, and M.S. (2010) degree from Georgia Institute of Technology in Atlanta, Georgia. His current research interests lie at the intersection of digital signal processing, machine learning, and optimization.

Justin Romberg is the Schlumberger Professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology, where he has been on the faculty since 2006. Dr. Romberg received the B.S.E.E. (1997), M.S. (1999) and Ph.D. (2004) degrees from Rice University in Houston, Texas. From Fall 2003 until Fall 2006, he was a Postdoctoral Scholar in Applied and Computational Mathematics at the California Institute of Technology. His current research interests lie at the intersection of statistical signal processing, machine learning, optimization, and applied probability.