Spectral Methods for Passive Imaging: Nonasymptotic Performance and Robustness[*]

Kiryung Lee[†], Felix Krahmer[‡], and Justin Romberg[†]

Abstract. We study the problem of passive imaging through convolutive channels. A scene is illuminated with an unknown, unstructured source, and the measured response is the convolution of this source with multiple channel responses, each of which is time-limited. Spectral methods based on the commutativity of convolution, first proposed and analyzed in the 1990s, provide an elegant mathematical framework for attacking this problem. However, these now classical methods are very sensitive to noise, especially when working from relatively small sample sizes. In this paper, we show that a linear subspace model on the coefficients of the impulse responses of the channels can make this problem well-posed. We derive nonasymptotic error bounds for the generic subspace model by analyzing the spectral gap of the cross-correlation (CC) matrix of the channels relative to the perturbation introduced by noise. Numerical results show that this modified spectral method offers significant improvements over the classical method and outperforms other competing methods for multichannel blind deconvolution.

Key words. passive imaging, blind deconvolution, perturbation analysis, random matrices

AMS subject classifications. 15B52, 93B30, 94A12

DOI. 10.1137/17M1143599

1. Introduction.
We give a rigorous analysis of a passive imaging problem, which can be cast as a multichannel blind deconvolution. A scene is illuminated by an ambient source that we cannot control or observe. This source signal is always active, having no discernible "on" or "off" time, and is unstructured. We observe the convolution of this source with M unknown *channel impulse response* sequences over a window of time. The goal is to estimate this ensemble of impulse responses, which in many applications reveals the structure of the environment being sensed. Problems of this type arise in a wide variety of applications including as opportunistic channel estimation in underwater acoustics [41, 42, 4, 43], seismic interferometry [8], and passive synthetic aperture imaging [13].

As described fully in section 2, this is a multichannel blind deconvolution problem, where we observe the output of a number of linear time-invariant systems all driven by a common source. We will focus entirely on estimating the responses of these system and treat the

[†]School of ECE, Georgia Institute of Technology, Atlanta, GA 30332 (kiryung@ece.gatech.edu, jrom@ece.gatech.edu).

[‡]Department of Mathematics, Technical University of Munich, Garching, Germany (felix.krahmer@tum.de).

(unknown) source signal as a supporting actor whose only role is to help us collect information about these channels.

When the channel impulse responses have a finite length K, then a fundamental technique for performing this estimation, developed in the signal processing literature in the 1990s (see, for example, [47, 37]), is to form a cross-correlation matrix from the channel outputs and then estimate the channel responses by estimating the null space of this matrix. This method is reviewed in section 2.2 below. This classical theory shows that the cross-convolution (CC) method is consistent: as the number of noisy observations we make increases, the channel estimates asymptotically become aligned with the true underlying impulse responses.

From a finite number of samples, the stability of this process, both in theory and in practice, depends critically on the *spectral gap* of this cross-correlation matrix. For even the simplest concrete instances of this problem, this gap tends to be vanishingly small; a typical example is shown later in Figure 2 (and described in sections 2.2 and 2.3 below).

The main contribution of this paper is to show that if an additional structural constraint on the channel responses is imposed, then this spectral gap provably widens, stabilizing this channel estimation procedure. In particular, we constrain the length-K channel responses to be members of a known D dimensional subspace. Enforcing this constraint requires only a straightforward modification to the CC method. Our results show that for a generic D dimensional subspace (i.e., a subspace chosen at random), the principal angle between the true channel responses and their estimates decreases as (a) the number of observations in each channel increases, (b) the signal-to-noise ratio (SNR) of the observations increases, (c) the number of channels increases, and (d) the model becomes more restrictive, meaning D decreases relative to K.

Our analysis of the subspace constrained cross-convolution method (SCCC) uses as its starting point the classical Davis–Kahan bound on the deviation of the eigenvectors computed from a perturbed observation of a positive semidefinite matrix. Bounding the size of this perturbation in terms of the observation noise involves bounding the spectral norms of random matrices with entries given as coupled high order polynomials of sub-Gaussian random variables. These norms are written as the suprema of second order chaos processes, for which there are recently developed concentration results [25, 27]. Application of these concentration results involves computing entropy estimates for various norms. In particular, the entropy estimate for a block norm in Appendix E is a novel result derived using the polytope approximation and polar duality, which might be of independent interest.

The Monte Carlo simulation results in section 4 demonstrate the practical gains that the SCCC method offers over the classical CC method. In practice, the estimator produces accurate results when the number of samples per channel L is a relatively small multiple of K. In this regime, the estimation error scales (as a function of L) in the same manner as the oracle solution, where the source is known and the channels are recovered using standard least squares. We also demonstrate that SCCC outperforms both the classical method and the recently proposed approach in [34] for an underwater acoustics simulation with a realistic (nonrandom) subspace model.

Related work. As mentioned above, the multichannel blind deconvolution problem was studied with intense interest in the signal and image processing literature in the 1990s; the methods most closely related to the work below are described in [47, 18, 37], and good

overviews of general work on this problem can be found in [35, 44]. Many of these algorithms use models on the source signal and channels and develop consistency results under different modeling assumptions; see [19, 14, 48, 17] for representative examples from image processing. To our knowledge, no theoretical results exist for these algorithms when there are a finite number of noisy samples. More recently, necessary and sufficient conditions for the generic identifiability of this problem under various geometric priors have been presented in [32].

A different linearization for the multichannel problem was introduced in [3, 36, 38] and recently studied thoroughly in [34]. The model presented there is different in that the channels are not limited in time, a key piece of beneficial structure that our method exploits. These methods also impose a structural constraint on the source signal, while we view the source signal as unstructured.

Single-channel blind deconvolution of signals belonging to low-dimensional subspaces has also been rigorously studied recently. Identifiability results under various models were studied in [6, 7, 31, 33, 23]. Convex optimization algorithms based on "lifting" were analyzed in [1], followed by a similar result for a gradient descent algorithm [30]. An alternating minimization algorithm for blind deconvolution under sparsity models also with subsampling has been analyzed in [29]. While it is possible to extend these methods to the multichannel scenario, unlike the context of passive imaging, one needs strong geometric priors on both the source and impulse responses. This scenario is different from what we consider in this paper. Simultaneous wavelet estimation and deconvolution of seismic reflection signals [22], autocalibrated parallel imaging [16], and motion deblurring using multiple images [49] are examples of relevant applications of multichannel blind deconvolution with geometric priors.

2. Spectral methods for blind deconvolution. In this section, we formulate the finite impulse response (FIR) multichannel blind deconvolution and describe spectral methods based on the cross convolution.

2.1. Problem formulation. We observe an unknown signal $\boldsymbol{x} \in \mathbb{C}^L$ convolved with multiple unknown channel impulse responses $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m \in \mathbb{C}^L$ with the observations corrupted by additive noise $\boldsymbol{w}_m \in \mathbb{C}^L$:

$$(1) \qquad \boldsymbol{y}_m = \boldsymbol{h}_m \circledast \boldsymbol{x} + \boldsymbol{w}_m, \quad m = 1, \ldots, M,$$

where the convolution $\circledast$ is circular,[1] i.e.,

$$y_m[\ell] = \sum_{k=1}^{L} h_m[k] x[(k - \ell) \bmod L] + w_m[\ell].$$

Our goal is to recover the channel responses $\{\boldsymbol{h}_m\}$ from the observations $\{\boldsymbol{y}_m\}$.

---

[1]In our model, the source is opportunistic and "always on," and so the observations in (1) might better be modeled by a windowed (time-limited) linear convolution. To make a strict correspondence, the source would need to be periodic, which is an additional structural assumption. Having access to the full circular convolution greatly simplifies the analysis, and the discrepancy between these two models is marginal when the number of observations L dominates the length of the impulses responses K. Spectral methods similar to the one presented above that are explicitly based on the time-limited linear convolution are presented in [47, 37]; in practice, small gains might be realized by using these closely related methods.

We will assume that the filters have impulse responses of length K; this simply means that the last $L - K$ entries of each $\boldsymbol{h}_m$ are zero. We denote these nonzero entries using $\underline{\boldsymbol{h}}_m \in \mathbb{C}^K$, with the relation

$$\boldsymbol{h}_m = \boldsymbol{S}^* \underline{\boldsymbol{h}}_m, \quad \boldsymbol{S} = \begin{bmatrix} \mathbf{I}_K & \mathbf{0}_{K,L-K} \end{bmatrix}.$$

The operator $\boldsymbol{S} : \mathbb{C}^L \to \mathbb{C}^K$ restricts a given vector of length L to its subvector with the first K elements. The adjoint $\boldsymbol{S}^* : \mathbb{C}^K \to \mathbb{C}^L$ pads $L - K$ zeros to a given vector of length K.

When $L \geq 2K$, the nonzero terms in the linear convolution $\underline{\boldsymbol{h}}_m * \boldsymbol{x}$ and the circular convolution $\boldsymbol{h}_m \circledast \boldsymbol{x}$ will match, so this model applies to scenarios when we have fixed channels that are being continuously excited by an unknown input, and we observe a "snapshot" of length L of their outputs.

2.2. Cross-convolution method. Our method is a modification of the CC method introduced 20 years ago in [47]. The core idea is simple: we use the fact that multiple convolutions commute with one another to impose a set of linear constraints that the channel responses must obey, and then find the (unique up to scale) set of channels responses that obey these linear constraints. To see how this is done, suppose that the measurements we make are free of noise, $\boldsymbol{y}_m = \boldsymbol{x} \circledast \boldsymbol{h}_m$. Then for any pair of channels m, n

$$(2) \qquad \boldsymbol{y}_m \circledast \boldsymbol{h}_n = \boldsymbol{x} \circledast \boldsymbol{h}_m \circledast \boldsymbol{h}_n = \boldsymbol{y}_n \circledast \boldsymbol{h}_m.$$

Thus the pair of observations $\boldsymbol{y}_n, \boldsymbol{y}_m$ can be used to construct a set of L constraints on the $2K$ variables in the channel coefficients $\underline{\boldsymbol{h}}_n, \underline{\boldsymbol{h}}_m$.

To make this more precise, let $\boldsymbol{T}_{\boldsymbol{v}}$ be the $L \times K$ matrix whose action $\boldsymbol{T}_{\boldsymbol{v}} \underline{\boldsymbol{h}}$ circularly convolves $\boldsymbol{v} = [v_1, \dots, v_L]^\top \in \mathbb{C}^L$ with $\underline{\boldsymbol{h}}$ after zero-padding:

$$\boldsymbol{T}_{\boldsymbol{v}} = \boldsymbol{C}_{\boldsymbol{v}} \boldsymbol{S}^*,$$

where $\boldsymbol{C}_{\boldsymbol{v}} \in \mathbb{C}^{L \times L}$ is a circulant matrix defined by

$$\boldsymbol{C}_{\boldsymbol{v}} := \begin{bmatrix} v_1 & v_L & v_{L-1} & \cdots & v_2 \\ v_2 & v_1 & v_L & \cdots & v_3 \\ \vdots & & & \ddots & \\ v_L & v_{L-1} & v_{L-2} & \cdots & v_1 \end{bmatrix}.$$

Then we can write (2) as

$$\boldsymbol{T}_{\boldsymbol{y}_m} \underline{\boldsymbol{h}}_n - \boldsymbol{T}_{\boldsymbol{y}_n} \underline{\boldsymbol{h}}_m = \boldsymbol{0}.$$

We can represent all $M(M-1)/2$ such constraints in one linear system. With
(3)

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \\ \vdots \\ \boldsymbol{Y}^{(M-1)} \end{bmatrix}, \quad \text{where} \quad \boldsymbol{Y}^{(i)} = \begin{bmatrix} \underbrace{\mathbf{0}_{L,K} \quad \dots \quad \mathbf{0}_{L,K}}_{(M-i) \times (i-1) \text{ blocks}} & \begin{matrix} \boldsymbol{T}_{\boldsymbol{y}_{i+1}} \\ \vdots \\ \boldsymbol{T}_{\boldsymbol{y}_M} \end{matrix} & \underbrace{\begin{matrix} -\boldsymbol{T}_{\boldsymbol{y}_i} & & \\ & \ddots & \\ & & -\boldsymbol{T}_{\boldsymbol{y}_i} \end{matrix}}_{(M-i) \times (M-i) \text{ block diagonal}} \end{bmatrix},$$

we know that $\underline{\boldsymbol{h}} = [\underline{\boldsymbol{h}}_1^\top, \dots, \underline{\boldsymbol{h}}_M^\top]^\top$ will be in the null space of $\boldsymbol{Y}$. Indeed, under the mild condition that the z-transforms of the $\underline{\boldsymbol{h}}_m$ do not share common zeros, the null space of $\boldsymbol{Y}$ is one-dimensional, containing only the scalar multiples of $\underline{\boldsymbol{h}}$ [47].

In any practical scenario, noise (and possibly other perturbations) will keep (2) from holding exactly, and $\boldsymbol{Y}$ will in general not have a null space. The channel estimates, then, are formed by finding the vector that is as close to a null vector as possible; after forming $\boldsymbol{Y}$ from the observations, we solve

$$(4) \qquad \underset{\underline{\boldsymbol{v}} \in \mathbb{C}^{MK}}{\text{minimize}} \; \|\boldsymbol{Y} \underline{\boldsymbol{v}}\|_2^2 \quad \text{subject to} \quad \|\underline{\boldsymbol{v}}\|_2 = 1.$$

The solution to the above is of course given by the eigenvector of $\boldsymbol{Y}^* \boldsymbol{Y}$ corresponding to the smallest eigenvalue.

The matrix $\boldsymbol{Y}$ can be unwieldy for large M; its dimensions are $M(M-1)L/2 \times KM$. However, we can form the smaller $KM \times KM$ matrix $\boldsymbol{Y}^* \boldsymbol{Y}$ in a computationally efficient way using fast convolutions. $\boldsymbol{Y}^* \boldsymbol{Y}$ can be thought of as an $M \times M$ array of $K \times K$ matrices; a quick calculation shows that $K \times K$ block $\boldsymbol{B}_{n,m}$, corresponding to rows $(n-1)K+1$ to nK and columns $(m-1)K+1$ to mK in $\boldsymbol{Y}^* \boldsymbol{Y}$, is given by

$$\boldsymbol{B}_{n,m} = \begin{cases} \sum_{m' \neq m} \boldsymbol{T}_{\boldsymbol{y}_{m'}}^* \boldsymbol{T}_{\boldsymbol{y}_m}, & m = n, \\ -\boldsymbol{T}_{\boldsymbol{y}_m}^* \boldsymbol{T}_{\boldsymbol{y}_n}, & m \neq n. \end{cases}$$

Thus $\boldsymbol{Y}^* \boldsymbol{Y}$ can be computed with $M(M+1)/2$ convolutions of length L. Computing the solution to (4) can be done with an eigenvalue decomposition in $O(M^3 K^3)$ time. For large values of MK, the solution can be computed with the power method, with each application of $\boldsymbol{Y}^* \boldsymbol{Y}$ computed using fast convolutions.

Under certain statistical assumptions on the noise, this estimate is consistent: as $L \to \infty$, the smallest eigenvector of $\boldsymbol{Y}^* \boldsymbol{Y}$ goes to (a scalar multiple of) $\underline{\boldsymbol{h}}$. However, to date there is no rigorous analysis of the *stability* of this procedure and there are no nonasymptotic accuracy bounds that tell us what kind of performance we should expect for a certain number of channels M, filter lengths K, and observation times L. Both of these problems will be addressed in this paper.

The effect of noise on the accuracy of the estimate given by (4) can be understood using the spectral properties of the "noise-free" cross-correlation matrix. We write the noisy measurements as

$$\boldsymbol{y}_m = \boldsymbol{s}_m + \boldsymbol{w}_m, \quad \text{where} \quad \boldsymbol{s}_m = \boldsymbol{h}_m \circledast \boldsymbol{x}.$$

The cross-correlation matrix $\boldsymbol{Y}$ is simply the sum of the cross-correlation matrix $\boldsymbol{Y}_{\text{s}}$ for the signals $\boldsymbol{s}_m$ (i.e., create $\boldsymbol{Y}_{\text{s}}$ as in (3) using $\boldsymbol{T}_{\boldsymbol{s}_i}$ in place of the $\boldsymbol{T}_{\boldsymbol{y}_i}$) and the cross-correlation matrix $\boldsymbol{Y}_w$ for the noise signals $\boldsymbol{w}_m$. The estimate of the channels is formed by solving

$$(5) \qquad \underset{\underline{\boldsymbol{v}} \in \mathbb{C}^{MK}}{\text{minimize}} \; \underline{\boldsymbol{v}}^* \left(\boldsymbol{Y}_{\text{s}}^* \boldsymbol{Y}_{\text{s}} + \boldsymbol{E} \right) \underline{\boldsymbol{v}} \quad \text{subject to} \quad \|\underline{\boldsymbol{v}}\|_2 = 1,$$

where

$$\boldsymbol{E} = \boldsymbol{Y}_{\text{s}}^* \boldsymbol{Y}_{\text{n}} + \boldsymbol{Y}_{\text{n}}^* \boldsymbol{Y}_{\text{s}} + \boldsymbol{Y}_{\text{n}}^* \boldsymbol{Y}_{\text{n}}.$$

From the discussion above, we know that in the noise-free case ($\boldsymbol{E} = \boldsymbol{0}$), we will recover the true channel responses. In expectation, the $\boldsymbol{E}$ matrix becomes a scalar multiple of the identity, and the eigenvectors (and relative order of the eigenvalues) do not change. From a finite number

of samples, how closely the solution to (5) matches the noise-free solution depends on the size of $\boldsymbol{E}$ relative to the *spectral gap* of $\boldsymbol{Y}_s^*\boldsymbol{Y}_s$, which is the size of its second smallest (or smallest nonzero) eigenvalue. This is codified in the classical Davis–Kahan $\sin\theta$-theorem [10].

Theorem 2.1 ($\sin\theta$ theorem [15, Corollary 7.2.6]). *Let* $\boldsymbol{A}, \boldsymbol{E} \in \mathbb{C}^{n\times n}$ *satisfy that* $\boldsymbol{A}$ *and* $\boldsymbol{A} + \boldsymbol{E}$ *are positive semidefinite. Let* $\boldsymbol{q}$ *(resp.,* $\widehat{\boldsymbol{q}}$*) denote the eigenvector of* $\boldsymbol{A}$ *(resp.,* $\boldsymbol{A} + \boldsymbol{E}$*) corresponding to the smallest eigenvalue. Suppose that* $\lambda_{n-1}(\boldsymbol{A}) > \lambda_n(\boldsymbol{A})$*. If*

$$(6) \qquad \|\boldsymbol{E}\| \leq \frac{\lambda_{n-1}(\boldsymbol{A}) - \lambda_n(\boldsymbol{A})}{5},$$

then

$$(7) \qquad \sin\angle(\boldsymbol{q}, \widehat{\boldsymbol{q}}) \leq \frac{4\|\boldsymbol{E}\boldsymbol{q}\|_2}{\lambda_{n-1}(\boldsymbol{A}) - \lambda_n(\boldsymbol{A})}.$$

Remark 2.1. The error bound in the Davis–Kahan theorem is known to be sharp for general perturbations. Recent results in [46, 39] have provided refined bounds for unstructured random perturbations but unfortunately do not apply to our perturbation matrix $\boldsymbol{E}$.

Since the eigenvectors are unit norm, having a bound on angle between them is almost the same as having an error bound (up to a global phase), i.e.,

$$(8) \qquad \sin\angle(\boldsymbol{q}, \widetilde{\boldsymbol{q}}) \leq \min_{\theta \in [0,2\pi)} \|\boldsymbol{q} - e^{\mathrm{i}\theta}\widetilde{\boldsymbol{q}}\|_2 \leq \sqrt{2}\sin\angle(\boldsymbol{q}, \widetilde{\boldsymbol{q}}).$$

As discussed above, when the channels are identifiable, $\lambda_{KM}(\boldsymbol{Y}_s^*\boldsymbol{Y}_s) = 0$, and so we will have guarantees for the robustness of (5) when $\lambda_{KM-1}(\boldsymbol{Y}_s^*\boldsymbol{Y}_s)$ is large compared to $\|\boldsymbol{E}\|$. Unfortunately, this smallest nonzero eigenvalue $\lambda_{KM-1}(\boldsymbol{Y}_s^*\boldsymbol{Y}_s)$ is typically very small in magnitude. Figure 1(a) shows a typical generic example; here we create $\boldsymbol{Y}_s$ from $M = 4$ channels of length $K = 256$, the channel impulse responses themselves were generated at random, and the length of observation was $L = 3K$. In this example, $\lambda_1(\boldsymbol{Y}_s^*\boldsymbol{Y}_s) = 1$ but $\lambda_{KM-1}(\boldsymbol{Y}_s^*\boldsymbol{Y}_s) = 3.3 \times 10^{-5}$, and so we only have robustness guarantees for the mildest perturbations. The ratio $\lambda_{KM-1}(\boldsymbol{Y}_s^*\boldsymbol{Y}_s)/\lambda_1(\boldsymbol{Y}_s^*\boldsymbol{Y}_s)$ increases in L. However, as shown in Figure 1(b), even for a long observation of length $L = 10^5 K$, the improvement is not significant. In fact, the practical performance of the estimator is poor in even a mild amount of additive noise, as the experiments shown later in Figure 2 suggest.

2.3. Subspace-constrained cross-convolution method. In this paper, we show that the introduction of a linear model for the channel responses can tangibly increase the size of this spectral gap. Using a linear subspace to model the channel responses has had some empirical success in the literature. For example, in [43] a data-driven linear model is constructed for underwater acoustic channels for the purpose of ocean tomography.

Along with having an impulse response of limited length, we will also assume that the $\underline{\boldsymbol{h}}_m$ lie in known subspaces of dimension $D < K$. This means that each $\underline{\boldsymbol{h}}_m$ can be expressed as $\underline{\boldsymbol{h}}_m = \boldsymbol{\Phi}_m\boldsymbol{u}_m$, where the columns of $\boldsymbol{\Phi}_m$ form a basis for the model subspace, and the $\boldsymbol{u}_m$ are the expansion coefficients in this basis—recovering the $\boldsymbol{u}_m$ is now the same as recovering the channel responses $\underline{\boldsymbol{h}}_m$. The concatenated channels are written

$$(9) \qquad \underline{\boldsymbol{h}} = \boldsymbol{\Phi}\boldsymbol{u}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Phi}_M \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_M \end{bmatrix}.$$
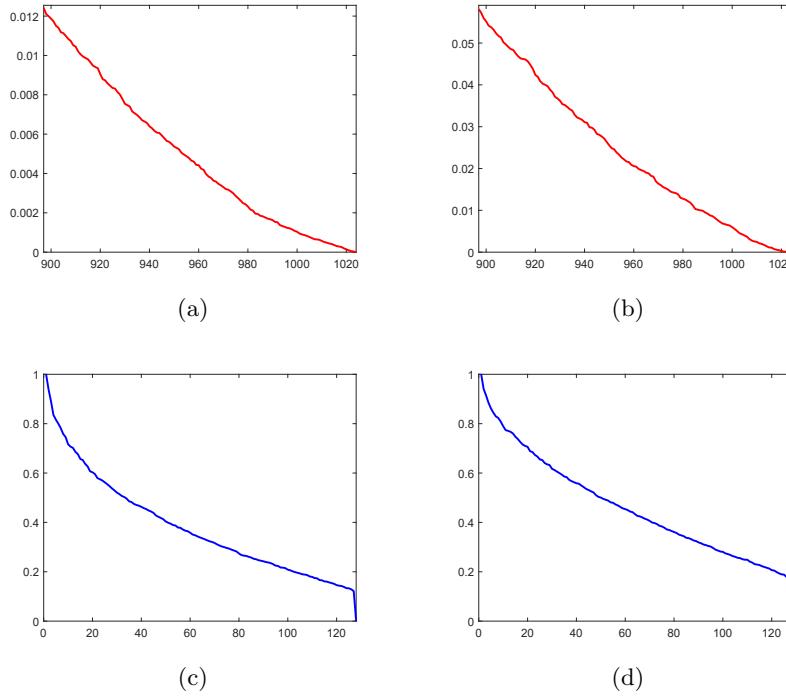
Figure 1. (a), (b) *The eigenvalue spectrum of an example cross-correlation matrix $\boldsymbol{Y}_s^*\boldsymbol{Y}_s \in \mathbb{C}^{KM \times KM}$ created from the observation of length $L = 3K$ (left column) and $L = 10^5 K$ (right column), where the number of channels is $M = 4$ and the channel length is $K = 256$. Eigenvalues are normalized so that the maximum eigenvalue is 1. The smallest 128 eigenvalues are displayed. Note that there is almost no gap between the second smallest eigenvalue and zero. (c), (d) The eigenvalue spectrum of $\boldsymbol{\Phi}^*\boldsymbol{Y}_s^*\boldsymbol{Y}_s\boldsymbol{\Phi} \in \mathbb{C}^{KD \times KD}$ after introducing a linear model of dimension $D = 32$. Note that the spectral gap is now pronounced.*

With this model in place, the channel coefficients $\boldsymbol{u}$ will be in the null space of $\boldsymbol{Y}_s\boldsymbol{\Phi}$.

The estimation procedure has to be modified to account for a slight bias introduced by the linear model. With random uncorrelated noise, $\mathbb{E}[\boldsymbol{w}_m] = \boldsymbol{0}$, $\mathbb{E}[\boldsymbol{w}_m\boldsymbol{w}_m^*] = \sigma_w^2\mathbf{I}$, we have

$$\mathbb{E}[\boldsymbol{Y}_n] = \boldsymbol{0} \quad \text{and} \quad \mathbb{E}[\boldsymbol{Y}_n^*\boldsymbol{Y}_n] = \sigma_w^2(M-1)L\mathbf{I},$$

and so

$$\mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}^*\boldsymbol{Y}\boldsymbol{\Phi}] = \boldsymbol{\Phi}^*\boldsymbol{Y}_s^*\boldsymbol{Y}_s\boldsymbol{\Phi} + \sigma_w^2(M-1)L \cdot \boldsymbol{\Phi}^*\boldsymbol{\Phi}.$$

To make the perturbation from the noise-free cross-correlation matrix zero mean, we will solve

$$(10) \qquad \underset{\boldsymbol{z}\in\mathbb{R}^{MD}}{\text{minimize}} \quad \boldsymbol{z}^*\boldsymbol{\Phi}^*(\boldsymbol{Y}^*\boldsymbol{Y} - \sigma_w^2(M-1)L\mathbf{I}_{MK})\boldsymbol{\Phi}\boldsymbol{z} \quad \text{subject to} \quad \|\boldsymbol{z}\|_2 = 1.$$

Again, the solution is the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{\Phi}^*(\boldsymbol{Y}^*\boldsymbol{Y} - \sigma_w^2(M-1)L\mathbf{I}_{MK})\boldsymbol{\Phi}$.

Remark 2.2. If $\boldsymbol{\Phi}^*\boldsymbol{\Phi} = \mathbf{I}_{MD}$, then since adding a scalar multiple of the identity does not perturb eigenvectors, we may ignore $\sigma_w^2(M-1)L\mathbf{I}_{MK}$ in (10). Otherwise, subtracting the

noise covariance $\sigma_w^2(M-1)L\mathbf{I}_{MK}$ from $\boldsymbol{Y}^*\boldsymbol{Y}$ further suppresses the error in the estimated impulse responses. In practice, the noise variance σ_w^2 needs to be estimated and the error in this estimate will propagate to the estimate of the impulse responses. For simplicity of analysis, we assume that σ_w^2 is known a priori.

Figure 1(c) shows the effect of the subspace constraint on the spectral gap. Here, a generic subspace was chosen by generating $\boldsymbol{\Phi}$ at random. The entries of $\boldsymbol{\Phi} \in \mathbb{C}^{256 \times 8}$ were generated as independent copies of a standard complex Gaussian random variable. The size of the smallest nonzero eigenvalue is now significantly more distinct ($\lambda_{MD-1}(\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi})/\lambda_1(\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}) = 0.4$). As the numerical results in section 4 show, adding subspace constraints of this nature does indeed lead to significant robustness of the method in the presence of noise.

Our main results, detailed in section 3, quantify this spectral gap for generic subspaces $\boldsymbol{\Phi}$.

3. Main results.

3.1. Nonasymptotic analysis.
Our main results give nonasymptotic performance guarantees for the SCCC method. We make the following two assumptions throughout:

(A1) *Generic subspaces.* The performance of our multichannel blind deconvolution method depends on properties of the subspaces underlying linear channel models. We will investigate the "generic" case, where the bases themselves are generated at random. In particular, we will assume that $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_M$ are independent copies of a K-by-D complex Gaussian matrix whose entries are independent and identically distributed (iid) as $\mathcal{CN}(0,1)$. Note that the resulting basis is not exactly orthonormal, but it is very close to an orthonormal basis, as the singular values of a tall Gaussian matrix are known to be well concentrated (e.g., [9]). While it may be more natural to work with orthonormal bases here, this would make the problem more technical, as one would lose independence. For this reason, we will stick to the Gaussian case here. An interesting question for follow-up research will be the conditions required for uniform guarantees; this will then also take care of this issue. Our theorems below hold with high probability with respect to this draw of the $\boldsymbol{\Phi}_m$; we might interpret this as saying that the results hold for "most" subspace models. In section 4 below, we empirically confirm this performance also for choices of $\boldsymbol{\Phi}_m$ with additional structure as it would appear in applications, even though they do not arise from the model analyzed in this paper.

(A2) *Random noise.* The perturbations to the measurements $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M \in \mathbb{C}^L$ are iid sub-Gaussian vectors with $\mathbb{E}[\boldsymbol{w}_m] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{w}_m \boldsymbol{w}_m^*] = \sigma_w^2 \mathbf{I}_L$ and are independent of the bases $\{\boldsymbol{\Phi}_m\}$.

We present two main theorems below. In the first one, we assume that the input itself is a white random process. In the second one, we study deterministic inputs with a mild structural assumption on the common source signal that essentially amounts to $\boldsymbol{x}$ being spread out in the frequency domain; the resulting error bounds are slightly weaker than for the random model.

The theorems provide sufficient conditions on the number of samples L we need to observe at the output of each channel in order to guarantee a certain level of accuracy in the estimate $\hat{\underline{\boldsymbol{h}}}$ found by solving (10) for $\hat{\boldsymbol{u}}$ and then taking $\hat{\underline{\boldsymbol{h}}} = \boldsymbol{\Phi}\hat{\boldsymbol{u}}$. The number of samples we need will

depend on the length of the filter responses K, their intrinsic dimensions D, the number of channels M, and the SNR defined as

$$(11) \qquad \eta := \frac{\mathbb{E}_{\boldsymbol{\phi}}[\sum_{m=1}^{M} \|\boldsymbol{h}_m \circledast \boldsymbol{x}\|_2^2]}{\mathbb{E}_{\boldsymbol{w}}[\sum_{m=1}^{M} \|\boldsymbol{w}_m\|_2^2]}.$$

Under (A1) and (A2), it follows from the commutativity of convolution and Lemma B.1 that η simplifies as

$$(12) \qquad \eta = \frac{K\|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^2}{ML\sigma_w^2}.$$

In addition, the bounds will depend on the channel impulse responses all being roughly the same size. We measure the disparity in impulse response energies using the flatness parameter

$$(13) \qquad \mu := \max_{1 \le m \le M} \frac{\sqrt{M}\|\boldsymbol{u}_m\|_2}{\|\boldsymbol{u}\|_2}.$$

When a small number of the M impulse responses are significantly greater than the others, we will have $\mu \approx \sqrt{M}$. In this case, we expect to have longer observation times, as we are only getting a small number of diverse looks at the signal. Our results are most interesting when μ is a constant on the order of 1. Qualitatively, this means that each channel is roughly as important as the others.

We now present the first of our main results. Theorem 3.1 below assumes a *random common source* signal $\boldsymbol{x}$. The bound on the number of observations L sufficient to guarantee a certain accuracy in the channel estimates is a complicated expression involving the number of channels M, their maximum impulse response lengths K, their intrinsic dimension D, the SNR η, the channel flatness μ, and the level of accuracy ϵ. But in reasonable scenarios where the noise is not extreme (η is a constant), and K and M are not too different, we have

$$\sin \angle(\widehat{\underline{\boldsymbol{h}}}, \underline{\boldsymbol{h}}) \le \epsilon \quad \text{when} \quad L \gtrsim \sqrt{KD}/\epsilon,$$

with the inequality on the right holding to within log factors. As a point of reference, we are estimating MD channel coefficients from ML samples at the outputs; we have the same number of observations as unknowns when $L \gtrsim D$. As $D \le K$, our estimate scales in a mildly unsatisfying way, though as the recovery procedure is highly nonlinear, it is unclear what form an optimal scaling would take.

Theorem 3.1 (random source). *We observe noisy channel outputs $\{\boldsymbol{y}_m\}$ as in (1), with SNR η as in (11), and form an estimate $\underline{\boldsymbol{h}}$ of the channel responses by solving (10). Suppose assumptions (A1) and (A2) above hold, and let $\boldsymbol{x}$ be a sequence of zero-mean iid sub-Gaussian random variables with variance σ_x^2, $\eta \ge 1$, $\mu = O(1)$, and $L \ge 3K$.[2] Then for any $\beta \in \mathbb{N}$,*

---

[2]Without the subspace prior, $L > K$ is necessary to claim that $\boldsymbol{Y}^* \boldsymbol{Y}$ has nullity 1 in the noiseless case. We used $L \ge 3K$ 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.

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,

$$M \geq C_1(\beta) \log^\alpha(MKL), \tag{14}$$

that are sufficiently long,

$$K \geq C_1(\beta)D \log^\alpha(MKL), \tag{15}$$

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

$$L \geq \frac{C_1(\beta) \log^\alpha(MKL)}{\eta}\left(\frac{K}{M^2} + D\right), \tag{16}$$

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

$$\sin \angle(\widehat{\underline{h}}, \underline{h}) \leq C_2(\beta) \log^\alpha(MKL)\left(\frac{1}{\sqrt{\eta L}}\left(\frac{\sqrt{K}}{M} + \sqrt{D}\right) + \frac{\sqrt{D}}{\eta\sqrt{ML}}\right). \tag{17}$$

Remark 3.1. The SNR requirement $\eta \geq 1$ was introduced to simplify the expressions in Theorem 3.1. The conditions in the low SNR regime $\eta < 1$ can be easily extracted from the proof of the theorem and Proposition 3.3 below.

Theorem 3.1 is interpreted as follows. When the dimension D of the subspaces in (9) is small (up to a fraction of the ambient dimension K), the number of channels M is large (depending weakly on the other dimension parameters only through log), and the length of observation is large enough ($L \geq 3K + C\sqrt{KD}$ for an absolute constant C), we can apply the Davis–Kahan theorem which provides an error bound for $\widehat{h}$. The error bound in (17) converges to 0 if either L or η grows toward infinity. Moreover, the error bound is nonasymptotic since it explicitly shows how the error depends on L and η when they are finite. In a heuristic argument that counts the number of unknown parameters and the number of given equations, a necessary condition for the unique recovery of $\boldsymbol{x}$ and $\{\underline{\boldsymbol{h}}_m\}_{m=1}^M$ from noise-free measurements is given as $L \geq MD/(M-1)$. It is unclear whether this is also a valid necessary condition under the finite impulse response structure. On the other hand, with some diversity in $\boldsymbol{x}$ (not necessarily under an explicit stochastic model), it has been shown that $L \geq 3K$ suffices in this scenario. In fact, if $L < K$, then the circular convolution modulo L of $\underline{\boldsymbol{h}}_m \in \mathbb{C}^K$ and $\boldsymbol{x} \in \mathbb{C}^L$ introduces aliasing. This turns the deconvolution problem into the demixing problem of separating a mixture of convolutions. It is not clear whether this still remains a valid necessary condition for the unique identification of the solutions when the extra subspace model in (9) is imposed. Nevertheless, the blind deconvolution approach in this paper does not apply due to the aforementioned aliasing and the requirement $L > K$ is the fundamental limitation of any approach that linearizes the problem using cross convolution.[3]

---

[3]The classical CC approaches (e.g., [47, 37, 18]) formulated the problem using the linear convolution followed by truncation. Since the source is not necessarily time-limited, $L > K$ is also necessary in these classical results.

To prove Theorem 3.1, we establish an intermediate result for the case where the input signal $\boldsymbol{x}$ is deterministic. In this case, our bounds depend on the spectral norm $\rho_{\boldsymbol{x}}$ of the (appropriately restricted) autocorrelation matrix of $\boldsymbol{x}$,

$$\rho_{\boldsymbol{x}} := \|\widetilde{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{S}}^*\|,$$

where

$$(18) \qquad \widetilde{\boldsymbol{S}} = \begin{bmatrix} \mathbf{0}_{K-1,L-K+1} & \mathbf{I}_{K-1} \\ \mathbf{I}_{2K-1} & \mathbf{0}_{2K-1,L-2K+1} \end{bmatrix}.$$

Then the deterministic version of our recovery result is as follows.

Theorem 3.2 (deterministic source). *Suppose that the same assumptions hold as in Theorem 3.1, only with $\boldsymbol{x}$ as a fixed sequence of numbers obeying*

$$(19) \qquad \rho_{\boldsymbol{x}} \leq C_3 \|\boldsymbol{x}\|_2^2.$$

If (15) and (14) hold, and

$$(20) \qquad L \geq \frac{C_1(\beta) \log^{\alpha}(MKL)}{\eta} \Big(\frac{K^2}{M^2} + KD \Big),$$

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

$$(21) \qquad \sin \angle(\widehat{\underline{\boldsymbol{h}}}, \underline{\boldsymbol{h}}) \leq \frac{C_2(\beta) \log^{\alpha}(MKL)}{\sqrt{\eta L}} \Big(\frac{K}{M} + \sqrt{KD} \Big).$$

The condition (19) can be interpreted as a kind of incoherence condition on the input signal $\boldsymbol{x}$. Since

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

where $\widehat{\boldsymbol{x}} \in \mathbb{C}^L$ is the normalized discrete Fourier transform of $\boldsymbol{x}$, it is sufficient that $\hat{\boldsymbol{x}}$ is approximately flat for (19) to hold. This is a milder assumption than imposing an explicit stochastic model on $\boldsymbol{x}$ as in Theorem 3.1. For the price of this relaxed condition, the requirement on L in (20) that activates Theorem 3.2 is more stringent compared to the analogous condition (16) in Theorem 3.1.

Theorems 3.1 and 3.2 distinguish from a recent result [34] in the following sense. Ling and Strohmer [34] analyzed the error bound for the least squares solution to a different linearized formulation in [3]. In their analysis, the unknown filters were assumed to follow stochastic subspace models, which span vectors fully supported on the entire observation period. Obviously, these models do not explain the FIR structures arising in applications. Unlike their analysis, we explicitly considered the case where the unknown filters are supported on a short interval. On the other hand, the number of observations L enabling the error bound in [34] scales near optimally, whereas L grows faster in Theorems 3.1 and 3.2. Again, models considered in these analyses are different and it is still open to verify whether a near optimal scaling can be achieved with limited randomness satisfying the FIR structure.

3.2. Proof of main results. The main results in Theorems 3.1 and 3.2 are obtained by the following proposition. Proposition 3.3 identifies a sufficient condition for (6), which enables Theorem 2.1 and provides an error estimate of computing the most dominant eigenvector from a noisy matrix. The sufficient condition is stated in terms of scaling of key parameters for the sake of interpretation.

Proposition 3.3. *Suppose the assumptions in* (A1) *and* (A2) *hold,* ρ_x *satisfies* (19), $L \geq 3K$, *and* $\mu \leq \sqrt{M}/2$.[4] *Let* $\rho_{x,w}$ *denote the cross correlation among the input* $\boldsymbol{x}$ *and the noise terms* $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M$ *defined by*

$$\rho_{x,w} := \max_{1 \leq m \leq M} \|\widetilde{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \widetilde{\boldsymbol{S}}^*\|,$$

where $\widetilde{\boldsymbol{S}} \in \mathbb{R}^{(3K-2) \times L}$ *is as in* (18). *For any* $\beta \in \mathbb{N}$, *there exist absolute constants* $C > 0, \alpha \in \mathbb{N}$ *and constants* $C_1(\beta), C_2(\beta)$ *that only depend on* β, *for which the following holds. If*

$$(22) \qquad K \geq C_1(\beta) \mu^4 D \log^\alpha(MKL),$$

$$(23) \qquad M \geq C_1(\beta) \mu^4 \log^\alpha(MKL),$$

and

$$(24) \qquad \frac{L}{C_1(\beta) \log^\alpha(MKL)} \geq \frac{\rho_{x,w}^2}{\eta K \sigma_w^2 \|\boldsymbol{x}\|_2^2} \left(\mu^2 \left(\frac{D^2}{K} + \frac{K}{M^2} \right) + D \right) + \frac{D}{\eta^2},$$

then

$$\sin \angle(\widehat{\underline{\boldsymbol{h}}}, \underline{\boldsymbol{h}})$$

$$(25) \qquad \leq C_2(\beta) \log^\alpha(MKL) \left(\frac{\rho_{x,w}}{\sqrt{\eta KL} \sigma_w \|\boldsymbol{x}\|_2} \left(\mu \left(\frac{D}{\sqrt{K}} + \frac{\sqrt{K}}{M} \right) + \sqrt{D} \right) + \frac{\sqrt{D}}{\eta \sqrt{ML}} \right)$$

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

Proof of Proposition 3.3. Recall that we first compute an estimate $\widehat{\boldsymbol{u}}$ of $\boldsymbol{u}$. Then $\widehat{\underline{\boldsymbol{h}}} = \boldsymbol{\Phi} \widehat{\boldsymbol{u}}$ serves as an estimate of $\underline{\boldsymbol{h}} = \boldsymbol{\Phi} \boldsymbol{u}$. Since the estimation error is measured in the principal angle, which is invariant under scalar multiplication, without loss of generality, we may assume that $\|\boldsymbol{u}\|_2 = 1$. Indeed, the errors in the estimates $\widehat{\underline{\boldsymbol{h}}}$ and $\widehat{\boldsymbol{u}}$ are related by

$$(26) \quad \sin \angle(\underline{\boldsymbol{h}}, \widehat{\underline{\boldsymbol{h}}}) = \frac{\|\boldsymbol{P}_{\widehat{\underline{\boldsymbol{h}}}}^\perp \underline{\boldsymbol{h}}\|_2}{\|\underline{\boldsymbol{h}}\|_2} \leq \frac{\|\underline{\boldsymbol{h}} - \widehat{\underline{\boldsymbol{h}}}\|_2}{\|\underline{\boldsymbol{h}}\|_2} \leq \frac{\sigma_{\max}(\boldsymbol{\Phi}) \|\boldsymbol{u} - \widehat{\boldsymbol{u}}\|_2}{\sigma_{\min}(\boldsymbol{\Phi}) \|\boldsymbol{u}\|_2} \leq \frac{\sigma_{\max}(\boldsymbol{\Phi})}{\sigma_{\min}(\boldsymbol{\Phi})} \cdot \sqrt{2} \sin \angle(\boldsymbol{u}, \widehat{\boldsymbol{u}}),$$

where the last step follows from (8).

By the assumption in (A1) and (22), the standard results on singular values of sub-Gaussian matrices (e.g., see [9, Theorem II.13]) imply that the condition number of $\boldsymbol{\Phi}_m$ is

---

[4] By definition, the parameter μ always satisfies $1 \leq \mu \leq \sqrt{M}$. In this perspective, $\mu \leq \sqrt{M}/2$ is a mild condition.

upper bounded by 3 for $m = 1, \ldots, M$ with high probability as we choose $C_1(\beta)$ in (22) large enough. We proceed with the proof under this event. Then the condition number of $\boldsymbol{\Phi}$ is also upper bounded by 3.

Therefore, it suffices to focus on bounding the estimation error in $\widehat{\boldsymbol{u}}$ in the principal angle. Note that $\widehat{\boldsymbol{u}}$ is computed as the least dominant eigenvector of $\widetilde{\boldsymbol{A}} = \boldsymbol{\Phi}^*(\boldsymbol{Y}^*\boldsymbol{Y} - \sigma_w^2(M - 1)L\mathbf{I}_{MK})\boldsymbol{\Phi}$. Furthermore, the target vector $\boldsymbol{u}$ is the unique null vector of $\boldsymbol{A} = \mathbb{E}[\boldsymbol{\Phi}^*(\boldsymbol{Y}^*\boldsymbol{Y} - \sigma_w^2(M - 1)L\mathbf{I}_{MK})\boldsymbol{\Phi}]$.

To see this, we decompose $\boldsymbol{Y}$ as $\boldsymbol{Y} = \boldsymbol{Y}_{\mathrm{s}} + \boldsymbol{Y}_{\mathrm{n}}$, where the noise-free portion $\boldsymbol{Y}_{\mathrm{s}}$ (resp., the noise portion $\boldsymbol{Y}_{\mathrm{n}}$) is obtained as we replace $\boldsymbol{y}_m = \boldsymbol{h}_m \circledast \boldsymbol{x} + \boldsymbol{w}_m$ in $\boldsymbol{Y}$ by its first summand $\boldsymbol{h}_m \circledast \boldsymbol{x}$ (resp., by its second summand $\boldsymbol{w}_m$) for all $m = 1, \ldots, M$. Consequently, we have

$$\mathbb{E}_{\boldsymbol{w}}[\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}}] = \sigma_w^2(M - 1)L\mathbf{I}_{MK} \quad \text{and} \quad \mathbb{E}_{\boldsymbol{w}}[\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{n}}] = \boldsymbol{0}$$

as well as

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{w}}[\widetilde{\boldsymbol{A}}] &= \mathbb{E}_{\boldsymbol{w}}[\boldsymbol{\Phi}^*(\boldsymbol{Y}^*\boldsymbol{Y} - \sigma_w^2(M - 1)L\mathbf{I}_{MK})\boldsymbol{\Phi}] \\
&= \boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi} + \boldsymbol{\Phi}^*\mathbb{E}_{\boldsymbol{w}}[\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{n}}]\boldsymbol{\Phi} + \boldsymbol{\Phi}^*\mathbb{E}_{\boldsymbol{w}}[\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{s}}]\boldsymbol{\Phi} \\
&\quad + \mathbb{E}_{\boldsymbol{w}}[\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \sigma_w^2(M - 1)L\mathbf{I}_{MK})\boldsymbol{\Phi}] \\
&= \boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}.
\end{aligned}$$

As shown in section 2.2, by the construction of $\boldsymbol{Y}_{\mathrm{s}}$, the vector $\underline{\boldsymbol{h}} = \boldsymbol{\Phi}\boldsymbol{u}$ with the true filter coefficients is in the null space of $\boldsymbol{Y}_{\mathrm{s}}$. Therefore, $\boldsymbol{u}$ is almost surely a null vector of the noise-free matrix $\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}$ and hence also of its expectation $\boldsymbol{A} = \mathbb{E}[\widetilde{\boldsymbol{A}}] = \mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}]$. The uniqueness follows from the first part of the following lemma, which is proved in section 5.1.

Lemma 3.4. *Under the hypothesis of Proposition 3.3, the following are true:* (i) *the nullity of $\mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}]$ is 1;* (ii) *nonzero eigenvalues of $\mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}]$ are no less than $K^2\|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^2/2 =: \delta$.*

This lemma also establishes a lower bound for the gap between the two smallest eigenvalues of $\boldsymbol{A}$. This spectral gap allows us to distinguish the corresponding eigenspaces of $\boldsymbol{A}$. Provided condition (6), that is, $\widetilde{\boldsymbol{A}}$ does not deviate too much from its expectation $\boldsymbol{A}$ in the spectral norm (this will be the main task of the remainder of this proof), this property also carries over to the eigenspaces of $\widetilde{\boldsymbol{A}}$ and it follows from Theorem 2.1 that the least dominant eigenspaces of $\widetilde{\boldsymbol{A}}$ and $\boldsymbol{A}$ are close to each other. Thus, up to a global phase, $\widehat{\boldsymbol{u}}$ is a good estimate of $\boldsymbol{u}$.

It remains to show that condition (6) is satisfied with high probability. To this end, we derive a tail estimate of the spectral norm of the random perturbation $\boldsymbol{E} = \widetilde{\boldsymbol{A}} - \boldsymbol{A}$ and show that the perturbation relative to the spectral gap δ satisfies

$$
\begin{aligned}
\frac{\|\boldsymbol{E}\|}{\delta} &\leq C(\beta)\log^\alpha(MKL)\Bigg(\left(\sqrt{\frac{1}{M}} + \sqrt{\frac{D}{K}}\right)\mu^2 \\
&\quad + \frac{\rho_{x,w}}{\sqrt{\eta KL}\sigma_w\|\boldsymbol{x}\|_2}\left(\mu\left(\frac{D}{\sqrt{K}} + \frac{\sqrt{K}}{M}\right) + \sqrt{D}\right) + \frac{\sqrt{D}}{\eta\sqrt{L}}\Bigg)
\end{aligned}
\tag{27}
$$

with probability $1 - CK^{-\beta}$, where $C(\beta)$ is a constant depending only on β. By choosing $C_1(\beta)$ in (22), (23), and (24) large enough, we can make the right-hand side of (27) less than $1/5$. Thus (6) is satisfied.

The derivation of (27) is rather involved for the following reasons. The entries of the perturbation matrix $\boldsymbol{E}$ are given as fourth order polynomials of sub-Gaussian random variables. In addition, the convolution structure in the construction of $\boldsymbol{Y}$ creates dependence relations between the matrix entries. To analyze the perturbation, we decompose $\boldsymbol{E}$ into three components of different polynomial order as follows:

$$\boldsymbol{E} = \underbrace{\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi} - \mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}]}_{\boldsymbol{E}_s}$$

$$+ \underbrace{\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{n}}\boldsymbol{\Phi}}_{\boldsymbol{E}_c} + \boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}$$

$$+ \underbrace{\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \sigma_w^2(M-1)L\mathbf{I}_{MK})\boldsymbol{\Phi}}_{\boldsymbol{E}_n}.$$

The following lemmas, the proofs of which will be presented in section 5, provide tail estimates of the components; the tail estimate in (27) is then obtained by combining these results via the triangle inequality.

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*

$$(28) \qquad \frac{\|\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi} - \mathbb{E}[\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}]\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq C(\beta)\log^\alpha(MKL)\left(\sqrt{\frac{1}{M}} + \sqrt{\frac{D}{K}}\right)\mu^2$$

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

Lemma 3.6. *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* $\boldsymbol{w}$,

$$(29) \qquad \frac{\|\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{n}}\boldsymbol{\Phi}\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq \frac{C(\beta)\rho_{x,w}}{\sqrt{\eta KL}\sigma_w\|\boldsymbol{x}\|_2}\left(\mu\left(\frac{D}{\sqrt{K}} + \frac{\sqrt{K}}{M}\right) + \sqrt{D}\right)$$

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

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*

$$(30) \qquad \frac{\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \sigma_w^2(M-1)L\mathbf{I}_{MK})\boldsymbol{\Phi}\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq \frac{C(\beta)\log^\alpha(MKL)}{\eta} \cdot \sqrt{\frac{D}{L}}$$

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

Finally, under the event where (6) is satisfied, Theorem 2.1 implies that

$$(31) \qquad \sin(\boldsymbol{u}, \widehat{\boldsymbol{u}}) \leq \frac{4\|\boldsymbol{E}\boldsymbol{u}\|_2}{\delta\|\boldsymbol{u}\|_2}.$$

To estimate the right-hand side, we again decompose $\boldsymbol{E}$ as in (28), so the triangle inequality yields

$$(32) \qquad \|\boldsymbol{E}\boldsymbol{u}\|_2 \leq \|\boldsymbol{E}_{\mathrm{s}}\boldsymbol{u}\|_2 + \|\boldsymbol{E}_{\mathrm{c}}\boldsymbol{u}\|_2 + \|\boldsymbol{E}_{\mathrm{c}}^*\boldsymbol{u}\|_2 + \|\boldsymbol{E}_{\mathrm{n}}\boldsymbol{u}\|_2.$$

To bound the first term, recall that $\boldsymbol{u}$ is in the null space of $\boldsymbol{\Phi}^* \boldsymbol{Y}_\text{s}^* \boldsymbol{Y}_\text{s} \boldsymbol{\Phi}$, so we obtain that

$$(33) \qquad \boldsymbol{E}_\text{s} \boldsymbol{u} = \boldsymbol{0}.$$

For the second and third summands, Lemma 3.6 yields that with probability $1 - CK^{-\beta}$

$$(34)$$

$$\max(\|\boldsymbol{E}_\text{c} \boldsymbol{u}\|_2, \|\boldsymbol{E}_\text{c}^* \boldsymbol{u}\|_2) \le \|\boldsymbol{E}_\text{c}\| \|\boldsymbol{u}\|_2 \le \frac{C(\beta) \rho_{x,w} K^{3/2} \|\boldsymbol{x}\|_2 \|\boldsymbol{u}\|_2^3}{\sqrt{\eta L} \sigma_w} \left(\mu \left(\frac{D}{\sqrt{K}} + \frac{\sqrt{K}}{M} \right) + \sqrt{D} \right).$$

A bound for the last summand is provided by the following lemma, which is proved in section 5.5.

Lemma 3.8. *Suppose that* (A1) *holds. For any* $\beta \in \mathbb{N}$, *there is a constant* $C(\beta)$ *that depends only on* β *such that*

$$(35) \qquad \frac{\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \sigma_w^2(M-1)L\mathbf{I}_{MK})\boldsymbol{\Phi}\boldsymbol{u}\|_2}{K^2 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^4} \le \frac{C(\beta) \log^\alpha(MKL)}{\eta} \cdot \sqrt{\frac{D}{ML}}$$

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

Inserting the bounds for the four summands into (32) yields the error bound in (25), which completes the proof. ∎

In the remainder of this section, we show how Theorems 3.1 and 3.2 can be deduced from Proposition 3.3.

Proof of Theorem 3.2. Since most assumptions of the theorem agree with the ones of Proposition 3.3 it only remains to bound $\rho_{x,w}$. This is achieved by the following lemma, which is proved in Appendix C.1.

Lemma 3.9. *Suppose* (A2) *holds and let* $\boldsymbol{x}$ *be a fixed sequence of numbers obeying* (19). *For any* $\beta \in \mathbb{N}$, *there exists an absolute constant* C *such that*

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

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

The theorem follows from a direct application of Proposition 3.3. ∎

Proof of Theorem 3.1. We again need to show a bound for $\rho_{x,w}$, but in addition we need to estimate ρ_x, as it is not part of the assumptions. The following lemma, which is proved in Appendix C.2, provides these two bounds.

Lemma 3.10. *Suppose* (A2) *holds and let* $\boldsymbol{x}$ *be a sequence of zero-mean iid sub-Gaussian random variables with variance* σ_x^2. *Then*

$$\frac{\rho_x}{\|\boldsymbol{x}\|_2^2} \le \frac{L + C_\beta \sqrt{KL} \log^5(KL)}{L - \sqrt{2L\beta \log K}}$$

and

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

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

Again, the theorem follows from a direct application of Proposition 3.3.

4. Numerical results. We compare the classical CC method and our modification with additional subspace prior, which is the SCCC method in a set of Monte Carlo simulations.

Our first tests concern the random signal model of Theorem 3.1 with additional subspace constraints. As expected, our method SCCC, which exploits this information, significantly outperforms the original CC, which does not; see Figure 2. Specifically, the estimation error monotonically decreases (resp., increases) with L and M (resp., D). This is consistent with the prediction in Theorem 3.1.

In order to make the comparison fair, we performed another experiment where CC and SCCC estimate the impulse responses following their respective models having the same number of parameters. CC assumes that the length of channels is $K = 64$. On the other hand, SCCC recovers channels of length $K = 256$ whose coefficients belong to generic subspaces of dimension $D = 64$. In Figure 3, besides the two models involving the same number of parameters, there is significant difference in the estimation error.

Figure 4 compares the empirical phase transition of SCCC and the least squares approach in the nonblind case (where $\boldsymbol{x}$ is known). As in the limit when M goes to infinity, (16) simplifies to $L/K \gtrsim \sqrt{D/K}$, and we provide the plot in terms of the quantities D/K and L/K. In the case of noisy measurements, our performance measure is the 95th percentile of the estimation error, i.e., we consider the worst case but ignore those 5% of the instances which performed worst. In Figure 4, we compare this error to the nonblind case Our color
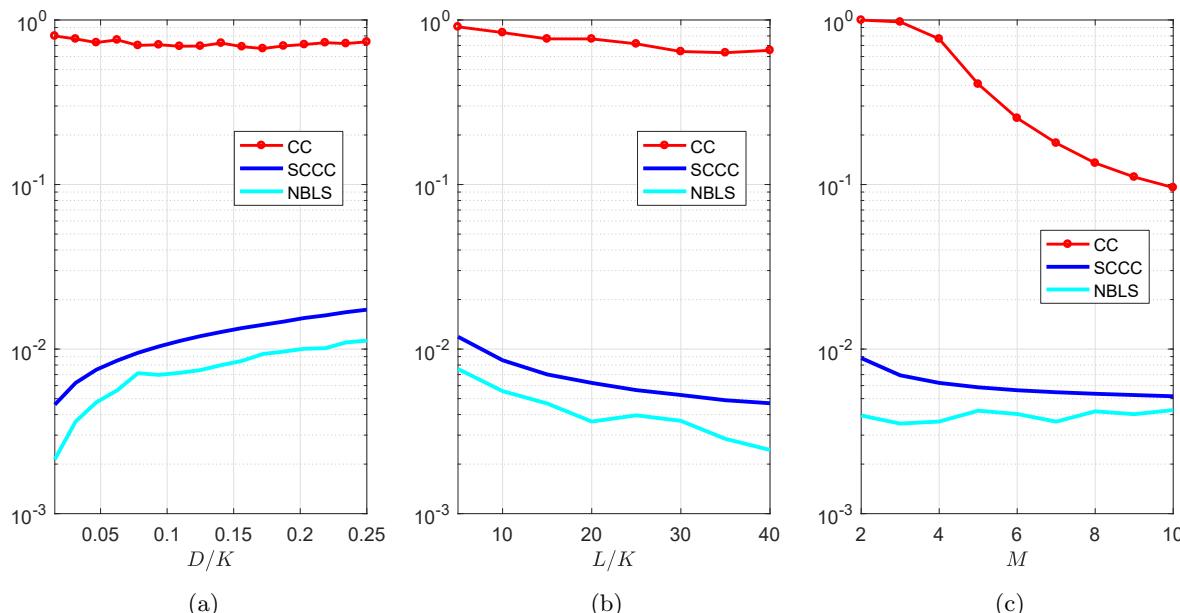
Figure 2. *Comparison of CC and SCCC. The 95th percentile of the estimation error $\sin \angle(\widehat{\boldsymbol{h}}, \boldsymbol{h})$ out of 1,000 trials with iid Gaussian basis is plotted. Default parameters: $K = 256$, $M = 4$, $D = 8$, $L = 20K$, $SNR = 20dB$. (a) For different dimensions. (b) For different observation lengths. (c) For different numbers of channels.*
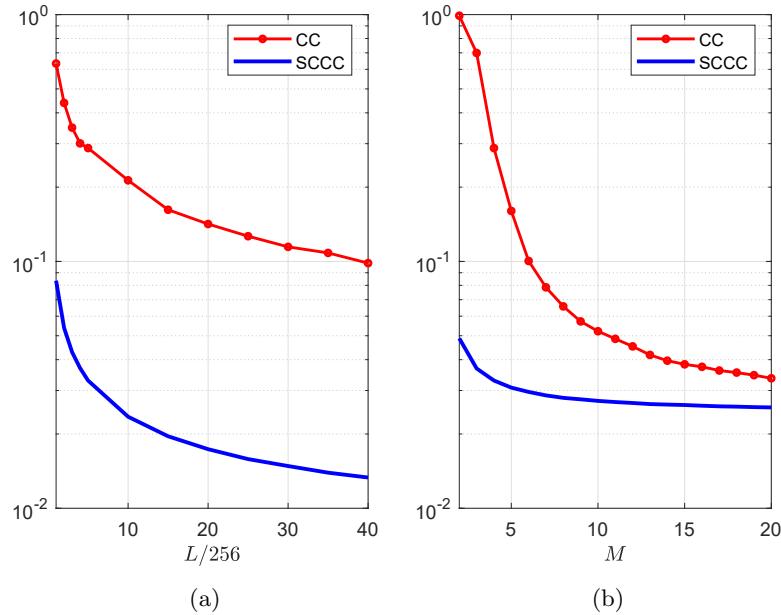
Figure 3. *Comparison of CC ($D = K = 64$) and SCCC ($D = 64$, $K = 256$) with the same number of unknown parameters. Estimation error $\sin \angle(\widehat{\boldsymbol{h}}, \boldsymbol{h})$ corresponding to the 95th percentile out of $1{,}000$ trials with iid Gaussian basis is plotted. Default parameters: $L = 1{,}280$, $M = 4$, $SNR = 20dB$. (a) For different observation lengths. (b) For different numbers of channels.*

coding uses a logarithmic scale with blue denoting the smallest and red the largest estimation error within the regime. We observe that, unlike the nonblind case, SCCC totally fails when D/K is larger than a certain threshold determined by M. This phenomenon clearly justifies the need to introduce a strong subspace prior that stabilizes the eigenvector estimation.

Finally we study data synthesized with a parametric channel impulse response model in underwater acoustics and apply SCCC under a subspace model obtained by principal component analysis (PCA) [43]. More precisely, we employed the ray-approximation model in [43], where each channel impulse response is obtained by sampling a superposition of shifts of a Gaussian-windowed pulse $p(t) = \sin(2\pi f_c t)e^{-(t/\tau)^2}$ with $f_c = 2$ kHz and the effective bandwidth of 400–600 Hz. When the arrival times that amount to the shifts (not necessarily on a uniform grid) are known up to respective uncertainty intervals, a linear approximation of the parametric model is obtained by applying PCA to random realizations (see [43, section III] for further details.) The source signal, subspace coefficients, and additive noise were generated as iid Gaussian random variables.

Figure 5 compares various methods in the case where the arrival times in the channels are aligned on a simple line with its slope known up to a given uncertainty interval. SCCC is compared to the classical CC method as well as to the least squares approach to a differently linearized formulation [3], which also incorporates the same prior models on the impulse responses. As shown in Figure 5, SCCC outperforms CC and LS in this scenario. Although the assumptions of Theorem 3.1 are not satisfied, similarly to the previous experiment, the estimation error for SCCC monotonically decreases with L and M. The other two methods
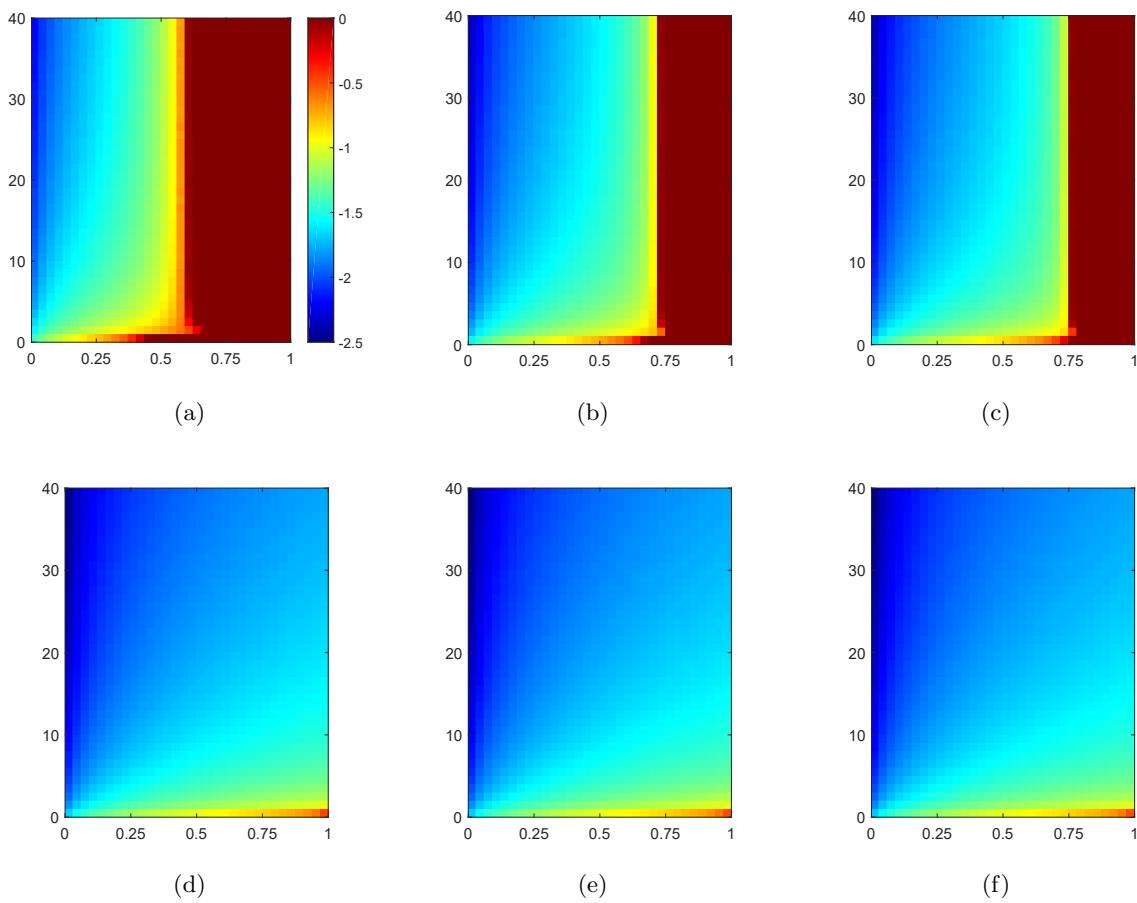
Figure 4. *Estimation error for SCCC (upper row) and nonblind deconvolution with known $\boldsymbol{x}$ (bottom row). (95th percentile of the log of the estimation error for 1,000 trials). Independent and identically distributed Gaussian basis. x-axis: D/K. y-axis: L/K. $K = 256$, $SNR = 20dB$. (a), (d) $M = 2$. (b), (e) $M = 4$. (c), (f) $M = 6$.*

did not perform satisfactorily even under a very high SNR of 80 dB. We have already explained why the classical method fails in terms of the spectral gap. The least squares method, which recovers both the input and filters simultaneously, was not successful because the system of convolution with multiple channels is highly ill-conditioned. Even when the filters are known, the corresponding system has condition number typically larger than 5,000. This happened since the known continuous function is close to a strict band-pass filter and the unknown signal has a white spectrum. Figure 5(c) demonstrates that even under moderate SNRs, SCCC provides stable recovery whereas the other methods totally failed in this regime.

Our study was motivated by underwater acoustic channel estimation in a shallow-water environment where collected measurements are severely corrupted with various sources. As a consequence, the ray-approximation model with a known pulse shape, which is employed in the state-of-the-art method in underwater acoustics [43], does not yet match the obtained data sufficiently well. For this reason, we do not yet obtain convincing recovery from real
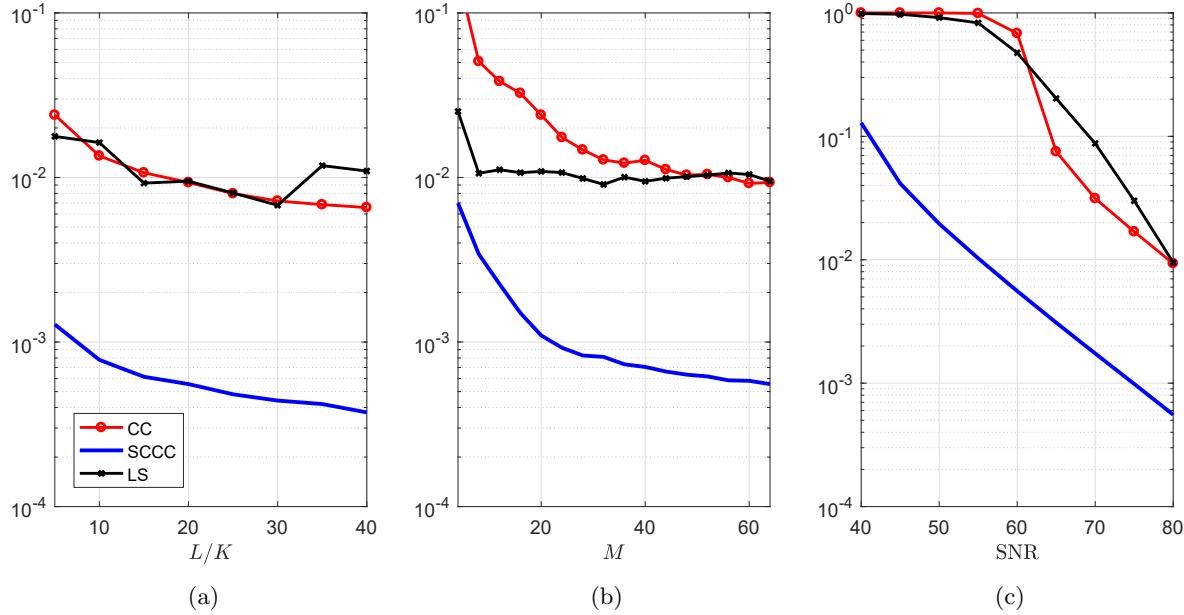
Figure 5. *Estimation error* $\sin \angle(\widehat{\underline{h}}, \underline{h})$ *for CC and SCCC (95th percentile for 1,000 trials). PCA basis. Default parameters:* $K = 64$, $M = 64$, $D = 6$, $L = 20K$, $SNR = 80dB$. (a) *For different observation lengths.* (b) *For different numbers of channels.* (c) *For different SNR.*

hydrophone measurements. We expect, however, that more advanced models to be developed in the future can also be incorporated in our recovery method, allowing us to move steadily toward convincing performance on real data. To illustrate that our results provide a step toward this goal, we provide an additional set of simulations for synthesized signals where the subspace model is constructed with the ray-approximation model by using the estimated arrival-time intervals from real measurements [42]. While the matrix Y_s in this experiment is highly ill-conditioned with nearly half of the eigenvalues almost zero, we demonstrate that the subspace model from the ray-approximation enables stable reconstruction, where the estimate is as good as the result by the state of the art in the underwater acoustics literature. The estimated arrival-time intervals are illustrated in Figure 6(a). From the subspace model constructed from this prior information, we synthesized noisy measurements of SNR 30 dB from a generic signal and generic impulse responses. Then we compared the reconstruction of the impulse responses in 16 channels by the three methods: SCCC, CC, and a low-rank recovery approach [43]. While CC failed in this scenario, both SCCC and the low-rank recovery provided accurate estimates as shown in Figures 6(c) and 6(d). The computational cost of SCCC is much lower than that of the low-rank recovery. Furthermore, no theoretic analysis has been shown for the latter yet.

5. Proof of key lemmas. In this section, we prove some important lemmas required in the proofs of our main theorems. For this it will be of particular importance to compute tail estimates of the spectral norms of several structured random matrices with entries given as functions of second order polynomials of Gaussian variables. In certain cases, such bounds

(a)

(b)

(c)

(d)
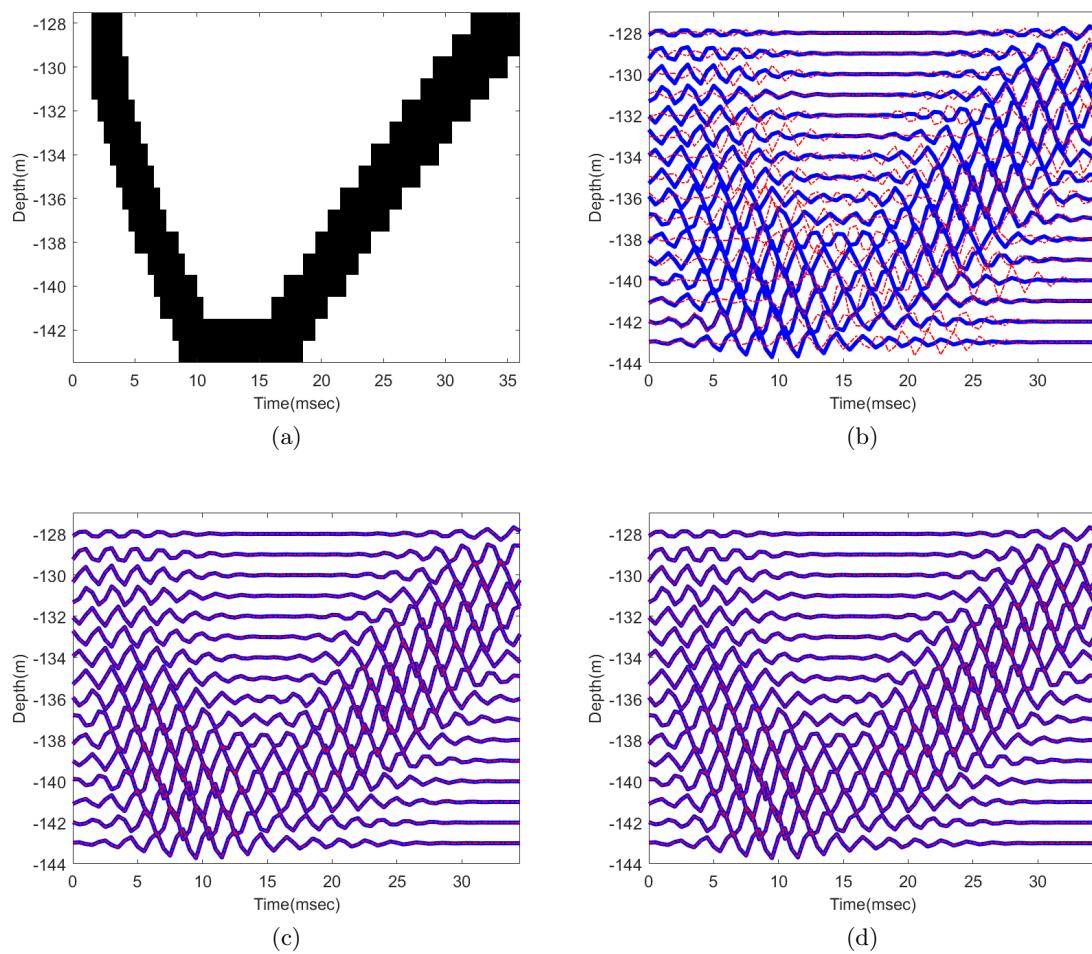
Figure 6. (a) *Arrival-time intervals estimated from real data.* (b), (c), (d) *Comparison of the synthesized ground-truth impulse responses (blue-solid) and the reconstruction (red-dash-dot);* (b) *CC.* (c) *SCCC.* (d) *Low-rank recovery.*

can be established via matrix concentration inequalities (e.g., [45, 24, 21]). These matrix concentration inequalities, however, require an upper estimate of spectral norms of summands and high order moments, the computation of which turns out to be rather complicated for those matrices arising in the proofs. Moreover, there are cases where these inequalities do not apply as the target matrix is not decomposed as a sum of independent variables. For these reasons, we rather proceed by writing the spectral norms of these random matrices in a variational form as suprema of corresponding chaos processes. Tail estimates for such suprema of second order chaos processes, as given in the following theorem, have been derived using chaining arguments [25].

Theorem 5.1 (tail estimates for suprema of chaos processes [25, Theorem 3.1]). *Let $\boldsymbol{\xi} \in \mathbb{C}^n$ be an L-sub-Gaussian vector with $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^*] = \mathbf{I}_n$, $\Delta \subset \mathbb{C}^{m \times n}$, and $0 < \zeta < 1$. There exists a constant $C(L)$ that only depends on L such that*

$$\sup_{\boldsymbol{M} \in \Delta} |\|\boldsymbol{M}\boldsymbol{\xi}\|_2^2 - \mathbb{E}[\|\boldsymbol{M}\boldsymbol{\xi}\|_2^2]| \leq C(L)(K_1 + K_2\sqrt{\log(2\zeta^{-1})} + K_3\log(2\zeta^{-1}))$$

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

$$K_1 := \gamma_2(\Delta, \|\cdot\|)\Big(\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)\Big) + d_{\mathrm{F}}(\Delta)d_{\mathrm{S}}(\Delta),$$
$$K_2 := d_{\mathrm{S}}(\Delta)(\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)),$$
$$K_3 := d_{\mathrm{S}}^2(\Delta).$$

Using the polarization identity, this result on the suprema of second order chaos processes has been extended from a sub-Gaussian quadratic form to a sub-Gaussian bilinear form [27].

Theorem 5.2 (a corollary of [27, Theorem 2.3]). *Let* $\boldsymbol{\xi} \in \mathbb{C}^n$ *be an* L-*sub-Gaussian vector with* $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^*] = \mathbf{I}_n$, $\Delta_2, \Delta_1 \subset \mathbb{C}^{m \times n}$, $0 < \zeta < 1$, *and* $a > 0$. *There exists a constant* $C(L)$ *that only depends on* L *such that*

$$\sup_{\boldsymbol{M}_1 \in \Delta_1, \boldsymbol{M}_2 \in \Delta_2} \big|\langle \boldsymbol{M}_1\boldsymbol{\xi}, \boldsymbol{M}_2\boldsymbol{\xi}\rangle - \mathbb{E}[\langle \boldsymbol{M}_1\boldsymbol{\xi}, \boldsymbol{M}_2\boldsymbol{\xi}\rangle]\big| \leq C(L)(\widetilde{K}_1 + \widetilde{K}_2\sqrt{\log(8\zeta^{-1})} + \widetilde{K}_3\log(8\zeta^{-1})),$$

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

$$\widetilde{K}_1 := \Big(a\gamma_2(\Delta_1, \|\cdot\|) + a^{-1}\gamma_2(\Delta_2, \|\cdot\|)\Big)\Big(a\gamma_2(\Delta_1, \|\cdot\|) + a^{-1}\gamma_2(\Delta_2, \|\cdot\|)$$
$$+ ad_{\mathrm{F}}(\Delta_1) + a^{-1}d_{\mathrm{F}}(\Delta_2)\Big) + \Big(ad_{\mathrm{F}}(\Delta_1) + a^{-1}d_{\mathrm{F}}(\Delta_2)\Big)\Big(ad_{\mathrm{S}}(\Delta_1) + a^{-1}d_{\mathrm{S}}(\Delta_2)\Big),$$
$$\widetilde{K}_2 := \Big(ad_{\mathrm{S}}(\Delta_1) + a^{-1}d_{\mathrm{S}}(\Delta_2)\Big)\Big(a\gamma_2(\Delta_1, \|\cdot\|) + a^{-1}\gamma_2(\Delta_2, \|\cdot\|) + ad_{\mathrm{F}}(\Delta_1) + a^{-1}d_{\mathrm{F}}(\Delta_2)\Big),$$
$$\widetilde{K}_3 := \Big(ad_{\mathrm{S}}(\Delta_1) + a^{-1}d_{\mathrm{S}}(\Delta_2)\Big)^2.$$

A special case of Theorem 5.2 where $a = 1$ was shown in [27, Theorem 2.3]. Note that the bilinear form satisfies

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

Moreover, the γ_2 functional and the radii with respect to the Frobenius and spectral norms are all 1-homogeneous functions. Therefore, Theorem 5.2 is a direct consequence of [27, Theorem 2.3].

Since $a > 0$ in Theorem 5.2 is arbitrary, one can minimize the tail estimate over $a > 0$.

5.1. Proof of Lemma 3.4. Note that $\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi} \in \mathbb{C}^{MD \times MD}$ is an M-by-M block matrix, where the size of each block is D-by-D. Then it follows from the special structure of $\boldsymbol{Y}$ (with each row consisting only of some $\boldsymbol{T}_{\boldsymbol{y}_k}$ in position j and the corresponding $-\boldsymbol{T}_{\boldsymbol{y}_j}$ in position k) that the (m, m)th diagonal block of $\boldsymbol{\Phi}^*\boldsymbol{Y}_{\mathrm{s}}^*\boldsymbol{Y}_{\mathrm{s}}\boldsymbol{\Phi}$ is given by

$$(36) \qquad \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{\Phi}_m^* \boldsymbol{T}_{\boldsymbol{x} \circledast \boldsymbol{h}_{m'}}^* \boldsymbol{T}_{\boldsymbol{x} \circledast \boldsymbol{h}_{m'}} \boldsymbol{\Phi}_m = \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}} \widetilde{\boldsymbol{\Phi}}_m,$$

where $\widetilde{\boldsymbol{\Phi}}_m = \boldsymbol{S}^* \boldsymbol{\Phi}_m$ for $m = 1, \ldots, M$. Similarly, the (m, m')th off-diagonal block of $\boldsymbol{\Phi}^* \boldsymbol{Y}_{\mathrm{s}}^* \boldsymbol{Y}_{\mathrm{s}} \boldsymbol{\Phi}$ for $m \neq m'$ is given by

$$(37) \qquad -\boldsymbol{\Phi}_m^* \boldsymbol{T}_{\boldsymbol{x} \circledast h_{m'}}^* \boldsymbol{T}_{\boldsymbol{x} \circledast h_m} \boldsymbol{\Phi}_{m'} = -\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_{m'}.$$

By Lemma B.3, the expectation of the (m, m)th diagonal block is given by

$$\mathbb{E} \left[\sum_{\substack{m'=1 \\ m' \neq m}}^M \widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}} \widetilde{\boldsymbol{\Phi}}_m \right] = \sum_{\substack{m'=1 \\ m' \neq m}}^M \|\boldsymbol{u}_{m'}\|_2^2 K^2 \|\boldsymbol{x}\|_2^2 \mathbf{I}_D$$

and a similar calculation yields that the expectation of the (m, m')th diagonal block for $m' \neq m$ is given by

$$-\mathbb{E} \left[\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_{m'} \right] = -K^2 \|\boldsymbol{x}\|_2^2 \boldsymbol{u}_m \boldsymbol{u}_{m'}^*.$$

Collecting the above expectations, we obtain that $\mathbb{E}[\boldsymbol{\Phi}^* \boldsymbol{Y}_{\mathrm{s}}^* \boldsymbol{Y}_{\mathrm{s}} \boldsymbol{\Phi}]$ satisfies

$$\frac{\mathbb{E}[\boldsymbol{\Phi}^* \boldsymbol{Y}_{\mathrm{s}}^* \boldsymbol{Y}_{\mathrm{s}} \boldsymbol{\Phi}]}{K^2 \|\boldsymbol{x}\|_2^2} = \|\boldsymbol{u}\|_2^2 \boldsymbol{P}_{\boldsymbol{u}^\perp} - \boldsymbol{\Upsilon},$$

where

$$\boldsymbol{\Upsilon} = \sum_{m=1}^M \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \|\boldsymbol{u}_m\|_2^2 \boldsymbol{P}_{\boldsymbol{u}_m^\perp},$$

where $\otimes$ denotes the Kronecker product and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M$ denote the standard basis vectors in $\mathbb{R}^M$.

Since $\boldsymbol{\Upsilon} \boldsymbol{u} = \boldsymbol{0}$ and hence $\boldsymbol{u}$ is in the kernel of $\mathbb{E}[\boldsymbol{\Phi}^* \boldsymbol{Y}_{\mathrm{s}}^* \boldsymbol{Y}_{\mathrm{s}} \boldsymbol{\Phi}]$, the gap between its two smallest eigenvalues is given by its smallest nonzero eigenvalue, which, using that $\|\boldsymbol{\Upsilon}\| \leq \max_{1 \leq m \leq M} \|\boldsymbol{u}_m\|_2^2$, is estimated from below by

$$(38) \qquad K^2 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^2 \left(1 - \frac{\max_{1 \leq m \leq M} \|\boldsymbol{u}_m\|_2^2}{\|\boldsymbol{u}\|_2^2} \right) \geq \frac{K^2 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^2}{2}.$$

Here the last inequality follows from our assumption that

$$\mu = \max_{1 \leq m \leq M} \frac{\sqrt{M} \|\boldsymbol{u}_m\|_2}{\|\boldsymbol{u}\|_2} \leq \frac{\sqrt{M}}{2}.$$

5.2. Proof of Lemma 3.5. We use the following notation in the proof. For $\boldsymbol{u} = [\boldsymbol{u}_1^\top, \ldots, \boldsymbol{u}_M^\top]^\top$, where $\boldsymbol{u}_m \in \mathbb{C}^D$ for $k = 1, \ldots, M$, we define a block (p, q) norm by

$$(39) \qquad \|\boldsymbol{u}\|_{p,q} = \begin{cases} \left(\displaystyle\sum_{m=1}^M \|\boldsymbol{u}_m\|_p^q \right)^{1/q}, & 1 \leq q < \infty, \\ \displaystyle\max_{1 \leq m \leq M} \|\boldsymbol{u}_m\|_p, & q = \infty. \end{cases}$$

By (36) and (37), $\boldsymbol{\Phi}^* \boldsymbol{Y}_\mathsf{s}^* \boldsymbol{Y}_\mathsf{s} \boldsymbol{\Phi} - \mathbb{E}[\boldsymbol{\Phi}^* \boldsymbol{Y}_\mathsf{s}^* \boldsymbol{Y}_\mathsf{s} \boldsymbol{\Phi}]$ is rewritten as the sum of its diagonal block portion (a) and the off-diagonal block portion (b), where

$$
\text{(a)} = \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \left(\sum_{\substack{m'=1 \\ m' \neq m}}^{M} \widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}} \widetilde{\boldsymbol{\Phi}}_m \right.
$$

$$
\tag{40} \left. - \mathbb{E}\left[\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}} \widetilde{\boldsymbol{\Phi}}_m \right] \right),
$$

$$
\text{(b)} = - \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes \left(\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_{m'} \right.
$$

$$
\left. - \mathbb{E}\left[\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_{m'} \right] \right).
$$

Therefore, by the triangle inequality, we have

$$
\| \boldsymbol{\Phi}^* \boldsymbol{Y}_\mathsf{s}^* \boldsymbol{Y}_\mathsf{s} \boldsymbol{\Phi} - \mathbb{E}[\boldsymbol{\Phi}^* \boldsymbol{Y}_\mathsf{s}^* \boldsymbol{Y}_\mathsf{s} \boldsymbol{\Phi}] \| \leq \|\text{(a)}\| + \|\text{(b)}\|.
$$

It remains to compute tail estimates for the spectral norms of (a) and (b).

Diagonal block portion. The (m,m)th block of (a) is written as $\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m$, where

$$
\boldsymbol{\Upsilon}_{\mathsf{s},m} = \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{S} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}} \boldsymbol{S}^*.
$$

Then due to the block diagonal structure, we have

$$
\|\text{(a)}\| = \max_{1 \leq m \leq M} \| \boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m - \mathbb{E}[\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m] \|.
$$

Since $\boldsymbol{\Phi}_m$ and $\boldsymbol{\Upsilon}_{\mathsf{s},m}$ are independent, $\mathbb{E}[\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m]$ is rewritten as

$$
\mathbb{E}[\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m] = \mathbb{E}_{\boldsymbol{\Phi}_m}[\boldsymbol{\Phi}_m^* \mathbb{E}_{\{\boldsymbol{\Phi}_{m'}\}_{m' \neq m}}[\boldsymbol{\Upsilon}_{\mathsf{s},m}] \boldsymbol{\Phi}_m] = \mathbb{E}_{\{\boldsymbol{\Phi}_{m'}\}_{m' \neq m}}[\operatorname{tr}(\boldsymbol{\Upsilon}_{\mathsf{s},m}) \mathbf{I}_D].
$$

Therefore, we have
$$
\tag{41}
\underbrace{\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m - \mathbb{E}[\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m]}_{} = \underbrace{\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathsf{s},m} \boldsymbol{\Phi}_m - \operatorname{tr}(\boldsymbol{\Upsilon}_{\mathsf{s},m}) \mathbf{I}_D}_{\text{(c)}} + \underbrace{(\operatorname{tr}(\boldsymbol{\Upsilon}_{\mathsf{s},m}) - \mathbb{E}[\operatorname{tr}(\boldsymbol{\Upsilon}_{\mathsf{s},m})]) \mathbf{I}_D}_{\text{(d)}}.
$$

We will compute tail estimates of the spectral norms of (c) and (d) separately and then combine them using the triangle inequality. First, we compute a tail bound of $\|\text{(c)}\|$ using the following lemmas, which are proved in Appendices D.1 and D.2.

Lemma 5.3. *Let* $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_D] \in \mathbb{C}^{K \times D}$ *satisfy that* $\operatorname{vec}(\boldsymbol{\Psi})$ *follows* $\mathcal{CN}(\boldsymbol{0}_{KD,1}, \mathbf{I}_{KD})$, *where* $\operatorname{vec}(\boldsymbol{\Psi}) = [\boldsymbol{\psi}_1^\top, \ldots, \boldsymbol{\psi}_D^\top]^\top$. *Then*

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

holds with probability $1 - \zeta$.

Lemma 5.4. *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*

$$\max_{1 \leq m \leq M} \|\boldsymbol{\Upsilon}_{\mathrm{s},m} - \mathbb{E}[\boldsymbol{\Upsilon}_{\mathrm{s},m}]\| \leq C(\beta)\rho_x K \|\boldsymbol{u}\|_{2,\infty} \|\boldsymbol{u}\|_2 \log^{\alpha}(MKL)$$

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

Lemma 5.3 is a direct consequence of the theory of suprema of second order chaos processes [25, 27]. By Lemma 5.3, conditioned on $\boldsymbol{\Upsilon}_{\mathrm{s},m}$,

$$\|(\mathrm{c})\| \leq C_1 \|\boldsymbol{\Upsilon}_{\mathrm{s},m}\| \sqrt{KD}(\log M + \beta \log K)$$

holds with probability $1 - M^{-1}K^{-\beta}$. Then by Lemmas 5.4 and B.1 with the triangle inequality, it follows that

$$\|(\mathrm{c})\| \leq C(\beta)(\rho_x + \|\boldsymbol{x}\|_2^2)K^{3/2}\sqrt{D}\|\boldsymbol{u}\|_2^2 \log^2(MK)$$

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

Next we consider (d). Note that

$$\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}}^*\boldsymbol{C}_{\boldsymbol{x}}^*\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}}\boldsymbol{S}^*\right) = K\|\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'}\|_2^2.$$

Therefore, the spectral norm of (d) is rewritten as

$$\|(\mathrm{d})\| = \left| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} K\|\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'}\|_2^2 - \mathbb{E}[K\|\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'}\|_2^2] \right|$$

$$= K \left| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \|(\boldsymbol{u}_{m'}^\top \otimes \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*)\boldsymbol{\phi}_{m'}\|_2^2 - \mathbb{E}[\|(\boldsymbol{u}_{m'}^\top \otimes \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*)\boldsymbol{\phi}_{m'}\|_2^2] \right|$$

$$= K \left| \underbrace{\left\| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} (\boldsymbol{e}_{m'}\boldsymbol{e}_{m'}^* \otimes \boldsymbol{u}_{m'}^\top \otimes \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*)\boldsymbol{\phi} \right\|_2^2}_{\boldsymbol{A}_m} - \mathbb{E}\left[\left\| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} (\boldsymbol{e}_{m'}\boldsymbol{e}_{m'}^* \otimes \boldsymbol{u}_{m'}^\top \otimes \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*)\boldsymbol{\phi} \right\|_2^2\right] \right|,$$

where $\boldsymbol{\phi}_m = \mathrm{vec}(\boldsymbol{\Phi}_m)$ for $m = 1, \ldots, M$ and $\boldsymbol{\phi} = [\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_M^\top]^\top$. In fact, we are computing a tail bound of the Gaussian quadratic form $\|\boldsymbol{A}_m\boldsymbol{\phi}\|_2^2$, which can be done by the Hanson–Wright inequality, Lemma A.2. Observing that the block diagonal matrix $\boldsymbol{U}$ with blocks $\boldsymbol{u}_m$ satisfies $\|\boldsymbol{U}^*\boldsymbol{U}\|_F \leq \|\boldsymbol{u}\|_{2,4}^2$, we obtain

$$\|(\mathrm{d})\| = K|\|\boldsymbol{A}_m\boldsymbol{\phi}\|_2^2 - \mathbb{E}[\|\boldsymbol{A}_m\boldsymbol{\phi}\|_2^2]|$$

$$\leq C_3 K(\|\boldsymbol{A}_m^*\boldsymbol{A}_m\|_{\mathrm{F}} \vee \|\boldsymbol{A}_m\|^2)(2\log M + 2\beta \log K)$$

$$\leq C_3 K(\sqrt{K}\|\boldsymbol{u}\|_{2,4}^2 \vee \|\boldsymbol{u}\|_{2,\infty}^2)\|\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{x}}^*\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{S}^*\|(2\log M + 2\beta \log K)$$

$$\leq C_4 K^{3/2}\|\boldsymbol{u}\|_{2,4}^2 \rho_x(2\log M + 2\beta \log K)$$

holds with probability $1 - M^{-1}K^{-\beta}$. Note that the tail bound of (c) dominates that of (d).

Collecting the above estimates, it follows that

$$\|(a)\| \leq C(\beta)\rho_x K^{3/2}\sqrt{D}\|\boldsymbol{u}\|_2^2 \log^\alpha(MKL)$$

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

If we normalize with the spectral gap given in (38), then the relative perturbation due to (a) is upper bounded by

$$(42) \qquad\qquad \frac{\|(a)\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq C(\beta) \log^\alpha(MKL)\sqrt{\frac{D}{K}}$$

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

Off-diagonal block portion. Unlike the diagonal block portion (a), the off-diagonal block portion (b) does not have a block diagonal structure and computing its tail bound is more involved.

To restrict the convolution of two short vectors of length K to its support, we introduce $\breve{\boldsymbol{S}} \in \mathbb{R}^{(2K-1)\times L}$ defined by

$$\breve{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{0}_{K-1,L-K+1} & \mathbf{I}_{K-1} \\ \mathbf{I}_K & \boldsymbol{0}_{K,L-K} \end{bmatrix}.$$

Then we have

$$(43) \qquad\qquad\qquad \boldsymbol{S}\breve{\boldsymbol{S}}^*\breve{\boldsymbol{S}} = \boldsymbol{S}$$

and

$$(44) \qquad\qquad \boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m}\widetilde{\boldsymbol{\Phi}}_m = \breve{\boldsymbol{S}}^*\breve{\boldsymbol{S}}\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m}\widetilde{\boldsymbol{\Phi}}_m \quad \forall m = 1,\dots,M.$$

Due to the commutativity of product of two circulant matrices, (43), and (44), we can rewrite (b) as

$$(b) = -\sum_{m=1}^M \sum_{\substack{m'=1 \\ m'\neq m}}^M \boldsymbol{e}_m\boldsymbol{e}_{m'}^* \otimes \Big((\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m)^*\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'} - \mathbb{E}[(\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m)^*\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}]\Big),$$

where

$$\boldsymbol{Z}_m = \breve{\boldsymbol{S}}\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m}\widetilde{\boldsymbol{\Phi}}_m, \quad m = 1,\dots,M.$$

Note that the summation in (b) runs over all distinct pairs (m, m') with $m \neq m'$. Our main trick here is to add and subtract the terms corresponding to pairs (m, m) for $m = 1,\dots,M$. This ends up with a diagonal sum and a full summation over all pairs (m, m'). The resulting full summation term now provides a nice factorization, which leads to an analysis using the techniques for the second order chaos processes.

Indeed, since the $\boldsymbol{\Phi}_m$'s are independent, we have

$$\mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]^*\mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}] - \mathbb{E}[(\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m)^*\boldsymbol{C}_{\boldsymbol{x}}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}] = \boldsymbol{0}_{D,D} \quad \forall m \neq m'.$$

Therefore, (b) is decomposed as (b) = (e) − (f), where

$$(e) = \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes ((\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m)^* \boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]^* \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]),$$

$$(f) = \sum_{m,m'=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes ((\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m)^* \boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]^* \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}]).$$

Note that $(\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m)^* \boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]^* \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}]$ is decomposed as

$$
\begin{aligned}
(45) \qquad & (\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m])^* (\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}]) \\
& + \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m^*](\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}]) \\
& + (\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m])^* \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}].
\end{aligned}
$$

Therefore, the spectral norm of (e), which corresponds to the extra diagonal term, is upper bounded by

$$
\begin{aligned}
\|(e)\| \leq & \left(\max_{1 \leq m \leq M} \|\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]\| \right)^2 \\
& + 2 \left(\max_{1 \leq m \leq M} \|\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]\| \right) \left(\max_{1 \leq m \leq M} \|\mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]\| \right).
\end{aligned}
$$

By Lemma B.2, we have

$$\|\mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m]\| = K\|\boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{e}_1 \boldsymbol{u}_m^*\| \leq K\|\boldsymbol{x}\|_2 \|\boldsymbol{u}_m\|_2 \leq K\|\boldsymbol{x}\|_2 \|\boldsymbol{u}\|_{2,\infty}.$$

We again use bounds for suprema of second order chaos processes [25, 27] to get a tail bound for $\|\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]\|$, as given in the following lemma, which is proved in Appendix D.3.

Lemma 5.5. *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*

$$\sup_{1 \leq m \leq M} \|\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]\| \leq C(\beta)\|\boldsymbol{u}\|_{2,\infty} K \log^{\alpha}(MKL)$$

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

By Lemma 5.5, it follows that the relative perturbation due to (e) is upper bounded by

$$(46) \qquad \frac{\|(e)\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq \frac{C(\beta) \log^{\alpha}(MKL)\mu^2}{M}$$

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

Similarly, (f), which corresponds to the full two-dimensional summation, is rewritten as

$$
\begin{aligned}
(f) = & \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m \right)^* \left(\sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes \boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'} \right) \\
& - \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m] \right)^* \left(\sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes \mathbb{E}[\boldsymbol{C}_{\boldsymbol{x}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}] \right)
\end{aligned}
$$

$$= \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]) \right)^* \left(\sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}]) \right)$$

$$+ \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m] \right)^* \left(\sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}]) \right)$$

$$+ \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]) \right)^* \left(\sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_{m'}] \right).$$

Therefore, by the triangle inequality, we have

$$\|(\mathrm{f})\| \le \left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]) \right\|^2$$

$$+ 2\left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]) \right\| \left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m] \right\|.$$

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M \in \mathbb{C}^D$ and $\boldsymbol{v} = [\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_M^\top]^\top$. Then, by Lemma B.2, we have

$$\left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m] \right\| = \sup_{\boldsymbol{v} \in B_2^{MD}} \left\| \sum_{m=1}^{M} \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m \boldsymbol{v}_m] \right\|_2$$

$$= \sup_{\boldsymbol{v} \in B_2^{MD}} K \left\| \sum_{m=1}^{M} \boldsymbol{C_x}\boldsymbol{e}_1 \boldsymbol{u}_m^* \boldsymbol{v}_m \right\|_2 = \sup_{\boldsymbol{v} \in B_2^{MD}} K\|\boldsymbol{x}\|_2 \left| \sum_{m=1}^{M} \boldsymbol{u}_m^* \boldsymbol{v}_m \right|$$

$$\le \sup_{\boldsymbol{v} \in B_2^{MD}} K\|\boldsymbol{x}\|_2 \sum_{m=1}^{M} \|\boldsymbol{u}_m\|_2 \|\boldsymbol{v}_m\|_2 \le K\|\boldsymbol{x}\|_2 \|\boldsymbol{u}\|_2.$$

On the other hand,

$$\left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\boldsymbol{Z}_m]) \right\| = \left\| \boldsymbol{C_x}\breve{\boldsymbol{S}}^* \left(\sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]) \right) \right\|$$

$$\le \|\breve{\boldsymbol{S}}\boldsymbol{C_x}^*\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\|^{1/2} \underbrace{\left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]) \right\|}_{(\ddagger)}.$$

As implied by (19), the first factor is bounded by $\sqrt{C_3}\|\boldsymbol{x}\|_2$. Hence it remains to show an upper bound on the last term ($\ddagger$). This is established in Lemma 5.6 using the results on suprema of second order chaos processes [25, 27] together with an entropy bound by polytope approximation of a unit ball [20], polar duality, and entropy duality [2]; see Appendix D.4 for the proof.

Lemma 5.6. *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*

$$\left\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]) \right\| \le C(\beta)\|\boldsymbol{u}\|_{2,\infty}(K + \sqrt{MKD}) \log^{\alpha}(MKL)$$

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

Collecting the above estimates and noting that, up to log factors, both factors are bounded by $\mu K \|\boldsymbol{x}\|_2 \|\boldsymbol{u}\|_2$, we obtain that the relative perturbation due to (f) is upper bounded by

$$(47) \qquad \frac{\|(\mathrm{f})\|}{K^2 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{u}\|_2^2} \leq C(\beta) \log^\alpha(MKL) \Big(\sqrt{\frac{1}{M}} + \sqrt{\frac{D}{K}} \Big) \mu^2,$$

potentially for an increased value of α, with probability $1 - CK^{-\beta}$.

Finally, (28) follows by combining (42), (46), and (47).

5.3. Proof of Lemma 3.6. The proof of Lemma 3.6 is similar to (and easier than) that of Lemma 3.5. We will reuse some of tail estimates obtained in section 5.2. On the other hand, we also need to derive tail estimates of the suprema of certain Gaussian processes, which did not arise in section 5.2. We use a moment-version of Dudley's inequality [12] to compute these tail estimates.

Similarly to the previous section, we use the following decomposition into the diagonal block portion and the off-diagonal block portion:

$$\boldsymbol{\Phi}^* \boldsymbol{Y}_\mathrm{s}^* \boldsymbol{Y}_\mathrm{n} \boldsymbol{\Phi} = (\mathrm{g}) + (\mathrm{h}),$$

where

$$(48) \qquad \begin{aligned} (\mathrm{g}) &= \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \left(\sum_{\substack{m'=1 \\ m' \neq m}}^{M} \widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \widetilde{\boldsymbol{\Phi}}_m \right), \\ (\mathrm{h}) &= - \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes \widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \widetilde{\boldsymbol{\Phi}}_{m'}. \end{aligned}$$

We derive upper bounds on the spectral norms of (g) and (h) respectively in the following.

Diagonal block portion. Note that (g) is a block diagonal matrix and its expectation is $\boldsymbol{0}$. Define

$$(49) \qquad \boldsymbol{\Upsilon}_{\mathrm{c},m} = \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{S} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{S}^*.$$

Then it follows from the block diagonal structure that

$$\|(\mathrm{g})\| \leq \max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{\Phi}_m\|.$$

Since $\boldsymbol{\Phi}_m$ and $\boldsymbol{\Upsilon}_{\mathrm{c},m}$ are independent, $\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{\Phi}_m$ is rewritten as

$$(50) \qquad \boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{\Phi}_m = \underbrace{\boldsymbol{\Phi}_m^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{\Phi}_m - \mathrm{tr}(\boldsymbol{\Upsilon}_{\mathrm{c},m}) \boldsymbol{I}_D}_{(\mathrm{i})} + \underbrace{\mathrm{tr}(\boldsymbol{\Upsilon}_{\mathrm{c},m}) \boldsymbol{I}_D}_{(\mathrm{j})}.$$

First, we compute a tail estimate of $\|(\mathrm{i})\|$. Similarly to the previous section, we use Lemma 5.3 conditioned on $\boldsymbol{\Upsilon}_{\mathrm{c},m}$. Then we apply the tail estimate of $\|\boldsymbol{\Upsilon}_{\mathrm{c},m}\|$ given in the following lemma, whose proof is in Appendix D.5.

Lemma 5.7. *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, conditional on the noise vector* $\boldsymbol{w}$,

$$\max_{1 \leq m \leq M} \|\boldsymbol{\Upsilon}_{c,m}\| \leq C(\beta)\rho_{x,w}\sqrt{K}\|\boldsymbol{u}\|_2 \log^\alpha(MKL)$$

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

By Lemmas 5.3 and 5.7, we obtain that

$$\|(i)\| \leq C(\beta)\rho_{x,w}K\sqrt{D}\|\boldsymbol{u}\|_2 \log^\alpha(MKL)$$

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

On the other hand, a direct calculation shows that (j) is expressed as

$$\overline{\mathrm{tr}(\boldsymbol{\Upsilon}_{c,m})} = \sum_{\substack{m'=1 \\ m'\neq m}}^{M} (\boldsymbol{u}_{m'}^\top \otimes K\boldsymbol{w}_{m'}^* \boldsymbol{C_x S^*})\mathrm{vec}(\boldsymbol{\Phi}_{m'}),$$

where $\overline{\mathrm{tr}(\boldsymbol{\Upsilon}_{c,m})}$ denotes the complex conjugate of $\mathrm{tr}(\boldsymbol{\Upsilon}_{c,m})$. The above expression implies that $\overline{\mathrm{tr}(\boldsymbol{\Upsilon}_{c,m})}$ is a linear function of the independent Gaussian matrix entries and hence is a zero-mean Gaussian random variable. The variance of $\overline{\mathrm{tr}(\boldsymbol{\Upsilon}_{c,m})}$ is given by

$$\sum_{\substack{m'=1 \\ m'\neq m}}^{M} K^2\|\boldsymbol{u}_{m'}\|_2^2\|\boldsymbol{w}_{m'}^* \boldsymbol{C_x S^*}\|_2^2 \leq K^2\|\boldsymbol{u}\|_2^2\rho_{x,w}^2.$$

Therefore, by a tail estimate of a Gaussian variable,

$$\|(j)\| \leq C\rho_{x,w}K\|\boldsymbol{u}\|_2\sqrt{1 + \log M + \beta \log K}$$

with probability $1 - M^{-1}K^{-\beta}$. Hence $\|(i)\|$ dominates $\|(j)\|$.

By collecting the above estimates, we obtain that the relative perturbation due to (g) is upper bounded by

$$(51) \qquad \frac{\|(g)\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} \leq \frac{C(\beta)\rho_{x,w}\log^\alpha(MKL)}{\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2}\frac{\sqrt{D}}{K} = \frac{C(\beta)\log^\alpha(MKL)}{\sqrt{\eta L}}\frac{\rho_{x,w}}{\sqrt{K}\sigma_w\|\boldsymbol{x}\|_2}\sqrt{\frac{D}{M}}$$

with probability $1 - CK^{-\beta}$, where in the last step we used (12).

Off-diagonal portion. Similarly to the analogous part of the proof of Lemma 3.5, we add and subtract the diagonal sum and obtain

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

where

$$(52) \qquad \begin{aligned} (k) &= \sum_{m=1}^{M} \boldsymbol{e}_m\boldsymbol{e}_m^* \otimes \boldsymbol{\Phi}_m^* \boldsymbol{S C_x^* C_{w_m}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_m, \\ (l) &= -\sum_{m,m'=1}^{M} \boldsymbol{e}_m\boldsymbol{e}_{m'}^* \otimes \boldsymbol{\Phi}_m^* \boldsymbol{S C_x^* C_{w_m}} \breve{\boldsymbol{S}}^* \boldsymbol{Z}_{m'}. \end{aligned}$$

By the triangle inequality,

$$\|(\mathrm{k})\| \leq \Big\| \underbrace{\sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \breve{\boldsymbol{S}}^* (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m])}_{(\mathrm{m})} \Big\|$$

$$+ \Big\| \underbrace{\sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \breve{\boldsymbol{S}}^* \mathbb{E}[\boldsymbol{Z}_m]}_{(\mathrm{n})} \Big\|.$$

We use the result by Davidson and Szarek [9, Theorem II.13] to get a tail estimate of $\|\boldsymbol{\Phi}_m\|$. Specifically, it follows from (A1) that

$$(53) \qquad \max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m\| \leq \|[\sqrt{2}\mathrm{Re}(\boldsymbol{\Phi}_m), \sqrt{2}\mathrm{Im}(\boldsymbol{\Phi}_m)]\| \leq \sqrt{K} + \sqrt{2D} + \sqrt{2\log M + 2\beta \log K}$$

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

By Lemma 5.5 and (53),

$$\|(\mathrm{m})\| = \max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \breve{\boldsymbol{S}}^* (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m])\|$$

$$\leq \Big(\max_{1 \leq m \leq M} \|\boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \breve{\boldsymbol{S}}\| \Big) \Big(\max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m\| \Big) \Big(\max_{1 \leq m \leq M} \|\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]\| \Big)$$

$$\leq \rho_{x,w} C(\beta) K^{3/2} \|\boldsymbol{u}\|_{2,\infty} \log^6(MK)$$

holds with probability $1 - K^{-\beta}$, where $\| \cdot \|_{2,\infty}$ is defined in (39). On the other hand, by Lemmas B.2 and A.3,

$$\|(\mathrm{n})\| = K \max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{w}_m \boldsymbol{u}_m^*\|$$

$$\leq K \|\boldsymbol{u}\|_{2,\infty} \max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{w}_m\|_2$$

$$= K \|\boldsymbol{u}\|_{2,\infty} \max_{1 \leq m \leq M} \|(\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \otimes \mathbf{I}_D) \mathrm{vec}(\boldsymbol{\Phi}_m^\top)\|_2$$

$$\leq K \|\boldsymbol{u}\|_{2,\infty} \max_{1 \leq m \leq M} \|\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \otimes \mathbf{I}_D\|_{\mathrm{F}} \sqrt{1 + \log M + \beta \log K}$$

$$= K \|\boldsymbol{u}\|_{2,\infty} \max_{1 \leq m \leq M} \sqrt{D} \|\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*\|_2 \sqrt{1 + \log M + \beta \log K}$$

$$\leq 2K \|\boldsymbol{u}\|_{2,\infty} \rho_{x,w} \sqrt{D} \sqrt{1 + \log M + \beta \log K}$$

holds with probability $1 - K^{-\beta}$, where the last inequality follows from the fact that $\|\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*\|_2 \leq \|\boldsymbol{S} \boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*\|$. Note that $\|(\mathrm{m})\|$ dominates $\|(\mathrm{n})\|$.

The spectral norm of (l) is upper bounded through a factorization by

$$\|(l)\| \le \underbrace{\Big\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \check{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m \Big\|}_{(o)} \underbrace{\Big\| \sum_{m'=1}^{M} \boldsymbol{e}_{m'}^* \otimes (\boldsymbol{Z}_{m'} - \mathbb{E}[\boldsymbol{Z}_{m'}]) \Big\|}_{(p)}$$

$$+ \underbrace{\Big\| \sum_{m,m'=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes \boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \check{\boldsymbol{S}}^* \mathbb{E}[\boldsymbol{Z}_{m'}] \Big\|}_{(q)}.$$

The spectral norm of (o) is written as the supremum of a Gaussian process and is bounded by the following lemma, which is proved in Appendix D.6.

Lemma 5.8. *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* $\boldsymbol{w}$,

$$\Big\| \sum_{m=1}^{M} \boldsymbol{e}_m^* \otimes \check{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m \Big\| \le C \sqrt{1+\beta} \rho_{x,w} (\sqrt{MD} + \sqrt{K}) \log K$$

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

By Lemma 5.6,

$$\|(p)\| \le C(\beta) \|\boldsymbol{u}\|_{2,\infty} (K + \sqrt{MKD}) \log^{\alpha}(MKL)$$

with probability $1 - K^{-\beta}$, where $\|\cdot\|_{2,\infty}$ is defined in (39). Note that $\|(o)\| \|(p)\|$ dominates $\|(m)\|$. Therefore, we may ignore $\|(m)\|$.

By Lemma B.2, the spectral norm of (q) is upper bounded by

$$\|(q)\|^2 = \Big\| \sum_{m,m'=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes K \boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{w}_m \boldsymbol{u}_{m'}^* \Big\|^2$$

$$= \Big\| \sum_{m=1}^{M} \boldsymbol{e}_m \otimes K \boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{w}_m \boldsymbol{u}^* \Big\|^2$$

$$= K^2 \|\boldsymbol{u}\|_2^2 \sum_{m=1}^{M} \|\boldsymbol{\Phi}_m^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{w}_m\|_2^2$$

$$= K^2 \|\boldsymbol{u}\|_2^2 \sum_{m=1}^{M} \|(\mathbf{I}_D \otimes \boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*) \mathrm{vec}(\boldsymbol{\Phi}_m)\|^2$$

$$= K^2 \|\boldsymbol{u}\|_2^2 \Big\| \Big(\sum_{m=1}^{M} \boldsymbol{e}_m^\top \otimes \mathbf{I}_D \otimes \boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \Big) [\mathrm{vec}(\boldsymbol{\Phi}_m)^\top, \dots, \mathrm{vec}(\boldsymbol{\Phi}_m)^\top]^\top \Big\|^2.$$

Therefore, by Lemma A.3,

$$\|(q)\| \le C(\beta) K \|\boldsymbol{u}\|_2 \sqrt{MD} \rho_{x,w} \sqrt{\log K}$$

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

By collecting the estimates with the fact that $\mu \le \sqrt{M}$, we obtain that with probability $1 - K^{-\beta}$, the relative perturbation due to (h) is upper bounded by

$$
\begin{aligned}
(54) \quad \frac{\|(\mathrm{h})\|}{K^2\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2^2} &\le \frac{C(\beta)\rho_{x,w}\log^\alpha(MKL)}{\|\boldsymbol{x}\|_2^2\|\boldsymbol{u}\|_2}\left(\mu\sqrt{MD+K}\left(\frac{1}{\sqrt{M}K} + \frac{\sqrt{D}}{K^{3/2}}\right) + \frac{\sqrt{MD}}{K}\right)\\
&\le \frac{C'(\beta)\log^\alpha(MKL)}{\sqrt{\eta L}}\frac{\rho_{x,w}}{\sqrt{K}\sigma_w\|\boldsymbol{x}\|_2}\left(\mu\left(\frac{\sqrt{K}}{M} + \frac{D}{\sqrt{K}}\right) + \sqrt{D}\right).
\end{aligned}
$$

Note that the tail estimate in (54) dominates that in (51). It therefore follows that (29) holds with probability $1 - CK^{-\beta}$. This completes the proof.

5.4. Proof of Lemma 3.7. The analysis of the noise term $\boldsymbol{E}_{\mathrm{n}}$ only involves second order chaos processes and can be reduced to bounds for suprema of such processes as they are established in [25, 27].

The first lemma used in the proof bounds the maximum cross-correlation deviation of the noise terms, which is given by

$$
\rho_w := \max_{1 \le m,m' \le M}\|\boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^*\boldsymbol{C}_{\boldsymbol{w}_{m'}} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^*\boldsymbol{C}_{\boldsymbol{w}_{m'}}])\boldsymbol{S}^*\|.
$$

See Appendix D.7 for the proof of the lemma.

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*

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

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

The second proof ingredient is a bound for the average autocorrelation deviation of the noise terms, which is given by

$$
\bar{\rho}_w := \left\|\frac{1}{M}\sum_{m=1}^M \boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^*\boldsymbol{C}_{\boldsymbol{w}_m} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^*\boldsymbol{C}_{\boldsymbol{w}_m}])\boldsymbol{S}^*\right\|.
$$

The bound is provided by the following lemma; see Appendix D.8 for its proof.

Lemma 5.10. Suppose that (A2) *holds. For any $\beta \in \mathbb{N}$, there is a constant $C(\beta)$ that depends only on β such that*

$$
(56) \quad \bar{\rho}_w \le C(\beta)\sigma_w^2 M^{-1/2}\sqrt{KL}\log^\alpha(MKL)
$$

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

For the remainder of the proof of Lemma 3.7, we condition on the event that (55) and (56) hold.

Define $\boldsymbol{\Lambda} \in \mathbb{C}^{MK \times MK}$ by

$$
\boldsymbol{\Lambda} = \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \ne m}}^M \|\boldsymbol{w}_{m'}\|_2^2 \boldsymbol{e}_m\boldsymbol{e}_m^* \otimes \mathbf{I}_K.
$$

Under the assumption in (A2), we have

$$\mathbb{E}[\|\boldsymbol{w}_{m'}\|_2^2] = \sigma_w^2 L \quad \forall m' = 1, \ldots, M.$$

Then it follows that $\mathbb{E}[\boldsymbol{\Lambda}]$ is a scalar multiple of the identity given by

$$\mathbb{E}[\boldsymbol{\Lambda}] = \sigma_w^2 (M-1) L \mathbf{I}_{MK}.$$

Here we assume that σ_w is known a priori or can be estimated from the data.

We decompose $\boldsymbol{\Phi}^*(\boldsymbol{Y}_n^* \boldsymbol{Y}_n - \sigma_w^2 (M-1) L \mathbf{I}_{MK}) \boldsymbol{\Phi}$ into two parts as follows:

$$(57) \qquad \boldsymbol{\Phi}^*(\boldsymbol{Y}_n^* \boldsymbol{Y}_n - \sigma_w^2 (M-1) L \mathbf{I}_{MK}) \boldsymbol{\Phi} = \underbrace{\boldsymbol{\Phi}^*(\boldsymbol{Y}_n^* \boldsymbol{Y}_n - \boldsymbol{\Lambda}) \boldsymbol{\Phi}}_{(r)} + \underbrace{\boldsymbol{\Phi}^*(\boldsymbol{\Lambda} - \mathbb{E}[\boldsymbol{\Lambda}]) \boldsymbol{\Phi}}_{(s)}.$$

Then we estimate the summands in the right-hand side of (57). The following lemma, which is proved in Appendix D.9, provides a tail estimate of $\|(r)\|$.

Lemma 5.11. *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, conditional on the noise vector* $\boldsymbol{w}$,

$$\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_n^* \boldsymbol{Y}_n - \boldsymbol{\Lambda}) \boldsymbol{\Phi}\| \leq C(\beta) M \sqrt{KD}(\bar{\rho}_w + 2\rho_w) \log^{\alpha}(MD)$$

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

Next, due to the block diagonal structure, the spectral norm of the second term (s) is upper bounded by

$$\max_{1 \leq m \leq M} \Big\| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} (\|\boldsymbol{w}_{m'}\|_2^2 - \mathbb{E}[\|\boldsymbol{w}_{m'}\|_2^2]) \boldsymbol{\Phi}_m^* \boldsymbol{\Phi}_m \Big\|$$

$$\leq \left(\max_{1 \leq m \leq M} \Big| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \|\boldsymbol{w}_{m'}\|_2^2 - \mathbb{E}[\|\boldsymbol{w}_{m'}\|_2^2] \Big| \right) \left(\max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{\Phi}_m\| \right).$$

The first factor divided by σ_w^2 is a χ^2 random variable with LM degrees of freedom. By Lemma A.1,

$$\max_{1 \leq m \leq M} \Big| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \|\boldsymbol{w}_{m'}\|_2^2 - \mathbb{E}[\|\boldsymbol{w}_{m'}\|_2^2] \Big| \leq C_1(1+\beta) \sigma_w^2 \sqrt{LM} \log(MK)$$

holds with probability $1 - K^{-\beta}$. On the other hand, by (53),

$$\max_{1 \leq m \leq M} \|\boldsymbol{\Phi}_m^* \boldsymbol{\Phi}_m\| \leq C_2(1+\beta) K \log(MK)$$

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

$$\|(s)\| \leq C(\beta) \sigma_w^2 \sqrt{M} K \sqrt{L} \log^2(MK)$$

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

Finally, (30) follows by collecting the above estimates. This completes the proof.

5.5. Proof of Lemma 3.8. Similarly to the proof of Lemma 3.7, through the triangle inequality, the left-hand side of (35) is upper bounded by

$$
(58) \qquad \frac{\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda})\boldsymbol{\Phi}\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2} + \frac{\|\boldsymbol{\Phi}^*(\boldsymbol{\Lambda} - \mathbb{E}[\boldsymbol{\Lambda}])\boldsymbol{\Phi}\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2}.
$$

For the first term in the right-hand side of (58), we modify Lemma 5.11 as follows; see Appendix D.10 for the proof.

Lemma 5.12. *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, conditional on the noise vector* $\boldsymbol{w}$,

$$
\frac{\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda})\boldsymbol{\Phi}\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2} \leq C(\beta)\sqrt{MKD}(\sqrt{M}\bar{\rho}_w + \rho_w)\log^\alpha(MD)
$$

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

Combining Lemmas 5.9, 5.10, and 5.12 implies that the first term in the right-hand side of (58) is smaller than $\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda})\boldsymbol{\Phi}\|$ by a factor of $\sqrt{M}$.

For the second term in the right-hand side of (58), we use the fact that it is no larger than $\|\boldsymbol{\Phi}^*(\boldsymbol{\Lambda} - \mathbb{E}[\boldsymbol{\Lambda}])\boldsymbol{\Phi}\|$. Then we may use the tail estimate derived in the proof of Lemma 3.7.

By collecting the estimates, we obtain that (35) holds with probability $1 - CK^{-\beta}$. This completes the proof.

6. Conclusion. We studied a passive imaging problem with multiple channels, which is formulated as multichannel blind deconvolution with noise-like source and time-limited impulse responses. Additionally, motivated by several real-world applications, we impose that the FIR coefficients of impulse responses belong to corresponding low-dimensional subspaces. For such a scenario, we proposed a spectral method called subspace-constrained cross-convolution (SCCC) that modifies and improves upon a classical method developed in the 1990s by overcoming the noise sensitivity. SCCC provides stable estimates of the impulse responses from finitely many samples and its performance is backed by theoretical error bounds under generic subspace models. In this scenario, SCCC also empirically outperforms competing approaches. The fundamental estimates in the analysis of this paper extend to the sparsity or low-rank cases with minor changes. Corresponding recovery results on these extended models will be left for follow-up work.

Appendix A. Toolbox.

A.1. Concentration of χ^2 random variables.

Lemma A.1 (complexification of [26, Lemma 1]). *Let* $g_1, \dots, g_n$ *be independent copies of a standard complex Gaussian variable. Let* $a_1, \dots, a_n$ *be nonnegative and* $\boldsymbol{a} = [a_1, \dots, a_n]^\top$. *Let* $Z = \sum_{k=1}^n a_k(|g_k|^2 - 1)$. *Then, for any* $t > 0$,

$$
\begin{aligned}
\mathbb{P}(Z \geq \sqrt{2}\|\boldsymbol{a}\|_2\sqrt{t} + \|\boldsymbol{a}\|_\infty t) &\leq \exp(-t), \\
\mathbb{P}(Z \leq -\sqrt{2}\|\boldsymbol{a}\|_2\sqrt{t}) &\leq \exp(-t).
\end{aligned}
$$

A.2. Hanson–Wright inequality.

Lemma A.2 (complexification of [40, Theorem 1.1]). *Let $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. Let $\boldsymbol{g} \in \mathbb{C}^n$ be a standard complex Gaussian vector. For any $0 < \zeta < 1$, there exists an absolute constant C such that*

$$\big| \|\boldsymbol{A}\boldsymbol{g}\|_2^2 - \mathbb{E}[\|\boldsymbol{A}\boldsymbol{g}\|_2^2] \big| \leq C(\|\boldsymbol{A}^*\boldsymbol{A}\|_{\mathrm{F}} \sqrt{\log(2\zeta^{-1})} \vee \|\boldsymbol{A}\|^2 \log(2\zeta^{-1}))$$

holds with probability $1 - \zeta$.

Lemma A.3 (complexification of [40, Theorem 2.1]). *Let $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. Let $\boldsymbol{g} \in \mathbb{C}^n$ be a standard complex Gaussian vector. For any $0 < \zeta < 1$, there exists an absolute constant C such that*

$$\big| \|\boldsymbol{A}\boldsymbol{g}\|_2 - \|\boldsymbol{A}\|_{\mathrm{F}} \big| \leq C\|\boldsymbol{A}\| \sqrt{\log(2\zeta^{-1})}$$

holds with probability $1 - \zeta$.

A.3. Complexification of Maurey's lemma.

The following lemma is a direct consequence of Maurey's empirical method [5]. Define a block norm on $\mathbb{R}^{md}$ by

$$\|[\boldsymbol{q}_1^\top, \dots, \boldsymbol{q}_m^\top]^\top\|_{\ell_\infty^m(\ell_2^d)} := \max_{1 \leq k \leq m} \|\boldsymbol{q}_k\|_2.$$

Let $\ell_\infty^m(\ell_2^d(\mathbb{R}))$ denote the corresponding Banach space. Similarly, $\ell_\infty^m(\ell_2^d(\mathbb{C}))$ is defined over the complex scalar field.

Lemma A.4 (Maurey's empirical method [20, Lemma 3.1]). *Let $k, m, n \in \mathbb{N}$ and $T : \ell_1^k(\mathbb{R}) \to \ell_\infty^m(\ell_2^d(\mathbb{R}))$ be a linear operator. Then*

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

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm.

Lemma A.4 extends to the complex field case, which is shown in the following corollary.

Corollary A.5. *Let $k, m, n \in \mathbb{N}$ and $T : \ell_1^k(\mathbb{C}) \to \ell_\infty^m(\ell_2^d(\mathbb{C}))$ be a linear operator. Then*

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

Proof of Corollary A.5. Let $\iota : \mathbb{C} \to \mathbb{R}^2$ be a natural map defined by

$$\iota(x) = [\mathrm{Re}(x), \mathrm{Im}(x)]^\top \quad \forall x \in \mathbb{C}.$$

By a slight abuse of notation, we apply ι elementwise to $\mathbb{C}^k$, i.e.,

$$\iota([x_1, \dots, x_k]^\top) = [\mathrm{Re}(x_1), \mathrm{Im}(x_1), \dots, \mathrm{Re}(x_k), \mathrm{Im}(x_k)]^\top.$$

Then we have

$$\iota B_{\ell_1^k(\mathbb{C})} \subset \sqrt{2} B_{\ell_1^{2k}(\mathbb{R})}$$

and

$$\iota B_{\ell_\infty^m(\ell_2^d(\mathbb{C}))} = B_{\ell_\infty^m(\ell_2^{2d}(\mathbb{R}))}.$$

Moreover, ι is bijective. Therefore,

$$\int_0^\infty \sqrt{\log N(T(B_{\ell_1^k(\mathbb{C})}), \|\cdot\|_{\ell_\infty^m(\ell_2^d(\mathbb{C}))}, t)} dt \leq \int_0^\infty \sqrt{\log N(T(\sqrt{2}B_{\ell_1^{2k}(\mathbb{R})}), \|\cdot\|_{\ell_\infty^m(\ell_2^{2d}(\mathbb{R}))}, t)} dt.$$

Then the assertion follows from Lemma A.4 with a change of variable in the integral. ∎

A.4. Suprema of Gaussian processes. We use the following lemma that provides tail estimates of suprema of first order chaos processes.

Lemma A.6. *Let $\boldsymbol{\xi} \in \mathbb{C}^n$ be a standard Gaussian vector with $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^*] = \mathbf{I}_n$, $\Delta \subset \mathbb{C}^n$, and $0 < \zeta < e^{-1/2}$. There is an absolute constant C such that*

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

holds with probability $1 - \zeta$.

Proof. Lemma A.6 is a direct consequence of the moment-version of Dudley's inequality [12, p. 263] and a version of Markov's inequality [12, Proposition 7.11]. ∎

Appendix B. Expectations.

Lemma B.1. *Under the assumption in* (A1),

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

Proof of Lemma B.1.

$$\mathbb{E}[\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m}] = \|\boldsymbol{u}_m\|_2^2\mathbb{E}[\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{g}}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}}],$$

where $\boldsymbol{g} \in \mathbb{C}^K$ is a standard complex Gaussian vector. Let g_k denote the kth entry of $\boldsymbol{g}$. Since $\mathbb{E}[g_k\overline{g_l}] = 0$ for all $k \neq l$, we have

$$\mathbb{E}[\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{g}}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}}] = \sum_{k=1}^K \boldsymbol{C}^*_{\boldsymbol{e}_k}\boldsymbol{C}_{\boldsymbol{e}_k} = K\mathbf{I}_L.$$

This completes the proof. ∎

Lemma B.2. *Under the assumption in* (A1),

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

Proof of Lemma B.2. By (A1), it follows that $\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m$ and $\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{q}$ are independent for any $\boldsymbol{q} \in \mathbb{C}^K$ such that $\boldsymbol{q}^*\boldsymbol{u}_m = 0$. Therefore,

$$\mathbb{E}[\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m] = \mathbb{E}\left[\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m \frac{\boldsymbol{u}_m\boldsymbol{u}_m^*}{\|\boldsymbol{u}_m\|_2^2}\right] = \frac{\mathbb{E}[\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m]}{\|\boldsymbol{u}_m\|_2^2}\boldsymbol{u}_m^*$$

$$= \frac{\mathbb{E}[\boldsymbol{C}^*_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m\boldsymbol{u}_m}\boldsymbol{e}_1]}{\|\boldsymbol{u}_m\|_2^2}\boldsymbol{u}_m^* = K\boldsymbol{e}_1\boldsymbol{u}_m^*,$$

where the last step follows from Lemma B.1. ∎

Lemma B.3. *Under the assumption in* (A1),

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

Proof of Lemma B.3. Suppose that $m \neq m'$. Then, by the independence of $\boldsymbol{\Phi}_m$ and $\boldsymbol{\Phi}_{m'}$, it follows that

$$\begin{aligned}
\mathbb{E}[\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}} \widetilde{\boldsymbol{\Phi}}_m] &= \mathbb{E}_{\boldsymbol{\Phi}_m}\Big[\widetilde{\boldsymbol{\Phi}}_m^* \mathbb{E}_{\boldsymbol{\Phi}_{m'}}\Big[\boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_{m'}\boldsymbol{u}_{m'}}\Big] \widetilde{\boldsymbol{\Phi}}_m\Big] \\
&= K\|\boldsymbol{u}_{m'}\|_2^2 \mathbb{E}_{\boldsymbol{\Phi}_m}[\widetilde{\boldsymbol{\Phi}}_m^* \widetilde{\boldsymbol{\Phi}}_m] \\
&= K\|\boldsymbol{u}_{m'}\|_2^2 \mathbb{E}_{\boldsymbol{\Phi}_m}[\boldsymbol{\Phi}_m^* \boldsymbol{S}\boldsymbol{S}^* \boldsymbol{\Phi}_m] \\
&= K\|\boldsymbol{u}_{m'}\|_2^2 \mathrm{tr}(\boldsymbol{S}\boldsymbol{S}^*)\mathbf{I}_D = K^2\|\boldsymbol{u}_{m'}\|_2^2 \mathbf{I}_D,
\end{aligned}$$

where the second identity holds by Lemma B.1. This proves the first case.

Next we assume that $m' = m$. For notational simplicity, let $\boldsymbol{P} \in \mathbb{C}^{D \times D}$ denote the orthogonal projection onto the span of $\boldsymbol{u}_m$, i.e.,

$$\boldsymbol{P} = \frac{\boldsymbol{u}_m \boldsymbol{u}_m^*}{\|\boldsymbol{u}_m\|_2^2}.$$

Then by (A1) it follows that $\boldsymbol{\Phi}_m \boldsymbol{u}_m$ and $\boldsymbol{\Phi}_m(\mathbf{I}_D - \boldsymbol{P})$ are independent. Therefore

$$\begin{aligned}
&\mathbb{E}[\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m] \\
&= \mathbb{E}[(\boldsymbol{P} + \mathbf{I}_D - \boldsymbol{P})\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m(\boldsymbol{P} + \mathbf{I}_D - \boldsymbol{P})] \\
&= \underbrace{\mathbb{E}[\boldsymbol{P}\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m \boldsymbol{P}]}_{(\star)} \\
&\quad + \underbrace{\mathbb{E}[(\mathbf{I}_D - \boldsymbol{P})\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m} \widetilde{\boldsymbol{\Phi}}_m(\mathbf{I}_D - \boldsymbol{P})]}_{(\star\star)}.
\end{aligned}$$

The first term $(\star)$ is rewritten as

$$\mathbb{E}[\boldsymbol{g}^* \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}} \boldsymbol{S}^* \boldsymbol{g} \boldsymbol{u}_m \boldsymbol{u}_m^*],$$

where $\boldsymbol{g} \in \mathbb{C}^K$ is a standard complex Gaussian vector. On the other hand, we have

$$\begin{aligned}
\mathbb{E}[\boldsymbol{g}^* \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{g}} \boldsymbol{S}^* \boldsymbol{g}] &= \sum_{j_1,j_2,j_3,j_4=1}^{K} \mathbb{E}[\overline{g_{j_1}g_{j_2}}g_{j_3}g_{j_4}]\boldsymbol{e}_{j_1}^* \boldsymbol{C}_{\boldsymbol{e}_{j_2}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{e}_{j_3}} \boldsymbol{e}_{j_4} \\
&= 2K\|\boldsymbol{x}\|_2^2 + 2K(K-1)\|\boldsymbol{x}\|_2^2 = 2K^2\|\boldsymbol{x}\|_2^2.
\end{aligned}$$

Therefore,

$$(\star) = 2K^2\|\boldsymbol{x}\|_2^2 \boldsymbol{u}_m \boldsymbol{u}_m^*.$$

By the independence of $\boldsymbol{\Phi}_m \boldsymbol{u}_m$ and $\boldsymbol{\Phi}_m(\mathbf{I}_D - \boldsymbol{P})$ together with the commutativity of convolution, the second term $(\star\star)$ is computed as

$$(\star\star) = \mathbb{E}[(\mathbf{I}_D - \boldsymbol{P})\widetilde{\boldsymbol{\Phi}}_m^* \mathbb{E}[\boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}^* \boldsymbol{C}_{\widetilde{\boldsymbol{\Phi}}_m \boldsymbol{u}_m}] \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{\Phi}}_m(\mathbf{I}_D - \boldsymbol{P})]$$
$$= K\|\boldsymbol{u}_m\|_2^2 \mathbb{E}[(\mathbf{I}_D - \boldsymbol{P})\widetilde{\boldsymbol{\Phi}}_m^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{\Phi}}_m(\mathbf{I}_D - \boldsymbol{P})]$$
$$= K\|\boldsymbol{u}_m\|_2^2 \mathrm{tr}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*)(\mathbf{I}_D - \boldsymbol{P})$$
$$= K^2\|\boldsymbol{x}\|_2^2(\|\boldsymbol{u}_m\|_2^2\mathbf{I}_D - \boldsymbol{u}_m \boldsymbol{u}_m^*).$$

Collecting these results proves the second case and the proof is done. ∎

Appendix C. Proof of lemmas in section 3.

C.1. Proof of Lemma 3.9. By the homogeneity of $\rho_{x,w}$, we may assume that $\sigma_w = 1$. Let

$$\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}} = \boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{S}}^* \boldsymbol{q},$$

where $\widetilde{\boldsymbol{S}}$ is defined in (18). Then

$$\|\widetilde{\boldsymbol{S}}\boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \widetilde{\boldsymbol{S}}^*\| = \sup_{\boldsymbol{z},\boldsymbol{q} \in B_2^{3K}} |\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}}^* \boldsymbol{w}_m|.$$

We use Lemma A.6 to get a tail estimate.
Since

$$\|\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}} - \boldsymbol{a}_{\boldsymbol{z}',\boldsymbol{q}'}\|_2 \leq \|\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}} - \boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}'}\|_2 + \|\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}'} - \boldsymbol{a}_{\boldsymbol{z}',\boldsymbol{q}'}\|_2$$
$$\leq \|\boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{z}}^*\|\|\boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{S}}^*\|\|\boldsymbol{q} - \boldsymbol{q}'\|_2 + \|\boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{q}'}\|\|\boldsymbol{C}_{\boldsymbol{x}}^* \widetilde{\boldsymbol{S}}^*\|\|\boldsymbol{z} - \boldsymbol{z}'\|_2$$
$$\leq \sqrt{K\rho_x}(\|\boldsymbol{q} - \boldsymbol{q}'\|_2 + \|\boldsymbol{z} - \boldsymbol{z}'\|_2),$$

it follows that the γ_2 functional of the set $\{\boldsymbol{a}_{\boldsymbol{z},\boldsymbol{q}} \mid \boldsymbol{z},\boldsymbol{q} \in B_2^{3K}\}$ is upper bounded by

$$C_1\sqrt{K\rho_x}\int_0^\infty \sqrt{\log N(B_2^{3K}, \|\cdot\|_2, t)}dt \leq C_2 K\sqrt{\rho_x},$$

where we used a standard volume argument. The assertion follows from Lemma A.6 and a union bound argument.

C.2. Proof of Lemma 3.10. By the homogeneity, we may assume that $\sigma_x = \sigma_w = 1$. Then

$$\|\widetilde{\boldsymbol{S}}\boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{x}} \widetilde{\boldsymbol{S}}^*\| = \sup_{\boldsymbol{z} \in B_2^{3K}} |\boldsymbol{x}^* \boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{z}}^* \boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{z}} \boldsymbol{x}|$$

and

$$\|\widetilde{\boldsymbol{S}}\boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \widetilde{\boldsymbol{S}}^*\| = \sup_{\boldsymbol{z},\boldsymbol{q} \in B_2^{3K}} |\boldsymbol{x}^* \boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{q}}^* \boldsymbol{C}_{\widetilde{\boldsymbol{S}}^* \boldsymbol{z}} \boldsymbol{w}_m|$$

are suprema of second order chaos processes. We estimate their tail decay using Theorems 5.1 and 5.2. (For more details, see the proof of Lemma 5.9.) By Theorem 5.1 and a union bound argument, it follows that

$$\rho_x \leq C_\beta \sigma_x^2 \sqrt{KL} \log^5(KL)$$

holds with probability $1 - K^{-\beta}$. Similarly, by Theorem 5.2,

$$\rho_{x,w} \leq C_\beta \sigma_w \sigma_x \sqrt{KL} \log^5(MKL)$$

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

Moreover, by Lemma A.1, we also have that

$$\|\boldsymbol{x}\|_2^2 \geq \sigma_x^2 (L - \sqrt{2L\beta \log K})$$

holds with probability $1 - K^{-\beta}$. The assertions follow by assembling the above estimates.

Appendix D. Proof of Lemmas in section 5.

D.1. Proof of Lemma 5.3. First note that $\|\boldsymbol{\Psi}^* \boldsymbol{A} \boldsymbol{\Psi} - \mathbb{E}[\boldsymbol{\Psi}^* \boldsymbol{A} \boldsymbol{\Psi}]\|$ is written as a variational form given by

$$(59) \qquad \max_{\boldsymbol{q}, \boldsymbol{q}' \in B_2^D} |\boldsymbol{q}^* (\boldsymbol{\Psi}^* \boldsymbol{A} \boldsymbol{\Psi} - \mathbb{E}[\boldsymbol{\Psi}^* \boldsymbol{A} \boldsymbol{\Psi}]) \boldsymbol{q}'|.$$

For all $\boldsymbol{q} \in \mathbb{C}^D$, we have

$$\boldsymbol{\Psi} \boldsymbol{q} = (\boldsymbol{q}^\top \otimes \mathbf{I}_K) \mathrm{vec}(\boldsymbol{\Psi}).$$

Let $\boldsymbol{\psi} = \mathrm{vec}(\boldsymbol{\Psi})$. Then

$$\boldsymbol{q}^* \boldsymbol{\Psi}^* \boldsymbol{A} \boldsymbol{\Psi} \boldsymbol{q}' = \boldsymbol{\psi}^* (\boldsymbol{q}^\top \otimes \mathbf{I}_K)^* \boldsymbol{A} ((\boldsymbol{q}')^\top \otimes \mathbf{I}_K) \boldsymbol{\psi} = \boldsymbol{\psi}^* (\overline{\boldsymbol{q}} (\boldsymbol{q}')^\top \otimes \boldsymbol{A}) \boldsymbol{\psi}.$$

Then (59) becomes the supremum of the second order chaos process

$$\boldsymbol{\psi}^* (\overline{\boldsymbol{q}} (\boldsymbol{q}')^\top \otimes \boldsymbol{A}) \boldsymbol{\psi} - \mathbb{E}[\boldsymbol{\psi}^* (\overline{\boldsymbol{q}} (\boldsymbol{q}')^\top \otimes \boldsymbol{A}) \boldsymbol{\psi}]$$

over $\boldsymbol{q} \in B_2^D$. We obtain its tail estimate by applying Theorem 5.2 with

$$\Delta_1 = \{\boldsymbol{q}^\top \otimes \mathbf{I}_K \mid \boldsymbol{q} \in B_2^D\}$$

and

$$\Delta_2 = \{(\boldsymbol{q}')^\top \otimes \boldsymbol{A} \mid \boldsymbol{q}' \in B_2^D\}.$$

By direct calculation, we obtain

$$d_{\mathrm{S}}(\Delta_1) \leq 1, \quad d_{\mathrm{F}}(\Delta_1) \leq \sqrt{K},$$

and

$$d_{\mathrm{S}}(\Delta_2) \leq \|\boldsymbol{A}\|, \quad d_{\mathrm{F}}(\Delta_2) \leq \sqrt{K} \|\boldsymbol{A}\|.$$

Moreover

$$\gamma_2(\Delta_1, \|\cdot\|) \leq C_1 \int_0^\infty \sqrt{N(B_2^D, \|\cdot\|_2, t)} dt \leq C_1 \int_0^1 \sqrt{2D \log\left(1 + \frac{2}{t}\right)} dt \leq C_2 \sqrt{D}.$$

Similarly, we also have

$$\gamma_2(\Delta_2, \|\cdot\|) \leq C_2 \sqrt{D} \|\boldsymbol{A}\|.$$

The assertion follows from Theorem 5.2 with

$$(60) \qquad a = \sqrt{\frac{\gamma_2(\Delta_2, \|\cdot\|)d_{\mathrm{F}}(\Delta_2)}{\gamma_2(\Delta_1, \|\cdot\|)d_{\mathrm{F}}(\Delta_1)}}.$$

D.2. Proof of Lemma 5.4. Note that $\|\boldsymbol{\Upsilon}_{\mathrm{s},m} - \mathbb{E}[\boldsymbol{\Upsilon}_{\mathrm{s},m}]\|$ is written as a variational form given by

$$\sup_{\boldsymbol{z} \in B_2^K} |\boldsymbol{z}^*(\boldsymbol{\Upsilon}_{\mathrm{s},m} - \mathbb{E}[\boldsymbol{\Upsilon}_{\mathrm{s},m}])\boldsymbol{z}|.$$

By the commutativity of convolution, we have

$$
\begin{aligned}
\boldsymbol{z}^*\boldsymbol{\Upsilon}_{\mathrm{s},m}\boldsymbol{z} &= \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \boldsymbol{z}^*\boldsymbol{S}\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'}}\boldsymbol{C}^*_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'}}\boldsymbol{S}^*\boldsymbol{z} \\
&= \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \boldsymbol{u}^*_{m'}\boldsymbol{\Phi}^*_{m'}\boldsymbol{S}\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{C}^*_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{S}^*\boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'} \\
&= \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \mathrm{vec}(\boldsymbol{\Phi}_{m'})^*(\overline{\boldsymbol{u}_{m'}} \otimes \mathbf{I}_K)\boldsymbol{S}\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{C}^*_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{S}^*(\boldsymbol{u}^\top_{m'} \otimes \mathbf{I}_K)\mathrm{vec}(\boldsymbol{\Phi}_{m'}) \\
&= \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \mathrm{vec}(\boldsymbol{\Phi}_{m'})^*(\overline{\boldsymbol{u}_{m'}}\boldsymbol{u}^\top_{m'} \otimes \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{C}^*_{\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{S}^*)\mathrm{vec}(\boldsymbol{\Phi}_{m'}),
\end{aligned}
$$

where the third identity follows from

$$(61) \qquad \boldsymbol{\Phi}_{m'}\boldsymbol{u}_{m'} = (\boldsymbol{u}^\top_{m'} \otimes \mathbf{I}_K)\mathrm{vec}(\boldsymbol{\Phi}_{m'}).$$

Let

$$\boldsymbol{Q}(\boldsymbol{z}) = \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \boldsymbol{e}_{m'}\boldsymbol{e}^*_{m'} \otimes \boldsymbol{u}^\top_{m'} \otimes \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{C}^*_{\boldsymbol{S}^*\boldsymbol{z}}\boldsymbol{S}^*$$

and

$$\boldsymbol{\phi} = [\mathrm{vec}(\boldsymbol{\Phi}_1)^\top, \ldots, \mathrm{vec}(\boldsymbol{\Phi}_M)^\top]^\top.$$

Then $\boldsymbol{\phi}$ follows the distribution $\mathcal{CN}(\mathbf{0}_{MKD,1}, \mathbf{I}_{MKD})$ and

$$\boldsymbol{z}^*\boldsymbol{\Upsilon}_{\mathrm{s},m}\boldsymbol{z} = \boldsymbol{\phi}^*\boldsymbol{Q}(\boldsymbol{z})^*\boldsymbol{Q}(\boldsymbol{z})\boldsymbol{\phi}.$$

Therefore,

$$\sup_{\boldsymbol{z} \in B_2^K} |\boldsymbol{z}^*(\boldsymbol{\Upsilon}_{\mathrm{s},m} - \mathbb{E}[\boldsymbol{\Upsilon}_{\mathrm{s},m}])\boldsymbol{z}| = \sup_{\boldsymbol{z} \in B_2^K} |\boldsymbol{\phi}^*\boldsymbol{Q}(\boldsymbol{z})\boldsymbol{Q}(\boldsymbol{z})\boldsymbol{\phi}|.$$

We get a tail bound of the supremum of the second order chaos process $\boldsymbol{\phi}^*\boldsymbol{Q}(\boldsymbol{z})^*\boldsymbol{Q}(\boldsymbol{z})\boldsymbol{\phi} - \mathbb{E}[\boldsymbol{\phi}^*\boldsymbol{Q}(\boldsymbol{z})^*\boldsymbol{Q}(\boldsymbol{z})\boldsymbol{\phi}]$ by applying Theorem 5.1 with $\boldsymbol{M} = \boldsymbol{Q}(\boldsymbol{z})$ and $\Delta = \{\boldsymbol{Q}(\boldsymbol{z}) | \boldsymbol{z} \in B_2^K\}$.

Recall that in section 5.2 we defined $\breve{\boldsymbol{S}}$ by

$$\breve{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{0}_{K-1,L-K+1} & \boldsymbol{I}_{K-1} \\ \boldsymbol{I}_K & \boldsymbol{0}_{K,L-K} \end{bmatrix}.$$

Then the radius of Δ with respect to the Frobenius norm is upper bounded by

$$d_{\mathrm{F}}(\Delta) \leq \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{u}\|_2 \|\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\| \|\boldsymbol{S}\boldsymbol{C_{S^*z}}\|_{\mathrm{F}}$$

$$\leq \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{u}\|_2 \|\breve{\boldsymbol{S}}\boldsymbol{C_x^*}\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\|^{1/2}\sqrt{K}\|\boldsymbol{z}\|_2 \leq \|\boldsymbol{u}\|_2\sqrt{\rho_x K},$$

where the first inequality follows from the identity $\boldsymbol{C_{S^*z}^*}\boldsymbol{S}^* = \breve{\boldsymbol{S}}^*\breve{\boldsymbol{S}}\boldsymbol{C_{S^*z}^*}\boldsymbol{S}^*$.

Let $\| \cdot \|_{2,\infty}$ be defined in (39). Then the radius of Δ with respect to the spectral norm is upper bounded by

$$(62) \qquad \begin{aligned} d_{\mathrm{S}}(\Delta) &\leq \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{u}\|_{2,\infty} \|\boldsymbol{C_x}\breve{\boldsymbol{S}}^*\| \|\boldsymbol{S}\boldsymbol{C_{S^*z}}\| \\ &\leq \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{u}\|_{2,\infty}\sqrt{\rho_x}\sqrt{L}\|\boldsymbol{F}\boldsymbol{S}^*\boldsymbol{z}\|_\infty \\ &\leq \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{u}\|_{2,\infty}\sqrt{\rho_x}\|\boldsymbol{z}\|_1 \leq \|\boldsymbol{u}\|_{2,\infty}\sqrt{\rho_x K}, \end{aligned}$$

where the second inequality follows from the identity $\boldsymbol{C_{S^*z}} = \sqrt{L}\boldsymbol{F}^*\mathrm{diag}(\boldsymbol{F}\boldsymbol{S}^*\boldsymbol{z})\boldsymbol{F}$ and the third inequality follow from $\|\boldsymbol{F} : \ell_1^L \to \ell_\infty^L\| = \frac{1}{\sqrt{L}}$.

By (62) and Dudley's inequality, the γ_2 functional of Δ is upper bounded by

$$\gamma_2(\Delta, \| \cdot \|) \leq C_1 \|\boldsymbol{u}\|_{2,\infty}\sqrt{\rho_x}\sqrt{L} \int_0^\infty \sqrt{\log N(\boldsymbol{F}\boldsymbol{S}^*B_2^K, \| \cdot \|_\infty, t)}\,dt$$

$$\leq C_2 \|\boldsymbol{u}\|_{2,\infty}\sqrt{\rho_x}\sqrt{K}\sqrt{\log K}\log^{3/2} L,$$

where the last step follows from Corollary A.5. The assertion follows by applying these estimates to Theorem 5.1 together with a union bound argument over $m = 1, \ldots, M$.

D.3. Proof of Lemma 5.5. Note that $\|\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m]\|$ is written as a variational form given by

$$(63) \qquad \sup_{\boldsymbol{z} \in B_2^{2L-1}, \boldsymbol{q} \in B_2^D} |\boldsymbol{z}^*\boldsymbol{Z}_m\boldsymbol{q} - \mathbb{E}[\boldsymbol{z}^*\boldsymbol{Z}_m\boldsymbol{q}]|.$$

Let $\boldsymbol{\phi}_m = \mathrm{vec}(\boldsymbol{\Phi}_m)$. Then

$$\boldsymbol{z}^*\boldsymbol{Z}_m\boldsymbol{q} = \boldsymbol{z}^*\breve{\boldsymbol{S}}\boldsymbol{C_{S^*\boldsymbol{\Phi}_m\boldsymbol{u}_m}^*}\boldsymbol{S}^*\boldsymbol{\Phi}_m\boldsymbol{q} = \boldsymbol{\phi}_m^*(\overline{\boldsymbol{u}_m}\boldsymbol{q}^\top \otimes \boldsymbol{S}\boldsymbol{C_{\breve{\boldsymbol{S}}^*z}^*}\boldsymbol{S}^*)\boldsymbol{\phi}_m.$$

Then (63) becomes the supremum of the second order chaos process

$$\boldsymbol{\phi}_m^*(\boldsymbol{u}_m^\top \otimes \boldsymbol{S}\boldsymbol{C_{\breve{\boldsymbol{S}}^*z}^*}\boldsymbol{S}^*)^*(\boldsymbol{q}^\top \otimes \boldsymbol{I}_K)\boldsymbol{\phi}_m - \mathbb{E}[\boldsymbol{\phi}_m^*(\boldsymbol{u}_m^\top \otimes \boldsymbol{S}\boldsymbol{C_{\breve{\boldsymbol{S}}^*z}^*}\boldsymbol{S}^*)^*(\boldsymbol{q}^\top \otimes \boldsymbol{I}_K)\boldsymbol{\phi}_m]$$

over $\boldsymbol{z} \in B_2^{2K-1}$ and $\boldsymbol{q} \in B_2^D$. We obtain its tail estimate by applying Theorem 5.2 with

$$\Delta_1 = \{\boldsymbol{u}_m^\top \otimes \boldsymbol{S}\boldsymbol{C}_{\breve{\boldsymbol{S}}^*\boldsymbol{z}}^* \boldsymbol{S}^* \mid \boldsymbol{z} \in B_2^{2K-1}\}$$

and

$$\Delta_2 = \{\boldsymbol{q}^\top \otimes \mathbf{I}_K \mid \boldsymbol{q} \in B_2^D\}.$$

The radii of Δ_1 and Δ_2 are upper bounded by

$$d_{\mathrm{S}}(\Delta_1) \le d_{\mathrm{F}}(\Delta_1) \le \|\boldsymbol{u}\|_{2,\infty}\sqrt{K}$$

and

$$d_{\mathrm{S}}(\Delta_2) \le 1, \quad d_{\mathrm{F}}(\Delta_2) \le \sqrt{K}.$$

Moreover, since

$$\|\boldsymbol{u}_m^\top \otimes \boldsymbol{S}\boldsymbol{C}_{\breve{\boldsymbol{S}}^*\boldsymbol{z}}^* \boldsymbol{S}^*\| \le \|\boldsymbol{u}\|_{2,\infty}\sqrt{L}\|\boldsymbol{F}\breve{\boldsymbol{S}}^*\boldsymbol{z}\|_\infty,$$

we have

$$\begin{aligned}
\gamma_2(\Delta_1, \|\cdot\|) &\le C_1\|\boldsymbol{u}\|_{2,\infty}\sqrt{L}\int_0^\infty \sqrt{N(\boldsymbol{F}\breve{\boldsymbol{S}}^* B_2^{2K-1}, \|\cdot\|_\infty, t)}\, dt \\
&\le C_2\|\boldsymbol{u}\|_{2,\infty}\sqrt{LK}\int_0^\infty \sqrt{N(\boldsymbol{F}\breve{\boldsymbol{S}}^* B_1^{2K-1}, \|\cdot\|_\infty, t)}\, dt \\
&\le C_3\|\boldsymbol{u}\|_{2,\infty}\sqrt{K}\sqrt{\log(2K-1)}\log^{3/2} L,
\end{aligned}$$

where the last step follows from Corollary A.5.

On the other hand, we also have

$$\gamma_2(\Delta_2, \|\cdot\|) \le C_1\int_0^\infty \sqrt{N(B_2^D, \|\cdot\|_2, t)}\, dt \le C_1\int_0^1 \sqrt{2D\log\left(1+\frac{2}{t}\right)}\, dt \le C_4\sqrt{D}.$$

By applying the estimates to Theorem 5.2 with a given in (60), we obtain that the supremum is upper bounded by

$$\begin{aligned}
C(\|\boldsymbol{u}\|_{2,\infty}K^{3/2}\sqrt{D}\sqrt{\log K}\log^{3/2} L + \|\boldsymbol{u}\|_{2,\infty}K\sqrt{\log K}\log^{3/2} L \\
+ \|\boldsymbol{u}\|_{2,\infty}\sqrt{KD})\log(\log M + \beta\log K)
\end{aligned}$$

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

The assertion follows by applying a union bound argument over $m = 1, \ldots, M$.

D.4. Proof of Lemma 5.6. Note that the spectral norm of $\sum_{m=1}^M \boldsymbol{e}_m^* \otimes (\boldsymbol{Z}_m - \mathbb{E}[\boldsymbol{Z}_m])$ admits a variational form given by

$$(64) \qquad \sup_{\boldsymbol{z}\in B_2^{2L-1}, \boldsymbol{v}\in B_2^{MD}} \left| \sum_{m=1}^M \boldsymbol{z}^*\boldsymbol{Z}_m\boldsymbol{v}_m - \mathbb{E}[\boldsymbol{z}^*\boldsymbol{Z}_m\boldsymbol{v}_m] \right|,$$

where $\boldsymbol{v}_m \in \mathbb{C}^D$ for $m = 1, \ldots, M$ and $\boldsymbol{v} = [\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_M^\top]^\top$.

Let $\boldsymbol{\phi}_m = \mathrm{vec}(\boldsymbol{\Phi}_m)$ for $m = 1, \ldots, M$ and $\boldsymbol{\phi} = [\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_M^\top]^\top$. Then as before

$$
\begin{aligned}
\sum_{m=1}^{M} \boldsymbol{z}^* \boldsymbol{Z}_m \boldsymbol{v}_m &= \sum_{m=1}^{M} \boldsymbol{z}^* \boldsymbol{\breve{S}} C^*_{\boldsymbol{S}^* \boldsymbol{\Phi}_m \boldsymbol{u}_m} \boldsymbol{S}^* \boldsymbol{\Phi}_m \boldsymbol{v}_m \\
&= \sum_{m=1}^{M} \boldsymbol{\phi}_m^* (\overline{\boldsymbol{u}_m} \boldsymbol{v}_m^\top \otimes \boldsymbol{S} C^*_{\boldsymbol{\breve{S}}^* \boldsymbol{z}} \boldsymbol{S}^*) \boldsymbol{\phi}_m \\
&= \boldsymbol{\phi}^* \left(\sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{u}_m^\top \otimes \boldsymbol{S} C^*_{\boldsymbol{\breve{S}}^* \boldsymbol{z}} \boldsymbol{S}^* \right)^* \left(\sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{v}_m^\top \otimes \mathbf{I}_K \right) \boldsymbol{\phi}.
\end{aligned}
$$

Then (64) becomes the supremum of a second order chaos process. We obtain its tail estimate by applying Theorem 5.2 with

$$
\Delta_1 = \left\{ \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{u}_m^\top \otimes \boldsymbol{S} C^*_{\boldsymbol{\breve{S}}^* \boldsymbol{z}} \boldsymbol{S}^* \mid \boldsymbol{z} \in B_2^{2K-1} \right\}
$$

and

$$
\Delta_2 = \left\{ \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{v}_m^\top \otimes \mathbf{I}_K \mid \boldsymbol{v} \in B_2^{MD} \right\}.
$$

Let $\| \cdot \|_{2,\infty}$ be defined in (39). Then the radii of Δ_1 and Δ_2 are upper bounded by

$$
d_{\mathrm{S}}(\Delta_1) \leq \|\boldsymbol{u}\|_{2,\infty} \sqrt{K}, \quad d_{\mathrm{F}}(\Delta_1) \leq \|\boldsymbol{u}\|_2 \sqrt{K} \leq \|\boldsymbol{u}\|_{2,\infty} \sqrt{MK},
$$

and

$$
d_{\mathrm{S}}(\Delta_2) \leq 1, \quad d_{\mathrm{F}}(\Delta_2) \leq \sqrt{K}.
$$

Moreover, since

$$
\Big\| \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{u}_m^\top \otimes \boldsymbol{S} C^*_{\boldsymbol{\breve{S}}^* \boldsymbol{z}} \boldsymbol{S}^* \Big\| \leq \|\boldsymbol{u}\|_{2,\infty} \sqrt{L} \|\boldsymbol{F} \boldsymbol{\breve{S}}^* \boldsymbol{z}\|_\infty,
$$

we have

$$
\begin{aligned}
\gamma_2(\Delta_1, \| \cdot \|) &\leq C_1 \|\boldsymbol{u}\|_{2,\infty} \sqrt{L} \int_0^\infty \sqrt{N(\boldsymbol{F} \boldsymbol{\breve{S}}^* B_2^{2K-1}, \| \cdot \|_\infty, t)} dt \\
&\leq C_2 \|\boldsymbol{u}\|_{2,\infty} \sqrt{LK} \int_0^\infty \sqrt{N(\boldsymbol{F} \boldsymbol{\breve{S}}^* B_1^{2K-1}, \| \cdot \|_\infty, t)} dt \\
&\leq C_3 \|\boldsymbol{u}\|_{2,\infty} \sqrt{K} \sqrt{\log(2K-1)} \log^{3/2} L,
\end{aligned}
$$

where the last step follows from Corollary A.5.

On the other hand, since

$$
\Big\| \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{v}_m^\top \otimes \mathbf{I}_K \Big\| \leq \|\boldsymbol{v}\|_{2,\infty},
$$

we also have

$$\gamma_2(\Delta_2, \|\cdot\|) \leq C_1 \int_0^\infty \sqrt{N(B_2^{MD}, \|\cdot\|_{2,\infty}, t)}dt \leq C_4\sqrt{D}\sqrt{\log D}\log(MD).$$

By applying these estimates to Theorem 5.2 with a given in (60), we obtain that the supremum is upper bounded by

$$C'(\beta)\log^\alpha(MKL)\|\boldsymbol{u}\|_{2,\infty}(M^{1/4}K^{3/4}D^{1/4} + K + \sqrt{MKD})$$

with probability $1 - K^{-\beta}$. Finally, by the inequality of arithmetic and geometric means, we have

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

This completes the proof.

D.5. Proof of Lemma 5.7. Similar to the proof of Lemma 5.4, we rewrite $\|\boldsymbol{\Upsilon}_{\mathrm{c},m}\|$ as a variational form given by

$$\sup_{\boldsymbol{z}_1, \boldsymbol{z}_2 \in B_2^K} |\boldsymbol{z}_1^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{z}_2|.$$

By the commutativity of convolution and (61), $\boldsymbol{z}_1^* \boldsymbol{\Upsilon}_{\mathrm{c},m} \boldsymbol{z}_2$ is written as follows:

$$\sum_{\substack{m'=1 \\ m' \neq m}}^M \boldsymbol{z}_1^* \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{\Phi}_{m'} \boldsymbol{u}_{m'}}^* \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{S}^* \boldsymbol{z}_2 = \sum_{\substack{m'=1 \\ m' \neq m}}^M \mathrm{vec}(\boldsymbol{\Phi}_{m'})^*(\overline{\boldsymbol{u}_{m'}} \otimes \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}_1} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{S}^* \boldsymbol{z}_2),$$

which, conditional on $\boldsymbol{w}$, is a centered Gaussian process indexed by $(\boldsymbol{z}_1, \boldsymbol{z}_2) \in B_2^K \times B_2^K$. We compute a tail estimate of the supremum by applying Lemma A.6 with

$$\Delta = \{\boldsymbol{f}(\boldsymbol{z}_1, \boldsymbol{z}_2) \mid \boldsymbol{z}_1, \boldsymbol{z}_2 \in B_2^K\},$$

where

$$\boldsymbol{f}(\boldsymbol{z}_1, \boldsymbol{z}_2) = \sum_{\substack{m'=1 \\ m' \neq m}}^M \boldsymbol{e}_{m'} \otimes \overline{\boldsymbol{u}_{m'}} \otimes \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}_1} \boldsymbol{S}^* \boldsymbol{z}_2.$$

Since $\|\boldsymbol{C}_{\boldsymbol{S}^*(\boldsymbol{z}_j - \boldsymbol{z}_j')}\| = \sqrt{L}\|\boldsymbol{F}\boldsymbol{S}^*(\boldsymbol{z}_j - \boldsymbol{z}_j')\|_\infty$ for $j = 1, 2$ and hence

$$\|\boldsymbol{f}(\boldsymbol{z}_1, \boldsymbol{z}_2) - \boldsymbol{f}(\boldsymbol{z}_1', \boldsymbol{z}_2')\|_2 \leq \|\boldsymbol{f}(\boldsymbol{z}_1, \boldsymbol{z}_2) - \boldsymbol{f}(\boldsymbol{z}_1', \boldsymbol{z}_2)\|_2 + \|\boldsymbol{f}(\boldsymbol{z}_1', \boldsymbol{z}_2) - \boldsymbol{f}(\boldsymbol{z}_1', \boldsymbol{z}_2')\|_2$$
$$\leq \|\boldsymbol{u}\|_2 \rho_{x,w} \sqrt{L}(\|\boldsymbol{F}\boldsymbol{S}^*(\boldsymbol{z}_1 - \boldsymbol{z}_1')\|_\infty + \|\boldsymbol{F}\boldsymbol{S}^*(\boldsymbol{z}_2 - \boldsymbol{z}_2')\|_\infty),$$

it follows that

$$\int_0^\infty \sqrt{\log N(\Delta, \|\cdot\|_2, t)}dt \leq C_1\|\boldsymbol{u}\|_2\rho_{x,w}\sqrt{L}\int_0^\infty \sqrt{\log N(\boldsymbol{F}\boldsymbol{S}^* B_2^K, \|\cdot\|_\infty, t)}dt$$
$$\leq C_1\|\boldsymbol{u}\|_2\rho_{x,w}\sqrt{KL}\int_0^\infty \sqrt{\log N(\boldsymbol{F}\boldsymbol{S}^* B_1^K, \|\cdot\|_\infty, t)}dt$$
$$\leq C_2\|\boldsymbol{u}\|_2\rho_{x,w}\sqrt{K}\sqrt{\log K}\log^{3/2} L,$$

where the last step follows from Corollary A.5 together with the observation that $\|\boldsymbol{F}\|_{\ell_1^L \to \ell_\infty^L} \leq L^{-1/2}$. Then the assertion follows from Lemma A.6 with a union bound argument over $m = 1, \ldots, M$.

D.6. Proof of Lemma 5.8. Let $\phi_m = \text{vec}(\boldsymbol{\Phi}_m)$ for $m = 1, \ldots, M$ and $\boldsymbol{\phi} = [\phi_1^\top, \ldots, \phi_M^\top]^\top$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M \in \mathbb{C}^D$ and $\boldsymbol{v} = [\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_M^\top] \in B_2^{MD}$. Then the spectral norm of $\sum_{m=1}^M \boldsymbol{e}_m^* \otimes \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m$ is rewritten as

$$
\left\| \sum_{m=1}^M \boldsymbol{e}_m^* \otimes \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m \right\| = \sup_{\boldsymbol{z} \in B_2^{2K-1}} \sup_{\boldsymbol{v} \in B_2^{MD}} \left| \sum_{m=1}^M \boldsymbol{z}^* \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m \boldsymbol{v}_m \right|
$$

$$
= \sup_{\boldsymbol{z} \in B_2^{2K-1}} \sup_{\boldsymbol{v} \in B_2^{MD}} \left| \sum_{m=1}^M (\boldsymbol{v}_m^\top \otimes \boldsymbol{z}^* \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*) \phi_m \right|
$$

$$
= \sup_{\boldsymbol{z} \in B_2^{2K-1}} \sup_{\boldsymbol{v} \in B_2^{MD}} \left| \sum_{m=1}^M \boldsymbol{e}_m^* \otimes (\boldsymbol{v}_m^\top \otimes \boldsymbol{z}^* \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^*) \boldsymbol{\phi} \right|.
$$

Let

$$
\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v}) = \sum_{m=1}^M \boldsymbol{e}_m \otimes (\overline{\boldsymbol{v}_m} \otimes \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_m} \breve{\boldsymbol{S}}^* \boldsymbol{z}).
$$

Then we obtain

$$
\left\| \sum_{m=1}^M \boldsymbol{e}_m^* \otimes \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \boldsymbol{\Phi}_m \right\| = \sup_{\boldsymbol{v} \in B_2^{MD}} \sup_{\boldsymbol{z} \in B_2^{2K-1}} |\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v})^* \boldsymbol{\phi}|.
$$

Note that $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v})^* \boldsymbol{\phi}$, conditioned on $\boldsymbol{w}$, is a centered Gaussian process. We compute a tail estimate of this supremum by applying Lemma A.6 with

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

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

$$
\rho_{x,w} = \max_{1 \le m \le M} \| \widetilde{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{x}}^* \boldsymbol{C}_{\boldsymbol{w}_1} \widetilde{\boldsymbol{S}}^* \| \ge \| \breve{\boldsymbol{S}} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{S}^* \| \quad \forall m = 1, \ldots, M.
$$

By the triangle inequality, we obtain

$$
\| \boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v}) - \boldsymbol{f}(\boldsymbol{z}', \boldsymbol{v}') \|_2 \le \| \boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v}) - \boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v}') \|_2 + \| \boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v}') - \boldsymbol{f}(\boldsymbol{z}', \boldsymbol{v}') \|_2
$$

$$
\le \rho_{x,w} (\| \boldsymbol{z} \|_2 \| \boldsymbol{v} - \boldsymbol{v}' \|_2 + \| \boldsymbol{z} - \boldsymbol{z}' \|_2 \| \boldsymbol{v}' \|_2)
$$

$$
\le \rho_{x,w} (\| \boldsymbol{v} - \boldsymbol{v}' \|_2 + \| \boldsymbol{z} - \boldsymbol{z}' \|_2).
$$

The integral of the log-entropy number is computed as

$$
\sup_{\boldsymbol{v} \in B_2^{MD}} \sup_{\boldsymbol{z} \in B_2^{2K-1}} |\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{v})^* \boldsymbol{\phi}|
$$

$$
\le C_1 \int_0^\infty \sqrt{\log N(\Delta, \| \cdot \|_2, t)} dt
$$

$$
\le C_1 \rho_{x,w} \left(\int_0^\infty \sqrt{\log N(B_2^{MD}, \| \cdot \|_2, t)} dt + \int_0^\infty \sqrt{\log N(B_2^{2K-1}, \| \cdot \|_2, t)} dt \right)
$$

$$
\le C_2 \rho_{x,w} (\sqrt{MD} + \sqrt{K}),
$$

where the last step follows from a standard volume argument. Then the assertion follows from Lemma A.6.

D.7. Proof of Lemma 5.9. By the homogeneity of ρ_w, we may assume that $\sigma_w = 1$. We first consider the case that $m' = m$. Let $\boldsymbol{z} \in \mathbb{C}^K$. Then by the commutativity of convolution, we have

$$\boldsymbol{C}_{\boldsymbol{w}_m} \boldsymbol{S}^* \boldsymbol{z} = \boldsymbol{w}_m \circledast \boldsymbol{S}^* \boldsymbol{z} = \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \boldsymbol{w}_m.$$

Therefore, the spectral norm of $\boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m}]) \boldsymbol{S}^*$ is rewritten as

$$\|\boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m}]) \boldsymbol{S}^*\| = \sup_{\boldsymbol{z} \in B_2^K} |\boldsymbol{z}^* \boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m}]) \boldsymbol{S}^* \boldsymbol{z}|$$

$$= \sup_{\boldsymbol{z} \in B_2^K} |\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \boldsymbol{w}_m - \mathbb{E}[\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \boldsymbol{w}_m]|,$$

where the last term is the supremum of a second order chaos. We use Theorem 5.1 to get its tail estimate. Define

$$\Delta_{\boldsymbol{z}} = \{\boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \mid \boldsymbol{z} \in B_2^K\}.$$

Then the radii of $\Delta_{\boldsymbol{z}}$ with respect to the spectral and Frobenius norms are given by

$$d_{\mathrm{S}}(\Delta_{\boldsymbol{z}}) = \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}}\| = \sup_{\boldsymbol{z} \in B_2^K} \sqrt{L} \|\boldsymbol{F} \boldsymbol{S}^* \boldsymbol{z}\|_\infty = \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{z}\|_1 = \sqrt{K},$$

$$d_{\mathrm{F}}(\Delta_{\boldsymbol{z}}) = \sup_{\boldsymbol{z} \in B_2^K} \|\boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}}\|_{\mathrm{F}} = \sup_{\boldsymbol{z} \in B_2^K} \sqrt{L} \|\boldsymbol{z}\|_2 = \sqrt{L}.$$

Moreover, the γ_2 functional of $\Delta_{\boldsymbol{z}}$ is bounded by

$$\gamma_2(\Delta_{\boldsymbol{z}}, \|\cdot\|) \le C_1 \sqrt{L} \int_0^\infty \sqrt{\log N(\boldsymbol{F} \boldsymbol{S}^* B_2^K, \|\cdot\|_\infty, t)} dt$$

$$\le C_1 \sqrt{KL} \int_0^\infty \sqrt{\log N(\boldsymbol{F} \boldsymbol{S}^* B_1^K, \|\cdot\|_\infty, t)} dt$$

$$\le C_1 \sqrt{K} \sqrt{\log K} \log^{3/2} L,$$

where the last step follows from Lemma A.4. By Theorem 5.1,

$$\|\boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_m}]) \boldsymbol{S}^*\| \le C \sigma_w^2 K (1 + \log K) \log^3 L (1 + 2 \log M + \beta \log K)$$

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

Next we consider the case where $m' \ne m$. In this case, we have $\mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}}] = \boldsymbol{0}_{L,L}$. Similarly to the previous case, the spectral norm of $\boldsymbol{S} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{S}^*$ is rewritten as

$$\|\boldsymbol{S} \boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{S}^*\| = \sup_{\boldsymbol{z}, \boldsymbol{q} \in B_2^K} |\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{q}}^* \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \boldsymbol{w}_{m'}| = \sup_{\boldsymbol{z}, \boldsymbol{q} \in B_2^K} |\widetilde{\boldsymbol{w}}_{m,m'}^* \boldsymbol{L}_{\boldsymbol{q}}^* \boldsymbol{R}_{\boldsymbol{z}} \widetilde{\boldsymbol{w}}_{m,m'}|,$$

where

$$\widetilde{\boldsymbol{w}}_{m,m'} = \begin{bmatrix} \boldsymbol{w}_m \\ \boldsymbol{w}_{m'} \end{bmatrix}, \quad \boldsymbol{L}_{\boldsymbol{q}} = \begin{bmatrix} \boldsymbol{0}_{L,L} & \boldsymbol{0}_{L,L} \\ \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{q}} & \boldsymbol{0}_{L,L} \end{bmatrix}, \quad \boldsymbol{R}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{0}_{L,L} & \boldsymbol{C}_{\boldsymbol{S}^* \boldsymbol{z}} \\ \boldsymbol{0}_{L,L} & \boldsymbol{0}_{L,L} \end{bmatrix}.$$

Define

$$\Delta_{\boldsymbol{L}_{\boldsymbol{q}}} = \{\boldsymbol{L}_{\boldsymbol{q}} \mid \boldsymbol{q} \in B_2^K\}.$$

and

$$\Delta_{\boldsymbol{R_z}} = \{\boldsymbol{R_z} \mid \boldsymbol{z} \in B_2^K\}.$$

Then, the radii and γ_2 functional of $\Delta_{\boldsymbol{L_q}}$ and $\Delta_{\boldsymbol{R_z}}$ are identical to those of $\Delta_{\boldsymbol{z}}$. Therefore, by Theorem 5.2,

$$\|\boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}} - \mathbb{E}[\boldsymbol{C}_{\boldsymbol{w}_m}^* \boldsymbol{C}_{\boldsymbol{w}_{m'}}])\boldsymbol{S}^*\| \le C\sigma_w^2 K(1 + \log K)\log^3 L(1 + 2\log M + \beta \log K)$$

holds with probability $1 - M^{-2}K^{-\beta}$. The assertion follows by applying a union bound argument.

D.8. Proof of Lemma 5.10. Without loss of generality, we assume that $\sigma_w = 1$. Similarly to the proof of Lemma 5.9 in Appendix D.7, we can rewrite $M\bar{\rho}_w$ as the supremum of a second order chaos process as follows:

$$M\bar{\rho}_w = \sup_{\boldsymbol{z} \in B_2^K} \left| \sum_{m=1}^M \boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}} \boldsymbol{w}_m - \mathbb{E}[\boldsymbol{w}_m^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}} \boldsymbol{w}_m] \right|$$

$$= \sup_{\boldsymbol{z} \in B_2^K} \left| \boldsymbol{w}^*(\mathbf{I}_M \otimes \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}})\boldsymbol{w} - \mathbb{E}[\boldsymbol{w}^*(\mathbf{I}_M \otimes \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}^* \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}})\boldsymbol{w}] \right|,$$

where $\boldsymbol{w} = [\boldsymbol{w}_1^\top, \ldots, \boldsymbol{w}_M^\top]^\top$ is a standard complex Gaussian vector of length ML.

Define

$$\widetilde{\Delta}_{\boldsymbol{z}} = \{\mathbf{I}_M \otimes \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}} \mid \boldsymbol{z} \in B_2^K\}.$$

Then the radii of $\widetilde{\Delta}_{\boldsymbol{z}}$ with respect to the spectral and Frobenius norms are upper bounded respectively by

$$d_{\mathrm{S}}(\widetilde{\Delta}_{\boldsymbol{z}}) = \sup_{\boldsymbol{z} \in B_2^K} \sqrt{L}\|\boldsymbol{F}\boldsymbol{S}^*\boldsymbol{z}\|_\infty \le \sqrt{K},$$

$$d_{\mathrm{F}}(\widetilde{\Delta}_{\boldsymbol{z}}) = \sup_{\boldsymbol{z} \in B_2^K} \sqrt{M}\|\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\|_{\mathrm{F}} \le \sqrt{ML}.$$

Moreover, since $\|\mathbf{I}_M \otimes \boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\| = \|\boldsymbol{C}_{\boldsymbol{S}^*\boldsymbol{z}}\|$, the γ_2 functional of $\widetilde{\Delta}_{\boldsymbol{z}}$ is upper bounded by

$$\gamma_2(\widetilde{\Delta}_{\boldsymbol{z}}, \|\cdot\|) = \gamma_2(\Delta_{\boldsymbol{z}}, \|\cdot\|) \le C_1\sqrt{K}\sqrt{\log K}\log^{3/2} L,$$

where the last step has been shown in Appendix D.7. By Theorem 5.1,

$$\bar{\rho}_w \le C\beta\sigma_w^2 M^{-1/2}\sqrt{KL}(1 + \log K)\log^3 L \log K$$

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

D.9. Proof of Lemma 5.11. First we rewrite $\boldsymbol{\Phi v}$ as

$$\boldsymbol{\Phi v} = \boldsymbol{W_v \phi},$$

where

$$\boldsymbol{\phi} = [\mathrm{vec}(\boldsymbol{\Phi}_1)^\top, \ldots, \mathrm{vec}(\boldsymbol{\Phi}_1)^\top]^\top$$

and

$$\boldsymbol{W_v} = \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{v}_m^\top \otimes \mathbf{I}_K.$$

Then it follows that

$$\boldsymbol{v}^* \boldsymbol{\Phi}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{\Phi} \boldsymbol{v} = \boldsymbol{\phi}^* \boldsymbol{W_v}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} \boldsymbol{\phi},$$

where the latter, conditional on $\boldsymbol{w}$, is a quadratic Gaussian form. Furthermore, by direct calculation, we have

$$\mathbb{E}_{\boldsymbol{\phi}} [\boldsymbol{\phi}^* \boldsymbol{W_v}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} \boldsymbol{\phi}] = \mathbb{E}_{\boldsymbol{\phi}} [\boldsymbol{\Phi}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{\Phi}] = \text{tr}(\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \mathbf{I}_{MD} = \boldsymbol{0}_{MD,MD}.$$

Then $\|\boldsymbol{\Phi}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{\Phi}\|$ is written as

$$\sup_{\boldsymbol{v} \in B_2^{MD}} |\boldsymbol{\phi}^* \boldsymbol{W_v}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} \boldsymbol{\phi} - \mathbb{E}_{\boldsymbol{\phi}} [\boldsymbol{\phi}^* \boldsymbol{W_v}^* (\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} \boldsymbol{\phi}]|,$$

which is the supremum of a second order Gaussian chaos process. We compute its tail estimate by applying Theorem 5.2 with

$$\Delta_1 = \{\boldsymbol{W_v} \mid \boldsymbol{v} \in B_2^{MD}\}$$

and

$$\Delta_2 = \{(\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} \mid \boldsymbol{v} \in B_2^{MD}\}.$$

Let $\| \cdot \|_{p,q}$ be defined in (39). Then the radii of Δ_1 with respect to the Frobenius and spectral norms are upper bounded respectively by

$$d_\text{S}(\Delta_1) \leq \sup_{\boldsymbol{v} \in B_2^{MD}} \|\boldsymbol{v}\|_{2,\infty} \leq 1,$$

$$d_\text{F}(\Delta_1) \leq \sup_{\boldsymbol{v} \in B_2^{MD}} \sqrt{K} \|\boldsymbol{v}\|_2 \leq \sqrt{K}.$$

By Lemma E.1, the γ_2 functional is bounded by

$$\gamma_2(\Delta_1, \| \cdot \|) \leq C_1 \int_0^\infty \sqrt{\log N(B_2^{MD}, \| \cdot \|_{2,\infty}, t)} dt \leq C_2 \sqrt{D} \sqrt{\log D} \log(MD).$$

We repeat the calculation for Δ_2. Note that $(\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v}$ is expressed as

$$(\boldsymbol{Y}_\text{n}^* \boldsymbol{Y}_\text{n} - \boldsymbol{\Lambda}) \boldsymbol{W_v} = \sum_{m=1}^{M} \boldsymbol{e}_m \boldsymbol{e}_m^* \otimes \boldsymbol{v}_m^\top \otimes \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{S} (\boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* - \|\boldsymbol{w}_{m'}\|_2^2 \mathbf{I}_L) \boldsymbol{S}^*$$

$$- \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{e}_m \boldsymbol{e}_{m'}^* \otimes \boldsymbol{v}_{m'}^\top \otimes \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{w}_m} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* \boldsymbol{S}^*.$$

Noting that all the summands in this decomposition are $K \times K$ matrices and orthogonal with respect to the Frobenius inner product, we obtain that the Frobenius norm of $(\boldsymbol{Y}_{\mathrm{n}}^* \boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda}) \boldsymbol{W_v}$ is upper bounded by

$$\|(\boldsymbol{Y}_{\mathrm{n}}^* \boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda}) \boldsymbol{W_v}\|_{\mathrm{F}}^2 \leq \sum_{m=1}^{M} \|\boldsymbol{v}_m\|_2^2 K \Big\| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* - \|\boldsymbol{w}_{m'}\|_2^2 \mathbf{I}_L) \boldsymbol{S}^* \Big\|^2$$

$$+ \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \|\boldsymbol{v}_{m'}\|_2^2 K \|\boldsymbol{S} \boldsymbol{C}_{\boldsymbol{w}_m} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* \boldsymbol{S}^*\|^2$$

$$\leq 2\|\boldsymbol{v}\|_2^2 K(M^2 \bar{\rho}_w^2 + \rho_w^2) + \|\boldsymbol{v}\|_2^2 K M \rho_w^2,$$

which implies that

$$\|(\boldsymbol{Y}_{\mathrm{n}}^* \boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda}) \boldsymbol{W_v}\|_{\mathrm{F}} \leq C \|\boldsymbol{v}\|_2 \sqrt{K} (M \bar{\rho}_w + \sqrt{M} \rho_w).$$

On the other hand, it follows from the block Gershgorin disk theorem [11] that

$$\|(\boldsymbol{Y}_{\mathrm{n}}^* \boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda}) \boldsymbol{W_v}\|$$

$$\leq \max_{1 \leq m \leq M} \left(\|\boldsymbol{v}_m\|_2 \Big\| \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \boldsymbol{S}(\boldsymbol{C}_{\boldsymbol{w}_{m'}} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* - \|\boldsymbol{w}_{m'}\|_2^2 \mathbf{I}_L) \boldsymbol{S}^* \Big\| + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \|\boldsymbol{v}_{m'}\|_2 \|\boldsymbol{S} \boldsymbol{C}_{\boldsymbol{w}_m} \boldsymbol{C}_{\boldsymbol{w}_{m'}}^* \boldsymbol{S}^*\| \right)$$

$$\leq \|\boldsymbol{v}\|_{2,\infty} (M \bar{\rho}_w + \rho_w) + \|\boldsymbol{v}\|_{2,1} \rho_w \leq \|\boldsymbol{v}\|_{2,\infty} M(\bar{\rho}_w + 2\rho_w).$$

Therefore, the radii of Δ_2 with respect to the Frobenius and spectral norms are upper bounded respectively by

$$d_{\mathrm{S}}(\Delta_2) \leq M(\bar{\rho}_w + 2\rho_w),$$
$$d_{\mathrm{F}}(\Delta_2) \leq C \sqrt{K} (M \bar{\rho}_w + \sqrt{M} \rho_w).$$

Moreover, by Lemma E.1, the γ_2 functional is bounded by

$$\gamma_2(\Delta_2, \|\cdot\|) \leq C_1 M(\bar{\rho}_w + 2\rho_w) \int_0^\infty \sqrt{\log N(B_2^{MD}, \|\cdot\|_{2,\infty}, t)} dt$$

$$\leq C_2 M \sqrt{D} (\bar{\rho}_w + 2\rho_w) \sqrt{\log D} \log(MD).$$

Consequently, Theorem 5.2 yields that

$$\|\boldsymbol{\Phi}^* (\boldsymbol{Y}_{\mathrm{n}}^* \boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda}) \boldsymbol{\Phi}\| \leq C(\beta) M \sqrt{KD} (\bar{\rho}_w + 2\rho_w) \log D \log^2(MD)$$

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

D.10. Proof of Lemma 5.12. We modify the proof of Lemma 5.11 in Appendix D.9 as follows.

First note that

$$\frac{\|\boldsymbol{\Phi}^*(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda})\boldsymbol{\Phi}\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2} = \sup_{\boldsymbol{v}\in B_2^{MD}} \frac{|\phi^*\boldsymbol{W}_{\boldsymbol{v}}^*\boldsymbol{Q}_w^*\boldsymbol{Q}_w\boldsymbol{W}_{\boldsymbol{u}}\phi|}{\|\boldsymbol{u}\|_2}.$$

We only need to replace $\check{\Delta}_{\mathrm{R}}$ by the following singleton set:

$$\check{\Delta}_{\mathrm{R},\boldsymbol{u}} = \{(\boldsymbol{Y}_{\mathrm{n}}^*\boldsymbol{Y}_{\mathrm{n}} - \boldsymbol{\Lambda})\boldsymbol{W}_{\boldsymbol{u}}\}.$$

Indeed, the radii of $\check{\Delta}_{\mathrm{R},\boldsymbol{u}}$ and $\check{\Delta}_{\mathrm{R}}$ are the same for both the Frobenius and spectral norms. However, the γ_2 functional of $\check{\Delta}_{\mathrm{R},\boldsymbol{u}}$ is 0. The assertion follows by applying the modified estimates to Theorem 5.2.

Appendix E. Entropy estimates.

Lemma E.1.

(65) $$\int_0^\infty \sqrt{\log N(B_2^{MD}, \|\cdot\|_{2,\infty}, t)}\, dt \le C\sqrt{D}\sqrt{\log D}\log(MD).$$

Proof of Lemma E.1. Let us recall that the $(2,\infty)$-block norm of $\boldsymbol{v} \in \mathbb{C}^{MD}$ is defined by

$$\|\boldsymbol{v}\|_{2,\infty} = \max_{m\in[M]} \|\boldsymbol{v}_m\|_2,$$

where $\boldsymbol{v}_k \in \mathbb{C}^D$ for $k = 1,\ldots,M$ denotes the blocks of $\boldsymbol{v}$ such that $\boldsymbol{v} = [\boldsymbol{v}_1^\top,\ldots,\boldsymbol{v}_M^\top]$.

Since

$$\|\boldsymbol{v}\|_{2,\infty} \le \|\boldsymbol{v}\|_2,$$

the interval in the integral in (65) can be restricted to the unit interval $[0,1]$.

Indeed, the $(2,\infty)$-block norm of $\boldsymbol{v}$ is rewritten as

$$\|\boldsymbol{v}\|_{2,\infty} = \max_{m\in[M]} \max_{\boldsymbol{\zeta}\in B_2^D} \langle \boldsymbol{v}_m, \boldsymbol{\zeta}\rangle.$$

To compute an estimate of the entropy integral in (65), we adopt the strategy [20] that estimates a unit ball using a polytope. The original strategy [20] was developed for the restricted isometry property analysis for low-rank tensors and applied to the tensor nuclear norm. The same strategy applies to the block norm in this section too.

Lemma E.2. *There exist* $\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N \in \mathbb{S}^{D-1}$ *such that*

$$B_2^D \subset 2\,\mathrm{conv}\{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\}$$

and $\log N \le (2D+1)\log(4D+3)$.

Proof of Lemma E.2. Let $\{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\} \subset \mathbb{S}^{D-1}$ be an ϵ-net of $\mathbb{S}^{D-1}$. Then by the standard volume argument, we have $N \le (1 + 2/\epsilon)^{2D}$. Furthermore, it follows that

$$\mathbb{S}^{D-1} \subset \frac{1}{1-\epsilon}\cdot\mathrm{absconv}\{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\}.$$

Indeed, for any $\boldsymbol{w} \in \mathbb{S}^{D-1}$, we construct a sequence $(\widehat{\boldsymbol{w}}_k, \alpha_k)_{k\in\mathbb{N}} \subset \{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\}\times\mathbb{C}$ as follows. Let $\alpha_1 = 1$ and $\widehat{\boldsymbol{w}}_1$ be the closest vector to $\boldsymbol{w}$ in $\{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\}$. If $\boldsymbol{w} = \alpha_1\widehat{\boldsymbol{w}}_1$, then $\alpha_k = 0$

and $\widehat{\boldsymbol{w}}_k = \boldsymbol{0}$ for all $k \geq 2$. Otherwise, let $\alpha_2 = \|\boldsymbol{w} - \alpha_1\widehat{\boldsymbol{w}}_1\|_2$ and $\widehat{\boldsymbol{w}}_2$ be the closest vector to $\alpha_2^{-1}(\boldsymbol{w} - \alpha_1\widehat{\boldsymbol{w}}_1)$ in $\{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N\}$. If $\boldsymbol{w} = \alpha_1\widehat{\boldsymbol{w}}_1 + \alpha_2\widehat{\boldsymbol{w}}_2$, then $\alpha_k = 0$ and $\widehat{\boldsymbol{w}}_k = \boldsymbol{0}$ for all $k \geq 3$. Otherwise, let $\alpha_3 = \|\boldsymbol{w} - \alpha_1\widehat{\boldsymbol{w}}_1 - \alpha_2\widehat{\boldsymbol{w}}_2\|_2$ and $\widehat{\boldsymbol{w}}_3$ be the closest vector to $\alpha_3^{-1}(\boldsymbol{w} - \alpha_1\widehat{\boldsymbol{w}}_2 - \alpha_1\widehat{\boldsymbol{w}}_2)$ in $\{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N\}$. By continuing in this way, we have

$$\boldsymbol{w} = \sum_{k\in\mathbb{N}} \alpha_k\widehat{\boldsymbol{w}}_k,$$

where $|\alpha_k| \leq \epsilon^{k-1}$ and $\|\widehat{\boldsymbol{w}}_k\|_2 = 1$ for all $k \in \mathbb{N}$. Therefore,

$$\sum_{k\in\mathbb{N}} |\alpha_k| \leq \frac{1}{1-\epsilon}$$

and the assertion follows. By including $\pm\boldsymbol{\zeta}_k$ instead of $\boldsymbol{\zeta}_k$, we can replace the absolute convex hull by convex hull and the cardinality increases only by factor 2. Choosing $\epsilon = 1/(2D+1)$ completes the proof. ∎

By Lemma E.2, we approximate the $(2, \infty)$-block norm of $\boldsymbol{v}$ as a polytope norm as follows:

(66)
$$\begin{aligned}
\|\boldsymbol{v}\|_{2,\infty} &= \max_{\boldsymbol{\zeta}\in\mathbb{S}^{D-1}} \max_{m\in[M]} \langle\boldsymbol{v}_m, \boldsymbol{\zeta}\rangle \\
&\leq 2 \max_{\boldsymbol{\zeta}\in\text{conv}\{\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_N\}} \max_{m\in[M]} \langle\boldsymbol{v}_m, \boldsymbol{\zeta}\rangle \\
&= 2 \max_{n\in[N]} \max_{m\in[M]} |\langle\boldsymbol{v}_m, \boldsymbol{\zeta}_n\rangle| \\
&=: 2\,|||\boldsymbol{v}|||,
\end{aligned}$$

where $\log N \leq (D+1)\log(2D+3)$.

Define
$$S_{\zeta_n} = \left\{\boldsymbol{v} = [\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_M^\top]^\top \ \middle| \ \max_{m\in[M]} |\langle\boldsymbol{v}_m, \boldsymbol{\zeta}_n\rangle| \leq 1\right\}.$$

Then its polar set is given by

$$S_{\zeta_n}^\circ = \{\boldsymbol{z} \otimes \boldsymbol{\zeta}_n | \boldsymbol{z} \in B_1^M\} = \text{conv}\{\boldsymbol{e}_d \otimes \boldsymbol{\zeta}_n | d \in [D]\}.$$

Note that the unit ball with respect to $|||\cdot|||$ is given as $\bigcap_{n\in[N]} S_{\zeta_n}$. To compute the unit ball with respect to the dual norm, we will use a well-known polar duality result in the following lemma. Note that S_{ζ_n} is not bounded. As the lemma is typically stated for bounded sets, we provide the proof for completeness, verifying that boundedness is not a crucial assumption.

Lemma E.3. *Let A and B be convex sets. The polar set of the intersection of A and B is given by*

$$(A \cap B)^\circ = \text{conv}(A^\circ \cup B^\circ).$$

Proof of Lemma E.3. We first show

(67)
$$(\text{conv}(A^\circ \cup B^\circ))^\circ \subset A \cap B.$$

Suppose that $x \notin A \cap B$. Since A and B are convex, we have $A = A^{\circ\circ}$ and $B = B^{\circ\circ}$. Without loss of generality, we may assume that $x \notin A^{\circ\circ}$. Then there exists $w \in A^{\circ}$ such that

$$(68) \qquad |\langle x, w \rangle| > 1.$$

Since $A^{\circ} \subset A^{\circ} \cup B^{\circ} \subset \mathrm{conv}(A^{\circ} \cup B^{\circ})$, w also satisfies

$$(69) \qquad w \in \mathrm{conv}(A^{\circ} \cup B^{\circ}).$$

The existence of w satisfying both (68) and (69) implies that

$$x \notin (\mathrm{conv}(A^{\circ} \cup B^{\circ}))^{\circ}.$$

Then (67) follows by contraposition.

Next, we show the other inclusion, which is

$$(70) \qquad A \cap B \subset (\mathrm{conv}(A^{\circ} \cup B^{\circ}))^{\circ}.$$

Suppose that $x \in A \cap B = A^{\circ\circ} \cap B^{\circ\circ}$. Then for all $w_A \in A^{\circ}$, $w_B \in B^{\circ}$, and $t \in [0,1]$, it follows that

$$|\langle x, tw_A + (1-t)w_B \rangle| \leq t|\langle w, w_A \rangle| + (1-t)|\langle w, w_B \rangle| \leq 1.$$

Therefore

$$x \in \mathrm{conv}(A^{\circ} \cup B^{\circ})^{\circ}.$$

We have shown that

$$A \cap B = \mathrm{conv}(A^{\circ} \cup B^{\circ})^{\circ}.$$

The assertion follows from the definition of polar sets. ■

By the polar duality in Lemma E.3, the unit ball with respect to the dual of $|||\cdot|||$ is given as

$$\mathrm{conv}\left(\bigcup_{n \in [N]} S_{\zeta_n}^{\circ}\right) = \mathrm{conv}\left(\bigcup_{n \in [N]} \mathrm{conv}\{\boldsymbol{e}_d \otimes \boldsymbol{\zeta}_n | d \in [D]\}\right)$$
$$= \mathrm{conv}\{\boldsymbol{e}_d \otimes \boldsymbol{\zeta}_n | n \in [N], \ d \in [D]\}.$$

Collecting the above estimates, we bound the log entropy number in (65) as follows:

$$\log N(B_2^{MD}, \|\cdot\|_{2\infty}, t) \lesssim \log N(B_2^{MD}, |||\cdot|||, t/2)$$

$$\lesssim \log N\left(\mathrm{conv}\bigcup_{n \in [N]} S_{\zeta_n}^{\circ}, \|\cdot\|_2, t/2\right)$$

$$\lesssim \log N\left(\mathrm{conv}\{\boldsymbol{e}_d \otimes \boldsymbol{\zeta}_n | n \in [N], \ d \in [D]\}, \|\cdot\|_2, t/2\right),$$

where the first inequality follows from (66) and the second inequality holds by the entropy duality by Artstein, Milman, and Szarek [2].

Next we define a linear map $Q : \ell_1^{ND} \to \ell_2^{MD}$ so that the standard basis vectors in ℓ_1^{ND} are mapped to distinct elements in $\{e_d \otimes \zeta_n \mid n \in [N], \ d \in [D]\}$. In this construction, we only care about the one-to-one correspondence and we do not care how we enumerate the elements of $\{e_d \otimes \zeta_n \mid n \in [N], \ d \in [D]\}$. Although Q is not uniquely determined and there is ambiguity up to a permutation in ℓ_1^{ND}, every map Q constructed as above satisfies that $\|Q : \ell_1^{ND} \to \ell_2^{MD}\| = 1$. Fix Q and we get

$$\int_0^1 \sqrt{\log N\big(B_2^{MD}, \|\cdot\|_{2,\infty}, t\big)} dt \lesssim \int_0^1 \sqrt{\log N\Big(Q(B_1^{ND}), \|\cdot\|_2, t/2\Big)} dt$$
$$\lesssim \sqrt{\log(ND)} \log(MD) \lesssim \sqrt{D}\sqrt{\log D} \log(MD),$$

where the second inequality follows from Corollary A.5. ∎

REFERENCES

[1] A. AHMED, B. RECHT, AND J. ROMBERG, *Blind deconvolution using convex programming*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1711–1732.

[2] S. ARTSTEIN, V. MILMAN, AND S. J. SZAREK, *Duality of metric entropy*, Ann. of Math., 159 (2004), pp. 1313–1328.

[3] L. BALZANO AND R. NOWAK, *Blind calibration of sensor networks*, in Proceedings of the 6th International Conference on Information Processing in Sensor Networks, ACM, 2007, pp. 79–88.

[4] S.-H. BYUN, C. M. VERLINDEN, AND K. G. SABRA, *Blind deconvolution of shipping sources in an ocean waveguide*, J. Acoust. Soc. Amer., 141 (2017), pp. 797–807.

[5] B. CARL, *Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces*, in Ann. Inst. Fourier, 35 (1985), pp. 79–118.

[6] S. CHOUDHARY AND U. MITRA, *Fundamental Limits of Blind Deconvolution Part* I*: Ambiguity Kernel*, arxiv:1411.3810, 2014.

[7] S. CHOUDHARY AND U. MITRA, *Sparse blind deconvolution: What cannot be done*, in Proceedings of the International Symposium on Information Theory, IEEE, 2014, pp. 3002–3006.

[8] A. CURTIS, P. GERSTOFT, H. SATO, R. SNIEDER, AND K. WAPENAAR, *Seismic interferometry: Turning noise into signal*, Leading Edge, 25 (2006), pp. 1082–1092.

[9] K. R. DAVIDSON AND S. J. SZAREK, *Local operator theory, random matrices and Banach spaces*, Handb. Geom. Banach Spaces, 1 (2001), pp. 317–366.

[10] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.

[11] D. G. FEINGOLD AND R. S. VARGA, *Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem*, Pacific J. Math, 12 (1962), pp. 1241–1250.

[12] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Vol. 1, Birkhäuser, Basel, 2013.

[13] J. GARNIER AND G. PAPANICOLAOU, *Passive synthetic aperture imaging*, SIAM J. Imaging Sci., 8 (2015), pp. 2683–2705.

[14] G. B. GIANNAKIS AND R. W. HEATH, *Blind indentification of multichannel FIR blurs and perfect image restoration*, IEEE Trans. Image Process., 9 (2000), pp. 1877–1896.

[15] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2012.

[16] M. A. GRISWOLD, P. M. JAKOB, R. M. HEIDEMANN, M. NITTKA, V. JELLUS, J. WANG, B. KIEFER, AND A. HAASE, *Generalized autocalibrating partially parallel acquisitions (GRAPPA)*, Magnetic Resonance Medicine, 47 (2002), pp. 1202–1210.

[17] J. H. GUNTHER AND A. L. SWINDLEHURST, *On the use of kernel structure for blind equalization*, IEEE Trans. Signal. Process., 48 (2000), pp. 799–809.

[18] M. I. GURELLI AND C. L. NIKIAS, *EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals*, IEEE Trans. Signal Process., 43 (1995), pp. 134–149.

[19] G. HARIKUMAR AND Y. BRESLER, *Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms*, IEEE Trans. Image Process., 8 (1999), pp. 202–219.

[20] M. JUNGE AND K. LEE, *Generalized Notions of Sparsity and Restricted Isometry Property. Part I: A Unified Framework*, preprint, arXiv:1706.09410, 2017.

[21] M. JUNGE AND Q. ZENG, *Noncommutative Bennett and Rosenthal inequalities*, Ann. Probab., 41 (2013), pp. 4287–4316.

[22] K. F. KAARESEN AND T. TAXT, *Multichannel blind deconvolution of seismic signals*, Geophysics, 63 (1998), pp. 2093–2107.

[23] M. KECH AND F. KRAHMER, *Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems*, SIAM J. Appl. Algebra Geom., 1 (2017), pp. 20–37.

[24] V. KOLTCHINSKII, K. LOUNICI, AND A. B. TSYBAKOV, *Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion*, Ann. Statist., 39 (2011), pp. 2302–2329.

[25] F. KRAHMER, S. MENDELSON, AND H. RAUHUT, *Suprema of chaos processes and the restricted isometry property*, Comm. Pure Appl. Math., 67 (2014), pp. 1877–1904.

[26] B. LAURENT AND P. MASSART, *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist., 28 (2000), pp. 1302–1338.

[27] K. LEE AND M. JUNGE, *RIP-Like Properties in Subsampled Blind Deconvolution*, preprint, arXiv:1511.06146, 2015.

[28] K. LEE, F. KRAHMER, AND J. ROMBERG, *An eigen approach to stable multichannel blind deconvolution under an FIR subspace model*, in Proceedings of the International Conference on Sampling Theory and Applications, 2017, pp. 386–390, https://doi.org/10.1109/SAMPTA.2017.8024427.

[29] K. LEE, Y. LI, M. JUNGE, AND Y. BRESLER, *Blind recovery of sparse signals from subsampled convolution*, IEEE Trans. Inform. Theory, 63 (2017), pp. 802–821.

[30] X. LI, S. LING, T. STROHMER, AND K. WEI, *Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization*, preprint, arXiv:1606.04933, 2016.

[31] Y. LI, K. LEE, AND Y. BRESLER, *Identifiability in blind deconvolution with subspace or sparsity constraints*, IEEE Trans. Inform. Theory, 62 (2016), pp. 4266–4275.

[32] Y. LI, K. LEE, AND Y. BRESLER, *Optimal sample complexity for blind gain and phase calibration*, IEEE Trans. Signal Process., 64 (2016), pp. 5549–5556.

[33] Y. LI, K. LEE, AND Y. BRESLER, *Identifiability and stability in blind deconvolution under minimal assumptions*, IEEE Trans. Inform. Theory, 63 (2017), pp. 4619–4633.

[34] S. LING AND T. STROHMER, *Self-calibration and bilinear inverse problems via linear least squares*, SIAM J. Imaging Sci., 11 (2018), pp. 252–292.

[35] H. LIU, G. XU, L. TONG, AND T. KAILATH, *Recent developments in blind channel equalization: From cyclostationnarity to subspaces*, Signal Process., 50 (1996), pp. 83–99.

[36] R. L. MORRISON, M. N. DO, AND D. C. MUNSON, *MCA: A multichannel approach to sar autofocus*, IEEE Trans. Image Process., 18 (2009), pp. 840–853.

[37] E. MOULINES, P. DUHAMEL, J.-F. CARDOSO, AND S. MAYRARGUE, *Subspace methods for the blind identification of multichannel FIR filters*, IEEE Trans. Signal Process., 43 (1995), pp. 516–525.

[38] H. Q. NGUYEN, S. LIU, AND M. N. DO, *Subspace methods for computational relighting*, in Proceedings of Computational Imaging XI, 2013, 865703.

[39] S. O'ROURKE, V. VU, AND K. WANG, *Random Perturbation of Low Rank Matrices: Improving Classical Bounds*, preprint, arXiv:1311.2657, 2013.

[40] M. RUDELSON AND R. VERSHYNIN, *Hanson-Wright inequality and sub-Gaussian concentration*, Electron. Commun. Probab., 18 (2013), pp. 1–9.

[41] K. G. SABRA AND D. R. DOWLING, *Blind deconvolution in ocean waveguides using artificial time reversal*, J. Acoust. Soc. America, 116 (2004), pp. 262–271.

[42] K. G. SABRA, H.-C. SONG, AND D. R. DOWLING, *Ray-based blind deconvolution in ocean sound channels*, J. Acoust. Soc. America, 127 (2010), pp. EL42–EL47.

[43] N. TIAN, S.-H. BYUN, K. SABRA, AND J. ROMBERG, *Multichannel myopic deconvolution in underwater acoustic channels via low-rank recovery*, J. Acoust. Soc. America, 141 (2017), pp. 3337–3348.

[44] L. Tong and S. Perreau, *Multichannel blind identification: From subspace to maximum likelihood methods*, Proc. IEEE, 86 (1998), pp. 1951–1968.

[45] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math., 12 (2012), pp. 389–434.

[46] V. Vu, *Singular vectors under random perturbation*, Random Structures Algorithms, 39 (2011), pp. 526–538.

[47] G. Xu, H. Liu, L. Tong, and T. Kailath, *A least-squares approach to blind channel identification*, IEEE Trans. Signal Process., 43 (1995), pp. 2982–2993.

[48] J. Zhou and M. Do, *Multidimensional multichannel FIR deconvolution using Gröbner bases*, IEEE Trans. Image Process., 15 (2006), pp. 2998–3007.

[49] X. Zhu, F. Šroubek, and P. Milanfar, *Deconvolving PSFs for a better motion deblurring using multiple images*, in Proceedings of Computer Vision–ECCV, 2012, pp. 636–647.