



HAL
open science

A Random Matrix Approach to Low-Multilinear-Rank Tensor Approximation

Hugo Lebeau, Florent Chatelain, Romain Couillet, Romain Couillet

► **To cite this version:**

Hugo Lebeau, Florent Chatelain, Romain Couillet, Romain Couillet. A Random Matrix Approach to Low-Multilinear-Rank Tensor Approximation. 2024. hal-04673321

HAL Id: hal-04673321

<https://hal.science/hal-04673321v1>

Preprint submitted on 20 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A Random Matrix Approach to Low-Multilinear-Rank Tensor Approximation

Hugo Lebeau

HUGO.LEBEAU@UNIV-GRENOBLE-ALPES.FR

*Université Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG
Grenoble, 38000, France*

Florent Chatelain

FLORENT.CHATELAIN@GRENOBLE-INP.FR

*Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab
Grenoble, 38000, France*

Romain Couillet

ROMAIN.COUILLET@UNIV-GRENOBLE-ALPES.FR

*Université Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG
Grenoble, 38000, France*

Abstract

This work presents a comprehensive understanding of the estimation of a planted low-rank signal from a general spiked tensor model near the computational threshold. Relying on standard tools from the theory of large random matrices, we characterize the large-dimensional spectral behavior of the unfoldings of the data tensor and exhibit relevant signal-to-noise ratios governing the detectability of the principal directions of the signal. These results allow to accurately predict the reconstruction performance of truncated multilinear SVD (MLSVD) in the non-trivial regime. This is particularly important since it serves as an initialization of the higher-order orthogonal iteration (HOOI) scheme, whose convergence to the best low-multilinear-rank approximation depends entirely on its initialization. We give a sufficient condition for the convergence of HOOI and show that the number of iterations before convergence tends to 1 in the large-dimensional limit.

Keywords: random tensors, random matrix theory, spiked tensor model, truncated MLSVD, higher-order orthogonal iteration

1 Introduction

Information retrieval from large amounts of data has become a common task of signal processing and machine learning in the past decades. Often, these data have several modes as they may come from various sources, modalities, domains, and so on. Tensors (multi-way arrays) are therefore a natural representation for such datasets — they appear in multiple areas such as brain imaging (Zhou et al., 2013), neurophysiological measurements (Seely et al., 2016), community detection (Anandkumar et al., 2013), compression of hyperspectral images (Li and Li, 2010), spatio-temporal gene expression (Liu et al., 2022), recommender systems (Karatzoglou et al., 2010; Rendle and Schmidt-Thieme, 2010; Frolov and Oseledets, 2017) and topic modeling (Anandkumar et al., 2014). Indeed, tensors as multi-way arrays provide a more detailed representation of data than mere matrices (two-way arrays) as they convey a structural information. For instance, the modes of a data tensor can represent *pixel* \times *pixel* \times *wavelength* \times *sample* in hyperspectral imaging (Zhang et al., 2013; Kanatsoulis et al., 2018), *time* \times *spatial scale* \times *electrode* in the EEG analysis by Acar et al.

(2007) or *neuron* \times *time* \times *stimuli* in the study of the visual cortex by Rabinowitz et al. (2015).

In an information retrieval context, it is common to make use of tensor decompositions in order to estimate a sought signal. In their fMRI study, Hunyadi et al. (2017) perform a blind source separation via a joint tensor decomposition on a *channel* \times *time* \times *patient* array, whereas Williams et al. (2018) use a low-rank tensor approximation on a *neuron* \times *time* \times *trial* array as a dimensionality reduction technique to study neural dynamics. In fact, supposing that the signal has a low-rank structure is a natural sparsity assumption (Kadmon and Ganguli, 2018; Anandkumar et al., 2014), and low-rank tensor approximations are key tools to extract information from multi-way data.

In the present work, we propose a random matrix analysis of a general *low-rank information + noise* tensor model and precisely quantify the amount of information which can be recovered with a low-rank tensor approximation depending on the signal-to-noise ratio (SNR). For a general introduction to tensors, we refer the reader to Comon (2014, 2009); Landsberg (2011); Hackbusch (2012) and, for an emphasis on statistical learning applications, Sun et al. (2021). In the remainder of the introduction, the main concepts and challenges behind low-rank tensor estimation are presented in Section 1.1. Then, Section 1.2 introduces some important related works. Our main results are finally summarized in Section 1.3. All the notations are properly defined in Section 2.

1.1 Low-Rank Tensor Estimation

What is meant by a low-rank approximation of a tensor? And how is the *rank* of a tensor actually defined? Let us start with a familiar matrix case: a matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ is a two-way array (or order-2 tensor). A singular value decomposition (SVD) allows us to write \mathbf{M} in a compact way as the sum of R rank-1 terms, $\mathbf{M} = \sum_{i=1}^R \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ where \mathbf{U} , \mathbf{V} are respectively $n_1 \times R$ and $n_2 \times R$ matrices with *orthonormal* columns and $\mathbf{\Sigma}$ is the $R \times R$ *diagonal* matrix of singular values. The rank of \mathbf{M} is here the minimal number of rank-1 terms in which it can be exactly decomposed. Extending this notion to tensors therefore seems straightforward: a tensor¹ $\mathcal{J} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has rank R if it is the minimal number of rank-1 terms in which it can be exactly decomposed, $\mathcal{J} = \sum_{i=1}^R \sigma_i \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i$. What we have just described is the canonical polyadic decomposition (CPD) of \mathcal{J} , it dates back to Hitchcock (1927) and is unique under very mild conditions (Kolda and Bader, 2009). However, we have lost an important property in this process: the unit vectors \mathbf{a}_i (resp. \mathbf{b}_i , \mathbf{c}_i) are, in general, no longer orthonormal. Conversely, retaining the orthonormality property inevitably results in the loss of the diagonality property, $\mathcal{J} = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sum_{k=1}^{r_3} \mathcal{G}_{i,j,k} \mathbf{u}_i \otimes \mathbf{v}_j \otimes \mathbf{w}_k$. This latter decomposition is called a Tucker decomposition and dates back to Tucker (1966). In fact, the best way to represent \mathcal{J} with a Tucker decomposition is to choose the \mathbf{u}_i (resp. \mathbf{v}_i , \mathbf{w}_i) as the left singular vectors of the unfolding of \mathcal{J} along mode 1 (resp. 2, 3)². This is called the multilinear SVD (MLSVD, De Lathauwer et al., 2000b) and gives rise to a new definition of rank: the *multilinear-rank* (r_1, r_2, r_3) . Note that, in the matrix case, $r_1 = r_2 = R$ since both the diagonality and orthonormality properties are verified. However, r_1, r_2, r_3 are, in general, not equal in the tensor case, but $\max(r_1, r_2, r_3) \leq R \leq \min(r_1 r_2, r_2 r_3, r_1 r_3)$. See,

1. It is chosen of order 3 for simplicity of exposure.

2. This is properly defined in Section 2.2.

e.g., Sidiropoulos et al. (2017) for details. Other relevant references for the reader interested in tensor decompositions are Kolda and Bader (2009); Cichocki et al. (2015); Rabanser et al. (2017).

Given an order- d tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ of possibly very high rank, we are interested in finding a low-rank approximation, i.e., an $n_1 \times \dots \times n_d$ tensor \mathcal{X} which minimizes the distance $\|\mathcal{T} - \mathcal{X}\|_F$ on a set of low-rank tensors. Yet, the problem of the best rank- R approximation of a tensor is ill-posed as soon as $R > 1$ because the set of rank- R tensors is not closed (Kolda and Bader, 2009). Instead, we shall consider the best low-*multilinear*-rank problem, which is always well-posed,

$$\min_{\text{rank}(\mathcal{X}) \leq (r_1, \dots, r_d)} \|\mathcal{T} - \mathcal{X}\|_F^2. \quad (1)$$

It is well known in the matrix case that the best rank- R approximation can be easily computed by truncating the SVD to its R most energetic terms (Eckart and Young, 1936; Mirsky, 1960). Could this also be true for the MLSVD? Unfortunately, counter-examples exist (Kolda, 2003), showing that there is no tensor equivalent of the Eckart-Young-Mirsky theorem. Worse still, Problem (1) is in fact NP-hard (Hillar and Lim, 2013). Nevertheless, despite not being the best low-multilinear-rank approximation, the truncated MLSVD $\hat{\mathcal{T}}$ remains a very good “first guess” as it verifies $\|\mathcal{T} - \mathcal{T}_*\|_F \leq \|\mathcal{T} - \hat{\mathcal{T}}\|_F \leq \sqrt{d} \|\mathcal{T} - \mathcal{T}_*\|_F$ where \mathcal{T}_* denotes a solution to Problem (1) and d is the order of the tensor (Grasedyck et al., 2013; Hackbusch, 2012). It is a cheap (it consists only in d standard matrix SVDs) and quasi-optimal low-multilinear-rank approximation of \mathcal{T} . Moreover, it is often used as an initialization of numerical methods which estimate a solution to Problem (1), among which the most common is the higher-order orthogonal iteration (HOOI) algorithm (Kroonenberg and de Leeuw, 1980; Kapteyn et al., 1986; De Lathauwer et al., 2000a).

Another motivation for the analysis of the low-multilinear-rank approximation problem is that it has also a practical interest for the numerical computation of the canonical polyadic decomposition (CPD). Indeed, when dealing with large tensors, it is computationally more efficient to first compress the tensor with a low-multilinear-rank approximation and then compute the CPD on the smaller core tensor rather than computing the CPD of the large tensor directly (Bro and Andersson, 1998). This is done, e.g., by the `cpd` function of the MATLAB toolbox `Tensorlab` (Vervliet et al., 2016).

1.2 Related Work

Multilinear SVD (MLSVD) has a wide range of applications and is often used to extract relevant information from multi-way arrays. For instance, it has been used in human motion recognition (Vasilescu, 2002), face recognition (Vasilescu and Terzopoulos, 2003), handwritten digit classification (Savas and Eldén, 2007) but also genomics (Omberg et al., 2007, 2009; Muralidhara et al., 2011) and syndromic surveillance (Fanaee-T and Gama, 2015).

The analysis of *spiked* tensor models — i.e., low-rank perturbations of large random tensors — has started with the introduction by Montanari and Richard (2014) of the rank-1 symmetric spiked tensor model, $\mathcal{T} = \beta \mathbf{x}^{\otimes d} + \mathcal{N}$ with $\|\mathbf{x}\| = 1$, \mathcal{N} Gaussian noise and β a parameter controlling the signal-to-noise ratio (SNR). They show that estimation of \mathbf{x} from \mathcal{T} is *theoretically* possible as soon as β is above a certain threshold β_c behaving like $\sqrt{d \log d}$, which is reminiscent of the now well-known spiked *matrix* model where signal reconstruction is only possible above a critical threshold (Péché, 2006) — a phenomenon called the BBP

phase transition (Baik et al., 2005). The behavior of singular values and singular vectors of spiked matrix models is comprehensively studied by Benaych-Georges and Nadakuditi (2012). Contrary to the matrix case however, Montanari and Richard (2014) make the disturbing observation that none of the polynomial-time estimation algorithms among tensor unfolding, power iteration and approximate message passing (AMP) succeed unless β diverges as the dimensions of the tensor grow large. The results of Hopkins et al. (2015, 2017) suggest that no polynomial-time algorithm can succeed unless $\beta \gtrsim N^{\frac{d-2}{4}}$, where N scales as the dimensions of the data tensor. While Perry et al. (2020) show that the information-theoretic threshold is of order 1, this indicates the existence of a *computational-to-statistical gap* in spiked tensor estimation, as in a myriad of other problems (Bandeira et al., 2018; Zdeborová and Krzakala, 2016).

The landscape of the rank-1 symmetric spiked tensor model is studied by Ben Arous et al. (2019), who show that the number of local optima to Problem (1) grows exponentially with the size of the tensor, but all lie close to a subspace orthogonal to the sought solution, except for one if β exceeds a critical threshold β_c . Completing this analysis, Jagannath et al. (2020) show³ the existence of two close but different thresholds $\beta_s < \beta_c$ such that the solution aligned with the underlying signal is a local minimum of Problem (1) as soon as $\beta > \beta_s$ but becomes a *global* one only if $\beta > \beta_c$. Relying on the Kac-Rice method, Ros et al. (2019) thoroughly study such high-dimensional landscapes and classify the different behaviors and phase transitions which can occur.

So far, we have only referred to works dealing with the rank-1 symmetric case, but there are also some studies on higher-(low-)rank spiked models. Chevreuil and Loubaton (2018) give a sufficient (but not necessary) condition for the *non-detectability* of a rank- R asymmetric signal perturbed by an additive Gaussian noise. Chen et al. (2021) also discuss signal detectability in the rank- R symmetric case. The statistical inference of finite rank tensors is studied by Chen et al. (2022) who identify the limit free energy of the model in terms of a variational formula. Zhang and Xia (2018) consider a general *low-multilinear-rank signal* \mathcal{P} + *Gaussian noise* \mathcal{N} model and bring to light the same statistical-to-computational gap: if $\|\mathcal{P}\|_F$ is above a statistical threshold of order 1 then Problem (1) has a solution which is aligned with the signal but is computationally intractable unless $\|\mathcal{P}\|_F$ is above a computational threshold of order $N^{\frac{d-2}{4}}$. In this strong SNR regime, the higher-order orthogonal iteration (HOOI) algorithm (De Lathauwer et al., 2000a) is minimax-optimal. In fact, it is also proved by Ben Arous et al. (2020) that, with Langevin dynamics and gradient descent, the algorithmic threshold behaves like N^α with $\alpha > 0$. Other algorithmic thresholds have been shown as well for semi-definite and spectral relaxations of the maximum likelihood problem (Hopkins et al., 2015, 2016; Kim et al., 2017). AMP and tensor power iteration algorithms achieve $N^{\frac{d-1}{2}}$ (Lesieur et al., 2017; Huang et al., 2022) while tensor unfolding methods (truncated MLSVD and HOOI algorithm) achieve $N^{\frac{d-2}{4}}$ as already conjectured by Montanari and Richard (2014) and later proven by Hopkins et al. (2015); Ben Arous et al. (2023) in the rank-1 case. The convergence of the HOOI algorithm towards a local maximum for a *sufficiently close* initialization is proven by Xu (2018) and Feldman and Donoho

3. The setting considered by Jagannath et al. (2020) is more general than the one of Ben Arous et al. (2019) because the noise in their model is not necessarily symmetric but the perturbation is still a rank-1 symmetric tensor.

(2023) show that it achieves exact recovery of a rank-1 perturbation in the large N regime when it is initialized with the dominant singular vectors of the unfoldings.

Recently, a new approach relying on tools from random matrix theory has broadened the understanding of spiked tensor models. In particular, Goulart et al. (2022) study the rank-1 symmetric case and are able to recover explicitly the same β_s threshold as Jagannath et al. (2020) as well as to precisely quantify the alignment between a solution to Problem (1) and the signal. A similar analysis is carried out by Seddik et al. (2022) for the more general asymmetric case, relying solely on classical techniques from random matrix theory. Such tools show promise for the theoretical understanding of learning from tensor data (Seddik et al., 2023). In particular, the results of Ben Arous et al. (2023) and Feldman (2023) on *long* random matrices, similar to those considered in this work, provide instructive insight into the recovery performance of tensor unfolding methods.

1.3 Summary of Contributions

In low-rank tensor approximation, tensor unfolding methods achieve the best known performance among polynomial-time algorithms. Motivated by several works suggesting that such method could actually reach the computational threshold (Hopkins et al., 2015, 2017; Zhang and Xia, 2018; Wein et al., 2019), we propose a thorough random matrix analysis of the low-multilinear-rank tensor approximation problem.

Consider the general spiked tensor model,

$$\mathcal{T} = \mathcal{P} + \frac{1}{\sqrt{N}}\mathcal{N} \in \mathbb{R}^{n_1 \times \dots \times n_d}, \quad \mathcal{N}_{i_1, \dots, i_d} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad (2)$$

where \mathcal{N} is an additive Gaussian noise, N is a parameter controlling the size of the tensor such that $n_\ell/N = \Theta(1)$ ⁴ for all $\ell \in \{1, \dots, d\}$ (for instance, $N = n_1$ or $N = \sum_{\ell=1}^d n_\ell$) and \mathcal{P} is a low-multilinear-rank deterministic tensor, i.e., which can be decomposed as

$$\mathcal{P} = \sum_{q_1=1}^{r_1} \dots \sum_{q_d=1}^{r_d} \mathcal{H}_{q_1, \dots, q_d} [\mathbf{x}_{q_1}^{(1)} \otimes \dots \otimes \mathbf{x}_{q_d}^{(d)}] \stackrel{\text{def}}{=} \left[\mathcal{H}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)} \right], \quad (3)$$

with $\mathcal{H} \in \mathbb{R}^{r_1 \times \dots \times r_d}$ and $\mathbf{X}^{(\ell)}$ an $n_\ell \times r_\ell$ matrix with orthonormal columns $\mathbf{x}_{q_\ell}^{(\ell)}$. The range of $\mathbf{X}^{(\ell)}$ is the ℓ -th singular subspace of \mathcal{P} . This decomposition is illustrated for the case $d = 3$ in Figure 1. Model (2) with \mathcal{P} as in Equation (3) is the most general spiked tensor model — i.e., low-rank perturbation of a large random tensor. Indeed, any of the models referred to in the previous Section 1.2 fall into this definition since decomposition (3) always exists and low CPD-rank is equivalent to low multilinear-rank thanks to the inequality $\max_\ell \{r_\ell\} \leq R \leq \min_{\ell \neq \ell'} \{r_\ell r_{\ell'}\}$ (Sidiropoulos et al., 2017).

In the regime where $N \rightarrow +\infty$ — representing the fact that, in practice, the dimensions of the tensor are large compared to its rank —, we study the estimation of \mathcal{P} from \mathcal{T} with a truncated MLSVD, which serves as initialization of the HOOI algorithm. In particular,

4. That is, n_ℓ/N neither vanishes nor diverges as $N \rightarrow +\infty$. This ensures that the spectral norm of $\frac{1}{\sqrt{N}}\mathcal{N}$ is of order 1 (Tomioka and Suzuki, 2014).

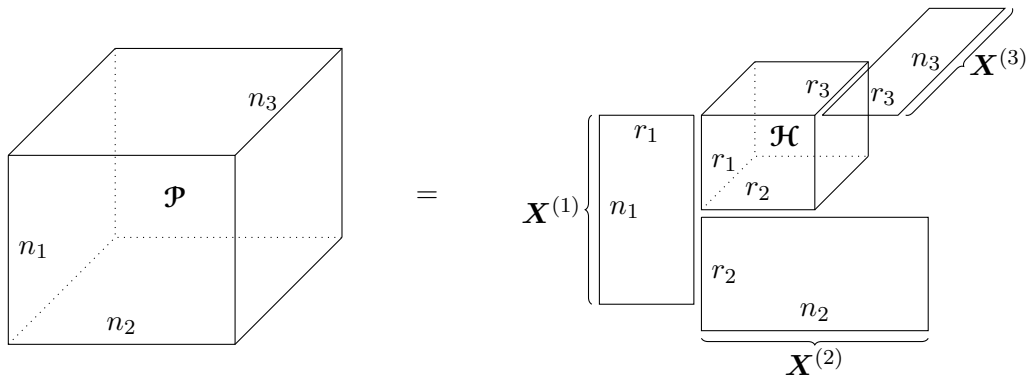


Figure 1: Illustration of the Tucker decomposition (3) of an $n_1 \times n_2 \times n_3$ tensor \mathcal{P} with multilinear-rank (r_1, r_2, r_3) . \mathcal{H} is the $r_1 \times r_2 \times r_3$ core tensor and $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, $\mathbf{X}^{(3)}$ are matrices with orthonormal columns spanning the singular subspaces of \mathcal{P} .

we reveal that the interesting *non-trivial* regime is characterized by the $\Theta(N^{\frac{d-2}{2}})$ quantity⁵ $\sigma_N = \frac{1}{N} \prod_{\ell=1}^d \sqrt{n_\ell}$.

- If $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N \xrightarrow{N \rightarrow +\infty} 0$, then the noise completely masks the signal, and truncated MLSVD fails to recover \mathcal{P} .
- If $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N \xrightarrow{N \rightarrow +\infty} +\infty$, then the signal clearly stands out from the noise, and reconstruction of \mathcal{P} with a truncated MLSVD is easy.
- If $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N = \Theta(1)$ as $N \rightarrow +\infty$, then we are precisely in the *non-trivial* regime between the two previous situations, and truncated MLSVD may partially recover \mathcal{P} .

It is the analysis of this last regime which is of practical interest. Given the low-multilinear-rank approximation $\hat{\mathcal{T}} = \llbracket \hat{\mathcal{G}}; \hat{\mathcal{U}}^{(1)}, \dots, \hat{\mathcal{U}}^{(d)} \rrbracket$ obtained with a truncated MLSVD of \mathcal{T} , we quantify how well $\hat{\mathcal{T}}$ reconstructs \mathcal{P} in this non-trivial regime. To do so, we study the spectral properties of the unfoldings (i.e., matricizations) of the tensor \mathcal{T} , i.e., the $n_\ell \times \prod_{\ell' \neq \ell} n_{\ell'}$ matrices $\mathbf{T}^{(\ell)}$ whose columns are mode- ℓ fibers of \mathcal{T} and the columns of $\hat{\mathcal{U}}^{(\ell)}$ are its dominant left singular vectors. Such *long* matrices (the second dimension grows faster than the first one) have already been studied by Ben Arous et al. (2023) in order to analyze the properties of tensor-unfolding methods in the particular setting of a rank-1 spike. Here, we tackle this problem with a highly different approach relying solely on classical tools from random matrix theory (Couillet and Liao, 2022) and give very general results on the spiked tensor model that go beyond the specific rank-1 one case. Moreover, we justify the practical use of truncated MLSVD as an initialization of the HOOI algorithm by showing its optimality in the large N regime.

5. Note that this corresponds to the regime of the algorithmic threshold but, here, our measure of the signal power is $\|\mathcal{P}\|_{\mathbb{F}}^2$ (and not $\|\mathcal{P}\|_{\mathbb{F}}$) hence the $\Theta(N^{\frac{d-2}{2}})$ instead of $\Theta(N^{\frac{d-2}{4}})$.

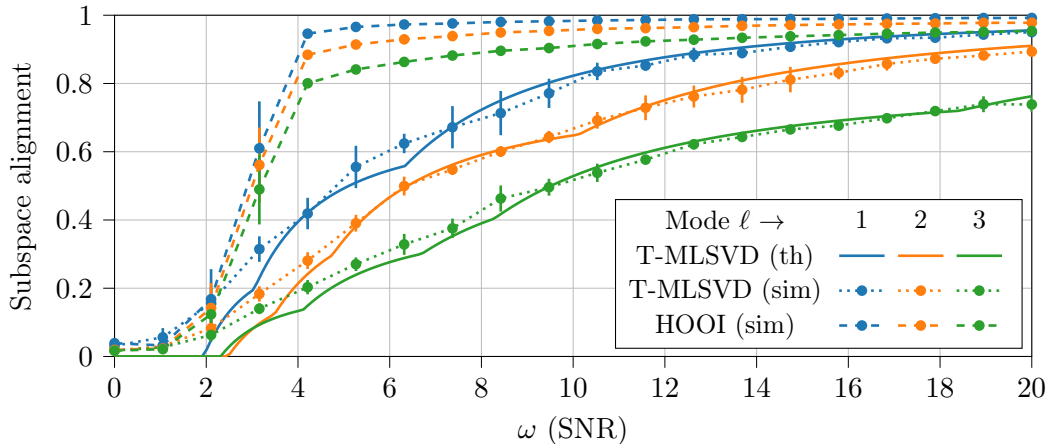


Figure 2: Alignments between singular subspaces (see Section 3.2) of the observation $\mathcal{J} = \sqrt{\omega}\mathcal{P}_\circ + \frac{1}{\sqrt{N}}\mathcal{N}$ and of the signal \mathcal{P}_\circ , with $\|\mathcal{P}_\circ\|_{\mathbb{F}}^2 = \frac{\sqrt{n_1 n_2 n_3}}{N}$, as a function of the signal-to-noise ratio ω . Theoretical alignments (Theorem 8) achieved with truncated MLSVD are compared with simulations and those achieved with the HOOI algorithm. Empirical results are averaged over 10 trials, with error bars representing standard deviation. **Experimental setting:** $d = 3$, $(n_1, n_2, n_3) = (100, 200, 300)$, $N = n_1 + n_2 + n_3$ and $(r_1, r_2, r_3) = (3, 4, 5)$.

Although the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$, $\ell \in \{1, \dots, d\}$, diverges as $N \rightarrow +\infty$, we show that its eigenvalues (i.e., the squared singular values of $\mathbf{T}^{(\ell)}$) gather in an interval $[\mu_N^{(\ell)} \pm 2\sigma_N]$ with $\mu_N^{(\ell)} = \frac{1}{N} \prod_{\ell' \neq \ell} n_{\ell'} = \Theta(N^{d-2})$. More precisely, the empirical spectral distribution of the centered-and-scaled matrix $\frac{1}{\sigma_N}[\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)}\mathbf{I}_{n_\ell}]$ converges weakly to the semicircle distribution on $[-2, +2]$ (see Figure 3, Theorem 5 and Corollary 6). Furthermore, we show that a BBP phase transition phenomenon occurs: for each singular value of $\mathbf{P}^{(\ell)}$ (the unfolding of \mathcal{P} along mode ℓ) which is above the threshold $\sqrt{\sigma_N}$, an eigenvalue of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ isolates itself on the right side of the *bulk* (see Figure 3) and its corresponding eigenvector (i.e., left singular vector of $\mathbf{T}^{(\ell)}$) is aligned with the corresponding singular subspace of \mathcal{P} . The position of the isolated eigenvalue and this alignment are efficiently predicted by Theorem 8 (see also Figure 3).

As a result, Figure 2 plots, for an order-3 tensor, as a function of the signal-to-noise ratio (SNR) $\omega = \|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N$, the alignments between the singular subspace of the signal \mathcal{P} spanned by $\mathbf{X}^{(\ell)}$ and the dominant singular subspace of the observation \mathcal{J} spanned by $\hat{\mathbf{U}}^{(\ell)}$. Solid curves are the alignments predicted by Theorem 8 while dotted curves are empirical alignments computed on a $100 \times 200 \times 300$ tensor with signal-rank $(3, 4, 5)$. If the SNR ω is too small, there is no alignment, meaning that truncated MLSVD fails to recover \mathcal{P} — the signal is masked by the noise. When it exceeds a critical value (see Theorem 8 and Section

3.2 for details), a phase transition phenomenon occurs⁶: the alignment starts to grow — i.e., truncated MLSVD now partially recovers \mathcal{P} — and converges to 1 as $\omega \rightarrow +\infty$.

Besides, Figure 2 also plots the empirical alignments between the singular subspaces of \mathcal{P} and those estimated with the HOOI algorithm (De Lathauwer et al., 2000a) given in Algorithm 1, whose truncated MLSVD serves as initialization. This yields much better alignments, especially close to the phase transition. In fact, we show in Theorem 14 that the HOOI algorithm converges to a solution to Problem (2) as soon as its initialization sufficiently preserves the underlying signal. This provides new insight into the computational barrier: initialization is the limiting factor here. Had one prior information on the solution, one could initialize the HOOI algorithm in the right basin of attraction and still be able to perfectly (i.e., with alignment 1) reconstruct the signal in the regime $1 \ll \|\mathcal{P}\|_F \ll N^{\frac{d-2}{4}}$, which is computationally hard but statistically easy (see details in Section 4.1 and discussion in Section 4.2).

In a nutshell, our contributions can be summarized as follows.

- We characterize, in the large N limit, the behavior of the singular values of the unfoldings of the tensor \mathcal{J} — denoted $\mathbf{T}^{(\ell)}$, $\ell \in \{1, \dots, d\}$ — when it follows the general spiked tensor model (2) (Theorem 5 and Corollary 6). This is performed through the analysis of the limiting spectral distribution of the symmetric matrix $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ using standard tools from the theory of large random matrices.
- We give a precise condition, depending on a $\Theta(N^{\frac{d-2}{4}})$ threshold on the signal, for the detectability of a principal direction of the ℓ -th singular subspace of \mathcal{P} from the unfolding $\mathbf{T}^{(\ell)}$. This corresponds to the presence of an isolated eigenvalue in the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ with associated eigenvector aligned with the sought singular subspace. We find similar formulae as Feldman (2023)⁷ for the asymptotic position of this isolated eigenvalue, as well as the quality of the alignment (Theorem 8).
- Relying on our random matrix analysis, we characterize the performance of truncated MLSVD in the reconstruction of the signal \mathcal{P} from the observation \mathcal{J} (Section 3.2).
- We show that exact reconstruction of \mathcal{P} from \mathcal{J} is possible in the large N regime with the HOOI algorithm (De Lathauwer et al., 2000a) as long as $\|\mathcal{P}\|_F \gg 1$ and it is initialized in the right basin of attraction (Theorem 14). Without prior information, this depends on the detectability of \mathcal{P} in the truncated MLSVD of \mathcal{J} , which is only possible above the $\Theta(N^{\frac{d-2}{4}})$ computational threshold. Moreover, as $N \rightarrow +\infty$, the number of iterations needed for the convergence of the algorithm converges to 1.

In section 2, we introduce our notations, tensor-related operations and decompositions and useful tools from random matrix theory. Section 3 presents the random matrix analysis of long matrices emerging from the unfoldings of tensors following the general spiked tensor

6. In fact, we will see in Section 3.2 that there is one phase transition for each principal direction of the singular subspaces of \mathcal{P} , resulting in $\sum_{\ell=1}^d r_\ell$ phase transitions. Their positions corresponds to sudden changes of slope in the solid curves of Figure 2.

7. Feldman (2023) studies the spiked model associated with long random matrices within a similar framework as Benaych-Georges and Nadakuditi (2011). Our approach is different in that it mostly relies on a *deterministic equivalent* which is introduced in Theorem 5.

model (2). These results are presented in the context of truncated MLSVD and exploited to quantitatively explain its reconstruction performances. Then, relying on these results, Section 4 deals with the numerical estimation of a solution to Problem (1) with the HOOI algorithm. We show its asymptotic optimality and provide insight into the limiting factors for numerical estimation below the computational threshold. We conclude and discuss our results in Section 5. Most proofs are deferred to the appendix.

2 Preliminaries on Tensors and Random Matrix Theory

We start by introducing some notation and the main tools which are needed to expose the results of the next sections.

2.1 General Notations

a , \mathbf{a} , \mathbf{A} and \mathcal{A} respectively denote a scalar, a vector, a matrix and a tensor. a_i , $A_{i,j}$ and $\mathcal{A}_{i_1, \dots, i_d}$ are their entries. For $x \in \mathbb{R}$, $[x]^+ = \max(0, x)$. The imaginary part of $z \in \mathbb{C}$ is $\Im[z]$. The set $\{1, \dots, n\}$ of positive integers smaller or equal to n is denoted $[n]$. δ_x is the Dirac measure at point x . The support of a probability measure μ is denoted $\text{Supp } \mu$. The notation $X \sim \mathcal{L}$ means that the random variable X is distributed according to the law \mathcal{L} . Given a sequence of random variables $(X_n)_{n \geq 0}$, its convergence in distribution to \mathcal{L} is denoted $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{L}$ and its almost sure convergence to L is denoted $X_n \xrightarrow[n \rightarrow +\infty]{\text{a.s.}} L$.

The normal distribution with mean μ and variance σ^2 is denoted $\mathcal{N}(\mu, \sigma^2)$. The span of an $n_1 \times n_2$ matrix \mathbf{A} is $\text{Span } \mathbf{A} = \{\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^{n_2}\} \subset \mathbb{R}^{n_1}$. The singular values of \mathbf{A} in *non-increasing* order are denoted $s_1(\mathbf{A}) \geq s_2(\mathbf{A}) \geq \dots \geq 0$. Given an $n \times n$ matrix \mathbf{B} , its trace is $\text{Tr } \mathbf{B} = \sum_{i=1}^n B_{i,i}$ and its spectrum, $\text{Sp } \mathbf{B}$, is the set of all its eigenvalues. $\|\cdot\|$ denotes the standard Euclidean norm for vectors and the corresponding operator norm (spectral norm) for matrices.

Given two real sequences $(u_n)_{n \geq 0}$ and $(v_n)_{n \geq 0}$, we write $u_n = \mathcal{O}_{n \rightarrow +\infty}(v_n)$ if there exists a constant $C > 0$ and an integer n_0 such that $|u_n| \leq C|v_n|$ as soon as $n \geq n_0$. If $u_n = \mathcal{O}_{n \rightarrow +\infty}(v_n)$ and $v_n = \mathcal{O}_{n \rightarrow +\infty}(u_n)$ ⁸, then we write $u_n = \Theta_{n \rightarrow +\infty}(v_n)$. We also write $u_n \ll_{n \rightarrow +\infty} v_n$ (or $v_n \gg_{n \rightarrow +\infty} u_n$) if, for all $\varepsilon > 0$, there exists an integer n_0 such that $|u_n| \leq \varepsilon|v_n|$ as soon as $n \geq n_0$. When it is clear from context that $n \rightarrow +\infty$, we simply write $u_n = \Theta(v_n)$, $u_n = \mathcal{O}(v_n)$, $u_n \ll v_n$ or $u_n \gg v_n$.

Unless stated otherwise, d represents the order of a tensor; ℓ is an index ranging from 1 to d and i_ℓ, q_ℓ range from 1 to n_ℓ, r_ℓ respectively.

2.2 Tensors, Related Operations and Decompositions

For our purposes, tensors are considered as multi-way arrays: $\mathcal{J} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is a collection of elements $\mathcal{J}_{i_1, \dots, i_d} \in \mathbb{R}$ with $i_\ell \in [n_\ell]$, $\ell \in [d]$. An ℓ -*fiber* of \mathcal{J} is the vector of \mathbb{R}^{n_ℓ} obtained by fixing all the indices of \mathcal{J} but the ℓ -th. This is the generalization of columns and rows of matrices, which are respectively 1- and 2-fibers. *Unfolding* (or *matricization*) is the process

8. I.e., if there exist two constants $c, C > 0$ and an integer n_0 such that $c|v_n| \leq |u_n| \leq C|v_n|$ as soon as $n \geq n_0$.

by which a matrix is built from a tensor — $\mathbf{T}^{(\ell)}$ is the unfolding of \mathcal{T} along mode ℓ , i.e., the $n_\ell \times \prod_{\ell' \neq \ell} n_{\ell'}$ matrix whose columns are ℓ -fibers of \mathcal{T} ⁹.

Given two $n_1 \times \dots \times n_d$ tensors \mathcal{A}, \mathcal{B} , their scalar product is

$$\langle \mathcal{A}, \mathcal{B} \rangle_{\mathbb{F}} \stackrel{\text{def}}{=} \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \mathcal{A}_{i_1, \dots, i_d} \mathcal{B}_{i_1, \dots, i_d}.$$

The Frobenius norm of \mathcal{A} is $\|\mathcal{A}\|_{\mathbb{F}} \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle_{\mathbb{F}}}$.

Just as a matrix is said to have rank 1 if it can be expressed as the *outer product* of two vectors, $\mathbf{xy}^\top \stackrel{\text{def}}{=} \mathbf{x} \otimes \mathbf{y}$, an order- d tensor is said to have rank 1 if it can be expressed as the outer product of d vectors, i.e., $\mathcal{T}_{i_1, \dots, i_d} = x_{i_1}^{(1)} \dots x_{i_d}^{(d)} \iff \mathcal{T} \stackrel{\text{def}}{=} \bigotimes_{\ell=1}^d \mathbf{x}^{(\ell)}$.

Given two matrices \mathbf{A} and \mathbf{B} of respective sizes $n_1 \times n_2$ and $p_1 \times p_2$, their *Kronecker product*, denoted $\mathbf{A} \boxtimes \mathbf{B}$, is the $n_1 p_1 \times n_2 p_2$ matrix such that $[\mathbf{A} \boxtimes \mathbf{B}]_{p_1(i_1-1)+j_1, p_2(i_2-1)+j_2} = A_{i_1, i_2} B_{j_1, j_2}$ for all $(i_1, i_2) \in [n_1] \times [n_2]$, $(j_1, j_2) \in [p_1] \times [p_2]$. It is also defined for n -dimensional vectors, seen as $n \times 1$ matrices. This product is useful to express unfoldings of tensors defined as outer products. For instance,

$$[\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}]^{(1)} = \mathbf{x}(\mathbf{y} \boxtimes \mathbf{z})^\top, \quad [\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}]^{(2)} = \mathbf{y}(\mathbf{x} \boxtimes \mathbf{z})^\top, \quad [\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}]^{(3)} = \mathbf{z}(\mathbf{x} \boxtimes \mathbf{y})^\top. \quad (4)$$

Among the various properties of the Kronecker product (Abadir and Magnus, 2005, Chapter 10), we highlight that it is bilinear, associative, *non-commutative*, $(\mathbf{A} \boxtimes \mathbf{B})^\top = \mathbf{A}^\top \boxtimes \mathbf{B}^\top$ and $(\mathbf{A} \boxtimes \mathbf{B})(\mathbf{C} \boxtimes \mathbf{D}) = (\mathbf{AC}) \boxtimes (\mathbf{BD})$ when the matrix products \mathbf{AC} and \mathbf{BD} are defined.

Because of their multidimensional nature, tensors can be very large and cause computational and storage difficulties. Hence, tensor decompositions are an appropriate way to provide a sparse representation of these objects and reveal relevant knowledge from multi-way data arrays. We briefly present two of the most common decompositions, namely CPD and MLSVD, although many others exist (see, e.g., Kolda and Bader (2009); Vervliet et al. (2014)).

Canonical Polyadic Decomposition (CPD). \mathcal{T} is decomposed as a sum of R rank-1 terms, $\mathcal{T} = \sum_{i=1}^R \sigma_i \bigotimes_{\ell=1}^d \mathbf{a}_i^{(\ell)}$, with $\mathbf{a}_i^{(\ell)}$ of unit norm, $\sigma_i > 0$ and minimal R . Introduced by Hitchcock (1927), and also called CANDECOP or PARAFAC, this decomposition is, up to permutation, essentially unique under mild conditions (Kolda and Bader, 2009).

Multilinear Singular Value Decomposition (MLSVD). \mathcal{T} is written as the multilinear tensor-matrix product of an $r_1 \times \dots \times r_d$ core tensor \mathcal{G} with d factors $\mathbf{U}^{(\ell)}$, $\ell \in [d]$, of respective sizes $n_\ell \times r_\ell$ and orthonormal columns $\mathbf{u}_{q_\ell}^{(\ell)}$,

$$\mathcal{T} = \sum_{q_1=1}^{r_1} \cdots \sum_{q_d=1}^{r_d} \mathcal{G}_{q_1, \dots, q_d} \bigotimes_{\ell=1}^d \mathbf{u}_{q_\ell}^{(\ell)} \stackrel{\text{def}}{=} \left[\mathcal{G}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)} \right].$$

r_ℓ is the dimension of the subspace spanned by the ℓ -fibers of \mathcal{T} , whose an orthonormal basis is formed by the columns of $\mathbf{U}^{(\ell)}$. This decomposition, introduced by Tucker (1966), is also

9. The order of the columns does not matter, as long as other operations, such as the Kronecker product, are defined in a consistent manner.

called higher-order SVD (HOSVD, De Lathauwer et al., 2000b). From this decomposition, the unfoldings of \mathcal{J} can be expressed with matrix and Kronecker products between the factor matrices and the unfoldings of the core tensor \mathcal{G} . For instance, if $\mathcal{J} = \llbracket \mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$,

$$\mathbf{T}^{(1)} = \mathbf{U}\mathbf{G}^{(1)}(\mathbf{V} \boxtimes \mathbf{W})^\top, \quad \mathbf{T}^{(2)} = \mathbf{V}\mathbf{G}^{(2)}(\mathbf{U} \boxtimes \mathbf{W})^\top, \quad \mathbf{T}^{(3)} = \mathbf{W}\mathbf{G}^{(3)}(\mathbf{U} \boxtimes \mathbf{V})^\top.$$

Note how these expressions generalize (4).

The *contraction* of \mathcal{J} on $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}$ with $\mathbf{A}^{(\ell)} \in \mathbb{R}^{n_\ell \times p_\ell}$, $\ell \in [d]$, is the $p_1 \times \dots \times p_d$ tensor $\mathcal{J}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})$ whose (j_1, \dots, j_d) -entry is $\sum_{i_1=1}^{n_1} \dots \sum_{i_d=1}^{n_d} \mathcal{J}_{i_1, \dots, i_d} \prod_{\ell=1}^d A_{i_\ell, j_\ell}^{(\ell)}$. If $\mathcal{J} = \llbracket \mathcal{G}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)} \rrbracket$ then $\mathcal{G} = \llbracket \mathcal{J}; \mathbf{U}^{(1)\top}, \dots, \mathbf{U}^{(d)\top} \rrbracket = \mathcal{J}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)})$. The tensor contraction generalizes the matrix operation $\mathbf{A}^{(1)\top} \mathbf{M} \mathbf{A}^{(2)} = \mathbf{M}(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$.

Remark 1 (Uniqueness of the MLSVD up to isometries) *For all $r_\ell \times r_\ell$ orthogonal matrices $\mathbf{O}^{(\ell)}$, $\ell \in [d]$, we have the equivalent decomposition*

$$\llbracket \mathcal{G}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)} \rrbracket = \llbracket \check{\mathcal{G}}; \mathbf{U}^{(1)} \mathbf{O}^{(1)}, \dots, \mathbf{U}^{(d)} \mathbf{O}^{(d)} \rrbracket$$

where $\check{\mathcal{G}} = \mathcal{G}(\mathbf{O}^{(1)}, \dots, \mathbf{O}^{(d)})$ is the contraction of \mathcal{G} on $\mathbf{O}^{(1)}, \dots, \mathbf{O}^{(d)}$. Up to isometries, the multilinear singular value decomposition is unique (De Lathauwer et al., 2000b, Property 4).

The actual rank of a tensor is usually understood as its CPD-rank R whereas its MLSVD-rank (r_1, \dots, r_d) is called *multilinear-rank*. These two ranks are constrained by the following inequalities,

$$\max_{1 \leq \ell \leq d} \{r_\ell\} \leq R \leq \min_{1 \leq \ell \neq \ell' \leq d} \{r_\ell r_{\ell'}\}, \quad (5)$$

which implies one can generally speak of a “low-rank tensor” without having to mention the rank to which it refers.

2.3 Tools from Random Matrix Theory

The results presented in this work rely on tools from the theory of large random matrices. Given a certain $n \times n$ matrix \mathbf{M} whose entries are random variables, one is interested in the behavior its eigenvalues and eigenvectors as $n \rightarrow +\infty$. Common questions are, in this regime, how does the *empirical spectral distribution* (ESD) of \mathbf{M} , i.e., $\frac{1}{n} \sum_{\lambda \in \text{Sp } \mathbf{M}} \delta_\lambda$, behave? Does it (weakly) converge to a *limiting spectral distribution* (LSD)? Are there any isolated eigenvalues? If so, are the corresponding eigenvectors aligned with a relevant subspace?

In case \mathbf{M} is symmetric, its eigenvalues and eigenvectors are real and a key tool to answer the preceding questions is its resolvent, $\mathbf{Q}_\mathbf{M}(z) = (\mathbf{M} - z\mathbf{I}_n)^{-1}$, defined for all $z \in \mathbb{C} \setminus \text{Sp } \mathbf{M}$. Indeed, $\frac{1}{n} \text{Tr } \mathbf{Q}_\mathbf{M}(z) = \frac{1}{n} \sum_{\lambda \in \text{Sp } \mathbf{M}} \frac{1}{\lambda - z}$ is the *Stieltjes transform* of its empirical spectral distribution.

Definition 2 (Stieltjes transform) *The Stieltjes transform of a real probability measure μ is defined for all $z \in \mathbb{C} \setminus \text{Supp } \mu$ as $m_\mu(z) = \int_{\mathbb{R}} \frac{1}{t-z} d\mu(t)$.*

The knowledge of m_μ is equivalent to the knowledge of μ thanks to the inversion formula $\mu([a, b]) = \frac{1}{\pi} \lim_{\eta \downarrow 0} \int_a^b \Im[m_\mu(x + i\eta)] dx$ for any continuity points a, b of μ (Couillet and Liao,

2022, Theorem 2.1). Hence, following the behavior of $\frac{1}{n} \text{Tr} \mathbf{Q}_M(z)$ as $n \rightarrow +\infty$ gives insight into the limiting spectral distribution of \mathbf{M} via its Stieltjes transform. To do so, we seek first a *deterministic equivalent* of \mathbf{Q}_M .

Definition 3 (Matrix equivalent) For two (random) matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times n}$, we write $\mathbf{X} \leftrightarrow \mathbf{Y}$ if, for any deterministic matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ of bounded norms (spectral and Euclidean norms respectively)¹⁰,

$$\frac{1}{n} \text{Tr} \mathbf{A} (\mathbf{X} - \mathbf{Y}) \xrightarrow[n \rightarrow +\infty]{a.s.} 0, \quad \mathbf{a}^\top (\mathbf{X} - \mathbf{Y}) \mathbf{b} \xrightarrow[n \rightarrow +\infty]{a.s.} 0 \quad \text{and} \quad \|\mathbb{E}[\mathbf{X} - \mathbf{Y}]\| \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

If \mathbf{Y} is deterministic, then it is said to be a deterministic equivalent of \mathbf{X} . For more details on this notion, see Couillet and Liao (2022, §2.1.4).

The following lemma will be extensively used to derive such equivalents.

Lemma 4 (Stein, 1981) Let $Z \sim \mathcal{N}(0, 1)$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuously differentiable function. When the following expectations exist, $\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)]$.

Once a deterministic equivalent $\bar{\mathbf{Q}}_M$ of \mathbf{Q}_M is found, the limiting spectral distribution of \mathbf{M} is accessible through the limit of $\frac{1}{n} \text{Tr} \bar{\mathbf{Q}}_M$ as $n \rightarrow +\infty$. In the study of spiked models, eigenvalues are usually gathered in a *bulk*, described by the limiting spectral distribution (LSD), with a finite number of isolated eigenvalues, which do not appear in the LSD (see Figure 3 for example). The position of these isolated eigenvalues are singular points of the resolvent \mathbf{Q}_M . Hence, our deterministic equivalent provides equations defining their almost sure asymptotic position. Moreover, thanks to Cauchy's integral formula, for all $\mathbf{a} \in \mathbb{R}^n$, $|\mathbf{a}^\top \mathbf{u}|^2 = \frac{-1}{2i\pi} \oint_\gamma \mathbf{a}^\top \mathbf{Q}_M(z) \mathbf{a} dz$ where γ is a positively-oriented complex contour circling around an isolated eigenvalue (assuming it has multiplicity 1) with all other eigenvalues left outside and \mathbf{u} is the eigenvector associated to the corresponding eigenvalue. The almost sure asymptotic value of this contour integral can be computed with our deterministic equivalent, thus yielding formulae for the alignments of the *spike* eigenvectors with relevant directions of the model.

This briefly summarizes the main techniques we use in our proofs. A complete presentation of these tools can be found in Couillet and Liao (2022). Other valuable references for the reader interested in random matrix theory are Potters and Bouchaud (2020); Bai and Silverstein (2010); Pastur and Shcherbina (2011); Tao (2012).

3 Analysis of Truncated MLSVD under the General Spiked Tensor Model

This section presents a random matrix analysis of the general spiked tensor model introduced in Equation (2) using tools presented in Section 2.3. We give precise results on the spectral behavior of the unfoldings of the observed tensor \mathcal{J} , and specify the achievable performance in the estimation of the underlying signal \mathcal{P} with a truncated MLSVD. Although,

10. Strictly speaking, we refer to *sequences* of matrices and vectors whose size grow with the index n . Yet, this mathematical aspect is made implicit in order to simplify notation and be closer to practical considerations where these sequences, in fact, do not exist — the assumption $n \rightarrow +\infty$ models the fact that n is large but remains finite.

as explained in Section 1.1, this approach is only *quasi-optimal*, it is very easy to implement and represents an excellent “first guess” to initialize a numerical scheme converging to a solution to Problem (1), which is discussed in Section 4.

3.1 Random Matrix Results on the Model

Under the general spiked tensor model (2), we consider an $n_1 \times \dots \times n_d$ tensor $\mathcal{J} = \mathcal{P} + \frac{1}{\sqrt{N}}\mathcal{N}$ of order $d \geq 3$, modeling a low-rank deterministic signal \mathcal{P} corrupted by an additive Gaussian noise tensor \mathcal{N} whose entries are independent $\mathcal{N}(0, 1)$ random variables¹¹. We denote by (r_1, \dots, r_d) the multilinear-rank of \mathcal{P} and study this model in the asymptotic regime $N \rightarrow +\infty$ with $n_\ell = \Theta(N)$ and $r_\ell = \Theta(1)$, $\ell \in [d]$.

The estimation of \mathcal{P} with a truncated MLSVD on \mathcal{J} is simply the computation of the dominant singular subspaces of \mathcal{J} . Specifically, $\hat{\mathbf{U}}^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$ gathers the r_ℓ dominant left singular vectors of $\mathbf{T}^{(\ell)}$ — and thus, $\hat{\mathbf{U}}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)} = \mathbf{I}_{r_\ell}$. Then, a low-multilinear-rank approximation of \mathcal{J} is $\hat{\mathcal{J}} = \llbracket \hat{\mathcal{G}}; \hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)} \rrbracket$ with an $r_1 \times \dots \times r_d$ core tensor $\hat{\mathcal{G}} = \mathcal{J}(\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)})$. An equivalent expression is $\hat{\mathcal{J}} = \llbracket \mathcal{J}; \hat{\mathbf{U}}^{(1)} \hat{\mathbf{U}}^{(1)\top}, \dots, \hat{\mathbf{U}}^{(d)} \hat{\mathbf{U}}^{(d)\top} \rrbracket$, which explicitly shows that $\hat{\mathcal{J}}$ is the projection of \mathcal{J} on its dominant singular subspaces¹². Thus, the quality of this estimation hinges upon the alignments between the singular subspaces of \mathcal{P} and the dominant singular subspaces of \mathcal{J} . Namely, denoting $\hat{\mathbf{u}}_{q_\ell}^{(\ell)}$, for $q_\ell \in [r_\ell]$, the columns of $\hat{\mathbf{U}}^{(\ell)} = \begin{bmatrix} \hat{\mathbf{u}}_1^{(\ell)} & \dots & \hat{\mathbf{u}}_{r_\ell}^{(\ell)} \end{bmatrix}$ and given that $\mathcal{P} = \llbracket \mathcal{H}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)} \rrbracket$, the quantities of interest are $\|\mathbf{X}^{(\ell)\top} \hat{\mathbf{u}}_{q_\ell}^{(\ell)}\|^2$ since they represent how much of $\hat{\mathbf{u}}_{q_\ell}^{(\ell)}$ is in the ℓ -th singular subspace of the signal \mathcal{P} .

In order to understand how the singular subspaces of \mathcal{P} are perturbed by the addition of noise, we study the spectral properties of the unfoldings $\mathbf{T}^{(\ell)} = \mathbf{P}^{(\ell)} + \frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}$. In fact, since we are only interested in the *left* singular vectors of $\mathbf{T}^{(\ell)}$, it is more convenient to consider the $n_\ell \times n_\ell$ *symmetric* matrix $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$. Note that this is different from standard spiked matrix models (Benaych-Georges and Nadakuditi, 2011) because the second dimension of $\mathbf{T}^{(\ell)}$ grows at a faster polynomial rate than the first one ($\Theta(N^{d-1})$ versus $\Theta(N)$). Hence, it is easy to see that the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ should diverge as $N \rightarrow +\infty$: set $\mathcal{P} = \mathbf{0}$ for simplicity and consider the expected mean of the eigenvalues,

$$\frac{1}{n_\ell} \mathbb{E} \left[\sum_{\lambda \in \text{Sp}(\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top})} \lambda \right] = \frac{1}{n_\ell} \mathbb{E} \left[\text{Tr} \left(\frac{1}{N} \mathbf{N}^{(\ell)} \mathbf{N}^{(\ell)\top} \right) \right] = \frac{1}{N} \prod_{\ell' \neq \ell} n_{\ell'} \xrightarrow{N \rightarrow +\infty} +\infty.$$

Hence, we need to consider instead a *centered-and-scaled* version of our random matrix $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ to properly study the behavior of its spectrum. The quantities $\mu_N^{(\ell)}$ and σ_N

11. We highlight the fact that the *Gaussian* noise assumption is not restrictive. Firstly, the universality result of Gurau (2014) shows that, as $N \rightarrow +\infty$, the distribution of a random tensor with i.i.d. entries has the same limit than that of a tensor with i.i.d. *Gaussian* entries. Moreover, our results can be extended to non-Gaussian noise up to a control on the moments of the distribution with the “interpolation trick” of Lytova and Pastur (2009, Corollary 3.1). These are technical aspects which go beyond the scope of our work. Examples of results relying on this method are Pastur and Shcherbina (2011, Theorems 18.4.2 and 19.2.1); Merlevède et al. (2015); Banna et al. (2015).

12. In the more familiar matrix case ($d = 2$), the expression $\llbracket \mathbf{T}; \hat{\mathbf{U}}^{(1)} \hat{\mathbf{U}}^{(1)\top}, \hat{\mathbf{U}}^{(2)} \hat{\mathbf{U}}^{(2)\top} \rrbracket$ is equivalent to $\hat{\mathbf{U}}^{(1)} \hat{\mathbf{U}}^{(1)\top} \mathbf{T} \hat{\mathbf{U}}^{(2)} \hat{\mathbf{U}}^{(2)\top}$.

introduced in Theorem 5 below are such that the eigenvalues of $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}]$ neither diverge nor vanish but stay at a $\Theta(1)$ scale as $N \rightarrow +\infty$.

Theorem 5 (Deterministic equivalent) *For $\ell \in [d]$, define the following quantities,*

$$\mu_N^{(\ell)} = \frac{1}{N} \prod_{\ell' \neq \ell} n_{\ell'}, \quad \sigma_N = \frac{1}{N} \sqrt{\prod_{\ell \in [d]} n_\ell}.$$

As $N \rightarrow +\infty$, if the ratio $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N$ is bounded, then the resolvent of the centered-and-scaled matrix $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}]$ has the following deterministic equivalent¹³, defined for all $\tilde{z} \in \mathbb{C} \setminus \mathbb{R}$,

$$\begin{aligned} \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) &\stackrel{\text{def}}{=} \left(\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}] - \tilde{z} \mathbf{I}_{n_\ell} \right)^{-1} \\ &\longleftrightarrow \bar{\mathbf{Q}}^{(\ell)}(\tilde{z}) \stackrel{\text{def}}{=} \left(\frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} + \frac{1}{\tilde{m}(\tilde{z})} \mathbf{I}_{n_\ell} \right)^{-1} \end{aligned}$$

where $\tilde{m}(\tilde{z}) \stackrel{\text{def}}{=} \lim_{N \rightarrow +\infty} \frac{1}{n_\ell} \text{Tr} \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z})$ does not depend on $\ell \in [d]$ and satisfies the following equation,

$$\tilde{m}^2(\tilde{z}) + \tilde{z} \tilde{m}(\tilde{z}) + 1 = 0. \quad (6)$$

Proof See Appendix A. ■

This theorem is fundamental. It gives a deterministic equivalent $\bar{\mathbf{Q}}^{(\ell)}(\tilde{z})$ of the resolvent of the centered-and-scaled matrix $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}]$, which is our entry point into the precise characterization of its spectral behavior, following the approach presented in Section 2.3. First of all, notice that the “scaling parameters” $\mu_N^{(\ell)}$ and σ_N are respectively $\Theta(N^{d-2})$ and $\Theta(N^{\frac{d-2}{2}})$, meaning that the eigenvalues of $\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top}$ grow at a speed N^{d-2} and spread over an interval whose length grows as $\sqrt{N^{d-2}}$ and does not depend on the mode ℓ .

Moreover, the relation given in Equation (6) characterizes \tilde{m} , the Stieltjes transform of the limiting spectral distribution (LSD) of $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}]$. We see that this LSD is the same regardless of the low-rank perturbation $\mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top}$, as it is expected that the perturbation should only cause the presence of a *finite* number of isolated eigenvalues in the empirical spectral distribution. Notice that, if $\mathbf{P}^{(\ell)} = \mathbf{0}$, we simply have $\bar{\mathbf{Q}}^{(\ell)}(\tilde{z}) = \tilde{m}(\tilde{z}) \mathbf{I}_{n_\ell}$. As a corollary, we recover the limiting spectral distribution of long random matrices, which was first characterized by Bai and Yin (1988).

Corollary 6 (Limiting spectral distribution) *As $N \rightarrow +\infty$, the empirical spectral distribution of the centered-and-scaled matrix $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell}]$ converges weakly almost surely to μ_{SC} , the semicircle distribution on $[-2, +2]$,*

$$d\mu_{SC}(x) \stackrel{\text{def}}{=} \frac{1}{2\pi} \sqrt{[4 - x^2]^+} dx.$$

¹³. Recall Definition 3.

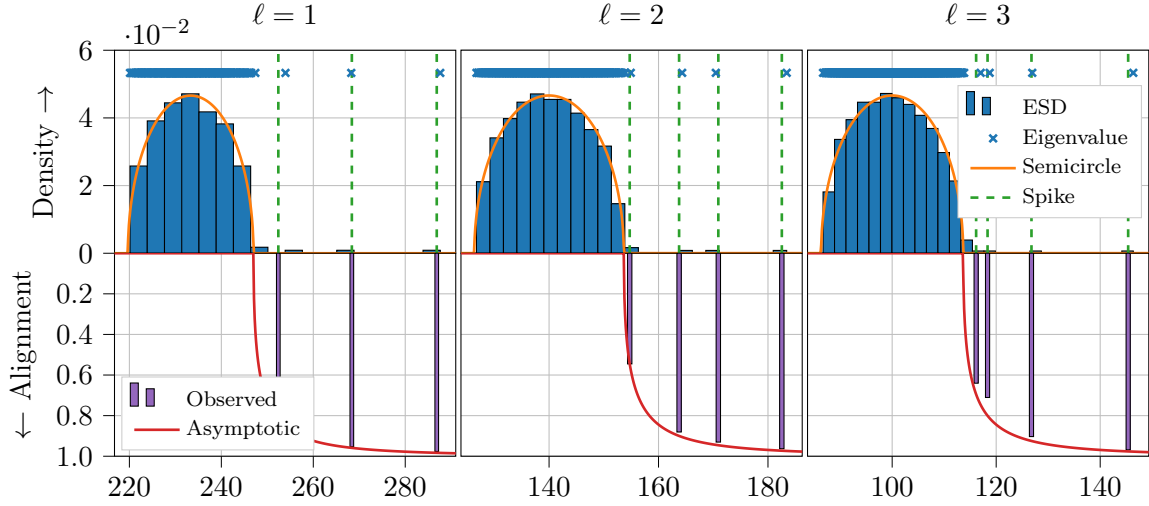


Figure 3: **Top:** empirical spectral distribution (ESD) of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$. The orange curve is the density of the stretched semicircle on $[\mu_N^{(\ell)} \pm 2\sigma_N]$ (Corollary 6). Green dashed lines represent asymptotic positions of spikes $\mu_N^{(\ell)} + \sigma_N \tilde{\xi}_{q_\ell}^{(\ell)}$ (Theorem 8). **Bottom:** Observed alignments between the dominant eigenvectors of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ and $\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}$ (purple bars) with their predicted asymptotic values $[\zeta_{q_\ell}^{(\ell)}]^+$ (red curve, Theorem 8). **Experimental setting:** $d = 3$, $(n_1, n_2, n_3) = (300, 500, 700)$, $N = n_1 + n_2 + n_3$, $(r_1, r_2, r_3) = (3, 4, 5)$ and $\|\mathcal{P}\|_{\mathbb{F}}^2 / \sigma_N = 15$.

Proof Following Equation (6), $\tilde{m}(\tilde{z}) = \frac{1}{2} \left[-z \pm \sqrt{z^2 - 4} \right]$ where the \pm sign is chosen so that \tilde{m} satisfies the properties of a Stieltjes transform, in particular $\Im[\tilde{z}]\Im[\tilde{m}(\tilde{z})] > 0$ for all $\tilde{z} \in \mathbb{C} \setminus \mathbb{R}$. Then, the result follows from the Stieltjes transform inversion formula introduced in Section 2.3 (see also Couillet and Liao (2022, Section 2.1.2)). \blacksquare

This result states that the limiting spectral distribution of $\frac{1}{\sigma_N} \left[\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_\ell} \right]$ is, in fact, the very-well-known semicircle distribution, first observed by Wigner (1955, 1958) in the study of certain special classes of random matrices arising in quantum mechanics. It indicates that, as $N \rightarrow +\infty$, the density of eigenvalues of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ is a stretched semicircle on $[\mu_N^{(\ell)} \pm 2\sigma_N]$. This phenomenon is illustrated in the first row of Figure 3, where the empirical spectral distribution (ESD) of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ is represented with the corresponding stretched semicircle for every mode ℓ of an order-3 tensor of size $300 \times 500 \times 700$ following the general spiked tensor model (2).

Remark 7 (From Marčenko-Pastur to Wigner) Given a random matrix $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$ with i.i.d. $\mathcal{N}(0, \frac{1}{p_2})$ entries, it is well known that the empirical spectral distribution of $\mathbf{X}\mathbf{X}^\top$ converges weakly to the Marčenko-Pastur distribution as $p_1, p_2 \rightarrow +\infty$ with $p_1/p_2 \rightarrow c > 0$ (Marčenko and Pastur, 1967; Pastur and Shcherbina, 2011, Corollary 7.2.5; Couillet and Liao,

2022, Theorem 2.4; Potters and Bouchaud, 2020, Chapter 4). On the other hand, the standard semicircle distribution μ_{SC} is known to be the limiting spectral distribution of symmetric $p \times p$ random matrices with i.i.d. (up to symmetry) $\mathcal{N}(0, \frac{1}{p})$ entries (Pastur and Shcherbina, 2011, Corollary 2.2.8; Couillet and Liao, 2022, Theorem 2.5; Potters and Bouchaud, 2020, Chapter 2). Here, Corollary 6 shows that if p_2 grows at a faster polynomial rate than p_1 , the matrix $\mathbf{X}\mathbf{X}^\top$ behaves asymptotically (up to a deterministic rescaling and shift) like a Wigner matrix, even if its entries are not independent. Experimentally, we observe that, if n_2 and n_3 are chosen small compared to n_1 (in contradiction with our assumption $n_1, n_2, n_3 = \Theta(N)$), e.g., $(n_1, n_2, n_3) = (1000, 40, 40)$, then the empirical spectral distribution of $\mathbf{T}^{(1)}\mathbf{T}^{(1)\top}$ is better modeled by a Marčenko-Pastur distribution than by a Wigner semicircle.

The empirical spectral distributions of Figure 3 also show isolated eigenvalues on the right side of each semicircle. They are caused by the low-rank perturbation \mathcal{P} which, in this setting, has multilinear-rank $(3, 4, 5)$. The estimate of \mathcal{P} given by a truncated MLSVD on \mathcal{J} has its singular subspaces spanned by the dominant eigenvectors of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$, i.e., precisely those associated with these isolated eigenvalues. Hence, a precise characterization of the behavior of these *spikes* is needed to plainly understand the recovery performance of this estimate. As explained in Section 2.3, this can be achieved with the deterministic equivalent given in Theorem 5.

Theorem 8 (Spike behavior) For $\ell \in [d]$ and $q_\ell \in [r_\ell]$, define the following quantities¹⁴,

$$\rho_{q_\ell}^{(\ell)} = \frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N}, \quad \tilde{\xi}_{q_\ell}^{(\ell)} = \rho_{q_\ell}^{(\ell)} + \frac{1}{\rho_{q_\ell}^{(\ell)}} \quad \text{and} \quad \zeta_{q_\ell}^{(\ell)} = 1 - \frac{1}{\left[\rho_{q_\ell}^{(\ell)}\right]^2}.$$

As $N \rightarrow +\infty$, if the ratio $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N$ is bounded and $\rho_{q_\ell}^{(\ell)} > 1$, then

$$\frac{1}{\sigma_N} \left[s_{q_\ell}^2(\mathbf{T}^{(\ell)}) - \mu_N^{(\ell)} \right] \xrightarrow{a.s.} \tilde{\xi}_{q_\ell}^{(\ell)} \quad \text{and} \quad \left\| \mathbf{X}^{(\ell)\top} \hat{\mathbf{u}}_{q_\ell}^{(\ell)} \right\|^2 \xrightarrow{a.s.} \zeta_{q_\ell}^{(\ell)}$$

where $\hat{\mathbf{u}}_{q_\ell}^{(\ell)}$ is the q_ℓ -th dominant left singular vector of $\mathbf{T}^{(\ell)}$.

Proof See Appendix C. ■

The first quantity defined in this theorem, $\rho_{q_\ell}^{(\ell)}$, should be understood as a signal-to-noise ratio (SNR). Indeed, the squared q_ℓ -th singular value of $\mathbf{P}^{(\ell)}$ (i.e., the q_ℓ -th eigenvalue of $\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}$), $s_{q_\ell}^2(\mathbf{P}^{(\ell)})$, measures the “strength” of the signal in its q_ℓ -th principal direction, whereas σ_N measures the spread of the noise, as seen in Theorem 5. The two quantities $\tilde{\xi}_{q_\ell}^{(\ell)}$ and $\zeta_{q_\ell}^{(\ell)}$ depend only on the value of this SNR and indicate respectively the position of an isolated eigenvalue in the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ and the alignment of the corresponding eigenvector with the sought signal. In fact, we observe a *phase transition* phenomenon: if the SNR is large enough, i.e., if $\rho_{q_\ell}^{(\ell)} > 1$, an eigenvalue of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ isolates itself from the semicircle¹⁵ around $\mu_N^{(\ell)} + \sigma_N \tilde{\xi}_{q_\ell}^{(\ell)}$. Moreover, recalling that $\mathcal{P} = \llbracket \mathcal{J}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)} \rrbracket$, the

14. We recall that the notation $s_i(\mathbf{A})$ denotes the i -th singular value of \mathbf{A} in non-increasing order.

15. Indeed, note that $\rho_{q_\ell}^{(\ell)} > 1 \implies \tilde{\xi}_{q_\ell}^{(\ell)} > 2$.

eigenvector associated with this isolated eigenvalue is aligned with the subspace spanned by $\mathbf{X}^{(\ell)}$, which is the ℓ -th singular subspace of \mathcal{P} . The quality of this alignment is given by $0 < \zeta_{q_\ell}^{(\ell)} \leq 1$.

Most importantly, this result reveals the *non-trivial regime* for the estimation of \mathcal{P} with a truncated MLSVD. Since $\sigma_N = \Theta(N^{\frac{d-2}{2}})$, it shows that $\|\mathcal{P}\|_F^2 = \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^2(\mathbf{P}^{(\ell)})$ must also be of the same order. Indeed, if $\|\mathcal{P}\|_F^2 \ll N^{\frac{d-2}{2}}$, then $\rho_{q_\ell}^{(\ell)} \rightarrow 0$, the SNR is too small and no signal can be recovered, whereas if $\|\mathcal{P}\|_F^2 \gg N^{\frac{d-2}{2}}$, then $\rho_{q_\ell}^{(\ell)} \rightarrow +\infty$, the SNR is very high and recovery of \mathcal{P} is easy. It is precisely between these two regimes, i.e., $\|\mathcal{P}\|_F = \Theta(N^{\frac{d-2}{4}})$, that the recovery is non-trivial. Note that this observation is in line with the results of Ben Arous et al. (2023) and Zhang and Xia (2018). In this non-trivial regime, the quantities $\zeta_{q_\ell}^{(\ell)}$ given in Theorem 8 precisely quantify how well the dominant eigenvectors of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ are aligned with the sought signal, i.e., the singular subspaces of \mathcal{P} . In section 3.2 below, this result is used to study the reconstruction performance of truncated MLSVD.

Remark 9 (Link with the spiked Wigner model) *The reader familiar with the spiked Wigner model may have recognized the expressions of $\tilde{\xi}_{q_\ell}^{(\ell)}$ and $\zeta_{q_\ell}^{(\ell)}$ given in Theorem 8. Indeed, given a symmetric $p \times p$ random matrix \mathbf{W} with i.i.d. (up to symmetry) $\mathcal{N}(0, \frac{1}{p})$ entries, the spectrum of $[\rho\mathbf{x}\mathbf{x}^\top + \mathbf{W}]$ with $\|\mathbf{x}\| = 1$ follows a semicircle distribution as $p \rightarrow +\infty$ with an isolated eigenvalue at $\rho + \frac{1}{\rho}$ if, and only if, $\rho > 1$ (Féral and Pécché, 2007; Edwards and Jones, 1976; Füredi and Komlós, 1981). Moreover, the corresponding eigenvector \mathbf{u} is such that $|\mathbf{x}^\top \mathbf{u}|^2 \rightarrow 1 - \frac{1}{\rho^2}$ almost surely as $p \rightarrow +\infty$ (Benaych-Georges and Nadakuditi, 2011). As discussed in Remark 7, up to a deterministic rescaling and shift, $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ asymptotically behaves like a spiked Wigner matrix.*

Theorem 8 is illustrated in Figure 3. In the first row, asymptotic positions of isolated eigenvalues $\mu_N^{(\ell)} + \sigma_N \tilde{\xi}_{q_\ell}^{(\ell)}$ are represented by the green dashed lines. In our experiment, \mathcal{P} has multilinear-rank (3, 4, 5). Hence 3, 4 and 5 isolated eigenvalues are expected in the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ for $\ell = 1, 2$ and 3 respectively. This is indeed the case for $\ell = 1$ and $\ell = 2$ but not $\ell = 3$ where there are only 4 spike eigenvalues. In fact, $s_5(\mathbf{P}^{(3)})$ is not “energetic enough” to extricate itself from the bulk of eigenvalues, i.e, the SNR $\rho_5^{(3)}$ is below the phase transition threshold. Hence the 5-th dominant left singular vector of $\mathbf{T}^{(\ell)}$ is not informative as it is not aligned with the 3-rd singular subspace of \mathcal{P} , spanned by $\mathbf{X}^{(3)}$. The second row of Figure 3 depicts the alignments of the spiked eigenvectors $\hat{\mathbf{u}}_{q_\ell}^{(\ell)}$ with the corresponding singular subspaces of \mathcal{P} as well as the asymptotic alignment given by Theorem 8 as a function of the position of the associated eigenvalue. It appears that the higher is the SNR $\rho_{q_\ell}^{(\ell)}$, the farther is the isolated eigenvalue from the bulk and the more is the corresponding eigenvector aligned with the span of $\mathbf{X}^{(\ell)}$. This assertion can be intuitively understood in terms of “energy” $s_{q_\ell}^2(\mathbf{P}^{(\ell)})$ of a given principal direction. More energy pushes the eigenvalue farther from the bulk and aligns the corresponding eigenvector with the corresponding singular subspace of \mathcal{P} .

3.2 Reconstruction Performance of Truncated MLSVD

Our random matrix results allow to accurately study the reconstruction performance of truncated MLSVD. Given a data tensor \mathcal{J} following the general spiked tensor model (2), we

consider its low-rank approximation $\hat{\mathcal{T}} = \llbracket \hat{\mathcal{G}}; \hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)} \rrbracket$ where $\hat{\mathbf{U}}^{(\ell)}$ is the $n_\ell \times r_\ell$ matrix whose columns are the r_ℓ dominant singular vectors of \mathcal{T} and $\hat{\mathcal{G}} = \mathcal{T}(\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)})$. $\hat{\mathcal{T}}$ is the projection of \mathcal{T} on its dominant singular subspaces, hence the name *truncated MLSVD* as it generalizes the truncated SVD of matrices. Given Model (2), the underlying signal estimated by $\hat{\mathcal{T}}$ is $\mathcal{P} = \llbracket \mathcal{H}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)} \rrbracket$. The reconstruction performance of $\hat{\mathcal{T}}$ hence depends on how well the *subspace*¹⁶ spanned by $\hat{\mathbf{U}}^{(\ell)}$ estimates the one spanned by $\mathbf{X}^{(\ell)}$.

Metrics between singular subspaces are often expressed in terms of *principal angles* (Björck and Golub, 1973; Stewart and Sun, 1990, II.4), which generalize the concept of angle between lines. Given two subspaces (here, $\text{Span } \mathbf{X}^{(\ell)}$ and $\text{Span } \hat{\mathbf{U}}^{(\ell)}$), one can define a set of mutual angles which are invariant under isometric transformation.

Definition 10 (Principal angles) *The principal angles $\theta_{q_\ell}^{(\ell)} \in [0, \frac{\pi}{2}]$ between the subspaces spanned by $\mathbf{X}^{(\ell)}$ and $\hat{\mathbf{U}}^{(\ell)}$ are recursively defined for $q_\ell = 1, \dots, r_\ell$ by*

$$\cos \theta_{q_\ell}^{(\ell)} = \mathbf{x}_{q_\ell}^\top \mathbf{u}_{q_\ell} \quad \text{with} \quad (\mathbf{x}_{q_\ell}, \mathbf{u}_{q_\ell}) \in \underset{\substack{(\mathbf{x}, \mathbf{u}) \in \text{Span } \mathbf{X}^{(\ell)} \times \text{Span } \hat{\mathbf{U}}^{(\ell)} \\ \mathbf{x}^\top \mathbf{x}_{q'_\ell} = 0, \mathbf{u}^\top \mathbf{u}_{q'_\ell} = 0, 1 \leq q'_\ell < q_\ell}}{\arg \max}} \mathbf{x}^\top \mathbf{u}.$$

Moreover, we have the following useful property.

Proposition 11 (Björck and Golub, 1973) *The q_ℓ -th singular value of $\mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)}$ in non-increasing order equals the cosine of the q_ℓ -th principal angle,*

$$s_{q_\ell}(\mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)}) = \cos \theta_{q_\ell}^{(\ell)}, \quad \ell \in [d], \quad q_\ell \in [r_\ell].$$

In fact, information about the alignment between the subspaces induced by $\mathbf{X}^{(\ell)}$ and $\hat{\mathbf{U}}^{(\ell)}$ are contained entirely in the $r_\ell \times r_\ell$ matrix $\mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)}$. Following Definition 10, we know from Theorem 8 that, as $N \rightarrow +\infty$, $\cos^2 \theta_{q_\ell}^{(\ell)} \rightarrow [\zeta_{q_\ell}^{(\ell)}]^+$ almost surely¹⁷, where we have used the handy notation $[\cdot]^+ = \max(\cdot, 0)$ since $\rho_{q_\ell}^{(\ell)} > 1 \iff \zeta_{q_\ell}^{(\ell)} > 0$. Hence, using Proposition 11,

$$\frac{1}{r_\ell} \left\| \mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)} \right\|_{\text{F}}^2 = \frac{1}{r_\ell} \sum_{q_\ell=1}^{r_\ell} \cos^2 \theta_{q_\ell}^{(\ell)} \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} \frac{1}{r_\ell} \sum_{q_\ell=1}^{r_\ell} [\zeta_{q_\ell}^{(\ell)}]^+. \quad (7)$$

Therefore, the quantity $\frac{1}{r_\ell} \left\| \mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)} \right\|_{\text{F}}^2 \in [0, 1]$ appears as a relevant measure of alignment between the singular subspaces of \mathcal{P} and $\hat{\mathcal{T}}$ and does not depend on the chosen orthonormal bases $\mathbf{X}^{(\ell)}$ and $\hat{\mathbf{U}}^{(\ell)}$. More details on metrics between subspaces can be found in Stewart and Sun (1990, II.4).

In Figure 2 are represented the alignments between the singular subspaces of $\hat{\mathcal{T}}$ and $\mathcal{P} = \sqrt{\omega} \mathcal{P}_\circ$ with $\|\mathcal{P}_\circ\|^2 = \sigma_N$ as a function of the signal-to-noise ratio ω . The fact that $\|\mathcal{P}_\circ\|^2 = \sigma_N$ ensures that the estimation problem is *non-trivial* (neither too easy nor too

16. The object of importance is indeed the subspace and not the matrix $\hat{\mathbf{U}}^{(\ell)}$ since any other matrix $\hat{\mathbf{U}}^{(\ell)} \mathbf{O}^{(\ell)}$, with $\mathbf{O}^{(\ell)}$ an $r_\ell \times r_\ell$ orthogonal matrix, would span the same subspace and therefore give the same approximation.

17. Indeed, since $\zeta_1^{(\ell)} \geq \dots \geq \zeta_{r_\ell}^{(\ell)}$, observe that, in Definition 10, $|\mathbf{x}_{q_\ell}^\top \mathbf{u}_{q_\ell}|^2$ is asymptotically bounded by $[\zeta_{q_\ell}^{(\ell)}]^+$.

hard) as $\rho_{q_\ell}^{(\ell)} = \Theta(1)$. Plain curves are the alignments given by Theorem 8 as $N \rightarrow +\infty$ (right-hand side of Equation (7)) whereas dotted curves are simulation results at finite N (left-hand side of Equation (7)). In this setting, \mathcal{P}_\circ has multilinear-rank $(3, 4, 5)$. Hence, its “energy” $\|\mathcal{P}_\circ\|^2$ is spread among 3, 4 and 5 directions along modes 1, 2 and 3 respectively. Each break in the plain curves correspond to a value of ω such that $\rho_{q_\ell}^{(\ell)} = 1$, that is, $\omega = \sigma_N/s_{q_\ell}^2(\mathcal{P}_\circ^{(\ell)})$. In other words, there are r_ℓ phase transitions along mode ℓ whose positions depend on the singular values of $\mathcal{P}_\circ^{(\ell)}$. If ω is too small (here, $\omega \lesssim 2$), truncated MLSVD fails to recover any direction of the singular subspaces of \mathcal{P}_\circ . As ω passes the first phase transition (here, at $\omega = \sigma_N/s_1^2(\mathcal{P}_\circ^{(1)}) \approx 2$), a first principal direction is partially reconstructed. Then, more and more phase transitions occur, corresponding to more and more principal directions being recovered as ω grows. Simultaneously, the reconstruction of previous directions keeps improving. Eventually, as ω is large, subspace alignments approach 1, indicating that truncated MLSVD accurately recovers the singular subspaces of \mathcal{P}_\circ .

The reader has not missed the dashed lines in Figure 2 showing much better subspace alignments than truncated MLSVD. They result from the numerical estimation of the *best* rank- $(3, 4, 5)$ approximation of \mathcal{J} with the HOOI algorithm, which is discussed in the next section.

Remark 12 (Reconstruction of \mathcal{H} of \mathcal{J}) *Guarantees on the recovery of the core tensor \mathcal{H} can be deduced from Theorem 8 as well. Without loss of generality, we can assume that $\mathbf{X}^{(\ell)\top} \hat{\mathbf{U}}^{(\ell)}$ is, up to an almost-surely vanishing additive term, a diagonal matrix with entries $\zeta_1^{(\ell)}, \dots, \zeta_{r_\ell}^{(\ell)}$ (otherwise replace $\hat{\mathbf{U}}^{(\ell)}$ by $\hat{\mathbf{U}}^{(\ell)} \mathbf{O}^{(\ell)}$ for a well-chosen orthogonal matrix $\mathbf{O}^{(\ell)}$). Then, $\hat{\mathcal{G}} = \mathcal{J}(\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)})$ is the corresponding estimator of \mathcal{H} and*

$$\hat{\mathcal{G}} = \hat{\mathcal{H}} + \frac{1}{\sqrt{N}} \mathcal{N}(\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)}) \quad \text{with} \quad \hat{\mathcal{H}} = \left[\left[\mathcal{H}; \hat{\mathbf{U}}^{(1)\top} \mathbf{X}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)\top} \mathbf{X}^{(d)} \right] \right].$$

We will see in Lemma 13 below that $\|\frac{1}{\sqrt{N}} \mathcal{N}(\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)})\|_F = \mathcal{O}(1)$ as $N \rightarrow +\infty$. On the other hand, we know that $\|\mathcal{H}\|_F = \|\mathcal{P}\|_F = \Theta(N^{\frac{d-2}{4}}) \gg \mathcal{O}(1)$ as soon as $d \geq 3$ and the entries of $\hat{\mathcal{H}}$ are proportional to those of \mathcal{H} :

$$\hat{\mathcal{H}}_{q_1, \dots, q_d} = \mathcal{H}_{q_1, \dots, q_d} \prod_{\ell=1}^d \zeta_{q_\ell}^{(\ell)} + \epsilon_{q_1, \dots, q_d}, \quad \ell \in [d], \quad q_\ell \in [r_\ell],$$

up to an almost-surely vanishing additive term $\epsilon_{q_1, \dots, q_d}$ as $N \rightarrow +\infty$.

Moreover, regarding the reconstruction of \mathcal{J} , we know from De Lathauwer et al. (2000b, Property 10) that

$$\left\| \mathcal{J} - \hat{\mathcal{J}} \right\|_F^2 \leq \sum_{\ell=1}^d \sum_{i_\ell=r_\ell+1}^{n_\ell} s_{i_\ell}^2(\mathbf{T}^{(\ell)}) = \sum_{\ell=1}^d \left(\|\mathcal{J}\|_F^2 - \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^2(\mathbf{T}^{(\ell)}) \right)$$

and the asymptotic behavior of $s_{q_\ell}^2(\mathbf{T}^{(\ell)})$ is given by Theorem 8.

4 Numerical Estimation of the Best Low-Multilinear-Rank Approximation

In search of an efficient estimator of the planted signal \mathcal{P} , one naturally considers the best low-multilinear-rank approximation of \mathcal{J} , that is, a solution to Problem (1). As explained in the introduction, this is NP-hard in general but numerical schemes can compute it in polynomial time above the computational threshold (Montanari and Richard, 2014; Zhang and Xia, 2018). In this section, we examine the most standard of these numerical schemes, namely the Higher Order Orthogonal Iteration (HOOI) algorithm (De Lathauwer et al., 2000a; Kroonenberg and de Leeuw, 1980; Kapteyn et al., 1986), and discuss the numerical difficulties faced in the computation of a solution to Problem (1).

4.1 Higher-Order Orthogonal Iteration

Following De Lathauwer et al. (2000a, Theorem 4.2), the maximum likelihood estimation formulated in Problem (1) is equivalent to

$$\left(\mathbf{U}_\star^{(1)}, \dots, \mathbf{U}_\star^{(d)}\right) \in \arg \max_{\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell}), \ell \in [d]} \frac{1}{2} \left\| \mathcal{J}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)}) \right\|_F^2 \quad (8)$$

where $V_{r_\ell}(\mathbb{R}^{n_\ell}) = \{\mathbf{U}^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell} \mid \mathbf{U}^{(\ell)\top} \mathbf{U}^{(\ell)} = \mathbf{I}_{r_\ell}\}$ is the set of $r_\ell \times n_\ell$ matrices with orthonormal columns, known as the *Stiefel manifold* (Chikuse, 2003; Absil et al., 2009). Then, since the Frobenius norm of $\mathcal{J}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)})$ is equal to the Frobenius norm of any of its unfoldings,

$$\left\| \mathcal{J}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)}) \right\|_F = \left\| \mathbf{U}^{(\ell)\top} \mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}^{(\ell')} \right\|_F, \quad \ell \in [d].$$

And we see from Problem (8) that $\mathbf{U}_\star^{(\ell)}$ is the matrix gathering the r_ℓ dominant left singular vectors of $\mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_\star^{(\ell')}$. This is precisely what motivates the HOOI algorithm presented in Algorithm 1. It performs fixed-point iterations to compute a solution $\mathbf{U}_\star^{(1)}, \dots, \mathbf{U}_\star^{(d)}$ satisfying the previous property¹⁸.

Algorithm 1: Higher-Order Orthogonal Iteration (De Lathauwer et al., 2000a)

```

for  $\ell = 1, \dots, d$  do  $\mathbf{U}_0^{(\ell)} \leftarrow r_\ell$  dominant left singular vectors of  $\mathbf{T}^{(\ell)}$ 
repeat
  | for  $\ell = 1, \dots, d$  do  $\mathbf{U}_{t+1}^{(\ell)} \leftarrow r_\ell$  dominant left singular vectors of  $\mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_t^{(\ell')}$ 
until convergence at  $t = T$ 
 $\mathcal{G}_{\text{HOOI}} \leftarrow \mathcal{J}(\mathbf{U}_T^{(1)}, \dots, \mathbf{U}_T^{(d)})$ 

```

The HOOI algorithm is initialized with $\mathbf{U}_0^{(1)}, \dots, \mathbf{U}_0^{(d)}$, the truncated MLSVD¹⁹ of \mathcal{J} . Given the results of Section 3, this is indeed a very good and easily computable first guess.

18. In fact, this property corresponds to first-order optimality conditions of Problem (8), with the squared singular values as Lagrange multipliers.

19. Consistently with the notations of Section 3, this is $\hat{\mathbf{U}}^{(1)}, \dots, \hat{\mathbf{U}}^{(d)}$.

Then, fixed-point iterations are repeated in order to find a solution $\mathbf{U}_\star^{(1)}, \dots, \mathbf{U}_\star^{(d)}$ such that $\mathbf{U}_\star^{(\ell)}$ spans the left r_ℓ -dimensional dominant singular subspace of $\mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_\star^{(\ell')}$ for all $\ell \in [d]$, which corresponds to a solution to Problem (8). In practice, the stopping criterion can be chosen as a negligible change in the norm of the estimated core tensor, $\|\mathfrak{T}(\mathbf{U}_t^{(1)}, \dots, \mathbf{U}_t^{(d)})\|_F$.

Xu (2018) showed that the convergence of this algorithm towards a local minimum of Problem (8) is guaranteed as long as its initialization is *sufficiently close* to this local minimum. In light of our previous results, we can provide further insight into this “sufficiently close” property. Indeed, in Theorem 14 below, we show that an initialization with non-zero alignment with the signal \mathcal{P} is sufficient to ensure that the HOOI algorithm perfectly reconstructs it *after a single iteration*.

Before introducing Theorem 14, we formulate an important preliminary result which essentially states that, given $\mathbf{A}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$, $\ell \in [d]$, the quantity $\frac{1}{\sqrt{N}} \|\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})\|_F$ is almost surely bounded as $N \rightarrow +\infty$.

Lemma 13 *With probability at least $1 - \delta$,*

$$\sup_{\mathbf{A}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell}), \ell \in [d]} \left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \right\|_F^2 \leq 16 \left[\left(\sum_{\ell=1}^d r_\ell \left(n_\ell - \frac{r_\ell + 1}{2} \right) \right) \log \frac{Cd}{\log \frac{3}{2}} + \log \left(\frac{1}{\delta} \max \left(1, e^{\frac{1}{2} \prod_{\ell=1}^d r_{\ell-1}} \right) \right) \right]$$

where $C > 0$ is a universal constant.

Proof See Appendix D. ■

This result is crucial to handle the behavior of the noise in our analysis of Algorithm 1 (Appendix E), which leads to the following result on the alignment between the singular subspaces of the signal \mathcal{P} (spanned by $\mathbf{X}^{(\ell)}$) and those estimated from the observation \mathfrak{T} after the first iteration of HOOI (spanned by $\mathbf{U}_1^{(\ell)}$).

Theorem 14 (Asymptotic optimality of HOOI) *As $N \rightarrow +\infty$, if $\|\mathcal{P}\|_F \gg 1$ and*

$$\min_{\ell \in [d], q_\ell \in [r_\ell]} \left\| \mathcal{P}(\mathbf{U}_0^{(1)}, \dots, \mathbf{x}_{q_\ell}^{(\ell)}, \dots, \mathbf{U}_0^{(d)}) \right\|_F \stackrel{\text{def}}{=} L_N \gg \|\mathcal{P}\|_F^{1/2}$$

where $\mathbf{x}_{q_\ell}^{(\ell)}$ is the q_ℓ -th column of $\mathbf{X}^{(\ell)}$, then,

$$\frac{1}{r_\ell} \left\| \mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)} \right\|_F^2 = 1 + \mathcal{O} \left(\frac{\|\mathcal{P}\|_F}{L_N^2} \right) \quad \text{almost surely.}$$

Proof See Appendix E. ■

It is important to carefully understand the assumptions of Theorem 14. Firstly, it assumes that $\|\mathcal{P}\|_F \gg 1$ as $N \rightarrow +\infty$, that is, the signal is not necessarily in the non-trivial regime $\Theta(N^{\frac{d-2}{4}})$ but can be smaller or bigger as long as $\|\mathcal{P}\|_F \rightarrow +\infty$, regardless its speed.

Then, the second assumption $\|\mathcal{P}(\mathbf{U}_0^{(1)}, \dots, \mathbf{x}_{q_\ell}^{(\ell)}, \dots, \mathbf{U}_0^{(d)})\|_{\mathbb{F}} \gg \|\mathcal{P}\|_{\mathbb{F}}^{1/2}$ means that each principal directions of the ℓ -th singular subspace are sufficiently preserved after contraction on $\{\mathbf{U}_0^{(\ell')}\}_{\ell' \neq \ell}$. When these assumptions are verified, Theorem 14 states that the matrices $\mathbf{U}_1^{(1)}, \dots, \mathbf{U}_1^{(d)}$ computed after the *first* iteration of HOOI *perfectly* reconstruct the singular subspaces of the sought signal \mathcal{P} as the dimensions of the tensor, n_1, \dots, n_d , grow large. More formally, as $N \rightarrow +\infty$, the alignment $\frac{1}{r_\ell} \|\mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)}\|_{\mathbb{F}}^2$ approaches 1 almost surely. Furthermore, the speed of this convergence behaves like $\|\mathcal{P}\|_{\mathbb{F}}/L_N^2$.

Theorem 14 does not assume a particular choice of initialization $\mathbf{U}_0^{(1)}, \dots, \mathbf{U}_0^{(d)}$ and gives a sufficient condition for it to ensure the convergence of the algorithm. Nevertheless, as it is presented in Algorithm 1, truncated MLSVD is a standard choice of initialization. In this case, the assumption $L_N = \Theta(\|\mathcal{P}\|)$ is verified as soon as enough principal directions are recovered. According to Theorem 8, this is only possible if $\|\mathcal{P}\|_{\mathbb{F}} \geq \sqrt{\sigma_N} = \Theta(N^{\frac{d-2}{4}})$ since a *necessary* condition is $\rho_1^{(\ell)} > 1$ for all $\ell \in [d]$, while a *sufficient* condition is $\rho_{q_\ell}^{(\ell)} > 1$ for all $\ell \in [d]$ and $q_\ell \in [d]$. In other words, convergence at speed $\|\mathcal{P}\|_{\mathbb{F}}^{-1}$ as assured above a critical signal-to-noise ratio lying between the first and the last phase transition of each mode, and which depends on the particular structure of the core tensor \mathcal{H} . Yet, in most cases, this happens quite early, right after the first phase transitions, see for example Figure 2.

We emphasize the fact that the assumption of Theorem 14 can already be verified as soon as $\min_\ell \rho_1^{(\ell)} > 1$. That is, there is no need for all the principal directions to be reconstructed at initialization. In fact, it could very well be that $\max_\ell \rho_2^{(\ell)} < 1$. If the singular subspaces of \mathcal{P} are sufficiently preserved with the initialization $\mathbf{U}_0^{(1)}, \dots, \mathbf{U}_0^{(d)}$ — i.e., if $L_N \gg \|\mathcal{P}\|^{1/2}$ —, then the other principal directions still emerge after the first iteration.

In the simpler rank-1 case, these technical considerations vanish and we recover the result of Feldman and Donoho (2023, Theorem 4.2): if $\mathcal{P} = \beta_N \otimes_{\ell \in [d]} \mathbf{x}^{(\ell)}$ then a *necessary and sufficient* condition for $L_N = \Theta(\|\mathcal{P}\|_{\mathbb{F}})$ is simply $\beta_N^2 > \sigma_N$. Indeed, $\rho_1^{(\ell)} = \beta_N^2/\sigma_N$ for all $\ell \in [d]$. Moreover, when this assumption is verified, Theorem 14 ensures the asymptotic exact reconstruction of $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(d)}$ in a single iteration with a $\|\mathcal{P}\|_{\mathbb{F}}^{-1} = \beta_N^{-1}$ speed of convergence.

Remark 15 (Practical implications) *In practice, one should still run several iterations of Algorithm 1 until a certain stopping criterion is verified as this effectively improves the final estimate and converges to a solution to the maximum likelihood estimation (1) (Xu, 2018). Theorem 14 states that the reconstruction performance of HOOI after the first iteration increases as we consider larger tensors, until it reaches perfect recovery in the large N limit. In other words, the number of iterations required to achieve a specific level of accuracy in maximum likelihood estimation tends to 1 as $N \rightarrow +\infty$.*

Theorem 14 is illustrated in Figure 4. As a function of N — the size of the tensor — we represent the subspace alignments observed at initialization and after the first iteration for a fixed signal-to-noise ratio $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N = 10$. The left panel compares the observed alignments achieved with truncated MLSVD (initialization of Algorithm 1) with the asymptotic alignments predicted by Theorem 8. As N grows, the observed alignments remain around their asymptotic values, with only a decrease in variance. The middle panel presents the alignments after the first iteration. Here, as N increases, we observe an increase in the

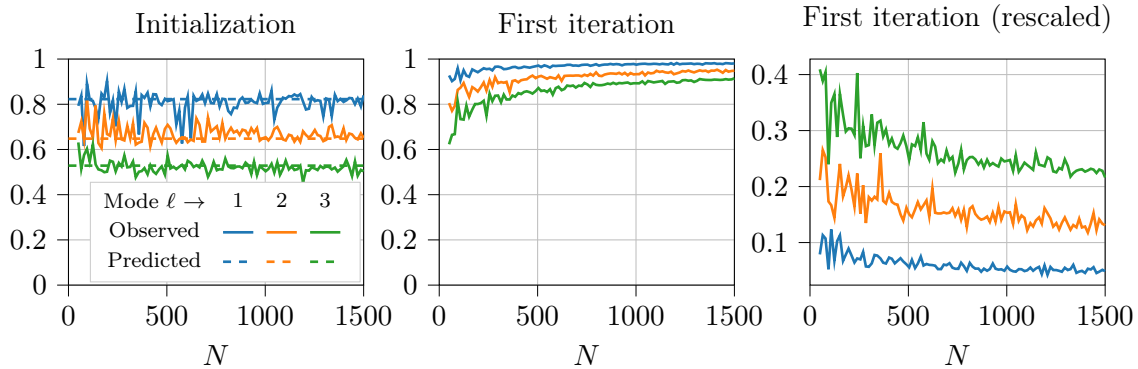


Figure 4: Alignments between singular subspaces of the observation $\mathcal{J} = \mathcal{P} + \frac{1}{\sqrt{N}}\mathcal{N}$ and of the signal \mathcal{P} , with $\|\mathcal{P}\|_F^2/\sigma_N = 10$, at initialization of Algorithm 1 (i.e., truncated MLSVD) and after the first iteration, as a function of the size of the tensor given by the parameter N . **Left:** $\frac{1}{r_\ell} \|\mathbf{X}^{(\ell)\top} \mathbf{U}_0^{(\ell)}\|_F^2$. **Middle:** $\frac{1}{r_\ell} \|\mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)}\|_F^2$. **Right:** $(1 - \frac{1}{r_\ell} \|\mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)}\|_F^2) \times \sqrt{\sigma_N}$. **Experimental setting:** $d = 3$, $(\frac{n_1}{N}, \frac{n_2}{N}, \frac{n_3}{N}) = (\frac{1}{6}, \frac{2}{6}, \frac{3}{6})$, $N = n_1 + n_2 + n_3$ and $(r_1, r_2, r_3) = (3, 4, 5)$.

values of the alignments, which approach 1, consistently with Theorem 14. This is specified in the right panel where the value $(1 - \frac{1}{r_\ell} \|\mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)}\|_F^2) \times \sqrt{\sigma_N}$ is plotted. According to Theorem 14, this value should be $\mathcal{O}(1)$ since $L_N = \Theta(\|\mathcal{P}\|_F) = \Theta(\sqrt{\sigma_N})$ here. The observed behavior confirms the $\|\mathcal{P}\|_F^{-1}$ speed of convergence asserted in Theorem 14.

4.2 Discussion on Signal Reconstructibility

Our results on truncated MLSVD (Section 3) and HOOI (Section 4.1) bring insight into the computational-to-statistical gap observed in the low-multilinear-rank approximation problem. Truncated MLSVD can only work efficiently if $\|\mathcal{P}\|_F$ is at least $\Theta(N^{\frac{d-2}{4}})$ and its reconstruction performance has been discussed in Section 3.2. However, Theorem 14 suggests that it is possible to perfectly reconstruct the signal \mathcal{P} from the observation \mathcal{J} as long as $\|\mathcal{P}\|_F \gg 1$ and HOOI is accurately initialized. Yet, it is known that, without prior information on \mathcal{P} , maximum likelihood estimation (1) is NP-hard below the $\Theta(N^{\frac{d-2}{4}})$ computational threshold (Zhang and Xia, 2018), which lies precisely in the non-trivial regime of truncated MLSVD.

In fact, what can be understood from Theorem 14 is that it suffices to have an initialization $\mathbf{U}_0^{(\ell)}$, $\ell \in [d]$, slightly aligned with the underlying signal \mathcal{P} to be in the right basin of attraction and allow the convergence of Algorithm 1 towards a solution to Problem (1). This complements the results of Xu (2018) on the conditions of convergence of HOOI. Furthermore, a solution to Problem (1) is aligned with \mathcal{P} as soon as $\|\mathcal{P}\|_F = \Theta(1)$ (Ben Arous et al., 2019; Jagannath et al., 2020; Zhang and Xia, 2018). Hence, as the HOOI algorithm is meant to compute a maximum likelihood estimator, with the assumption $\|\mathcal{P}\|_F \gg 1$ made in The-

orem 14, it is expected that these iterations allow to perfectly recover the signal asymptotically. Maximum likelihood estimation is indeed (theoretically) trivial if $\|\mathcal{P}\|_{\text{F}} \rightarrow +\infty$. It is more surprising however that this already happens at the first iteration.

As said previously, the choice of initialization $\mathbf{U}_0^{(\ell)}$, $\ell \in [d]$, does not matter in Theorem 14. In fact, without prior information, a truncated MLSVD is the best choice as it allows to partially reconstruct the signal at the $\Theta(N^{\frac{d-2}{4}})$ computational threshold. Nevertheless, had one prior information allowing such a reconstruction in the regime $1 \ll \mathcal{P} \ll N^{\frac{d-2}{4}}$ — where truncated MLSVD would not be fruitful —, HOOI would still be able to perfectly reconstruct the signal \mathcal{P} given this initialization.

Hence, HOOI initialized with a truncated MLSVD, as it is presented in Algorithm 1, allows to numerically compute a maximum likelihood estimator (solution to Problem (1)) but only above the phase transition of truncated MLSVD. Indeed, its initialization plays a determining role: it must place $\mathbf{U}_0^{(\ell)}$, $\ell \in [d]$, in the right basin of attraction, which, without prior information, is only possible above the $\Theta(N^{\frac{d-2}{4}})$ computational threshold.

Finally, we highlight the fact that these results concern the large N limit. In practice, it makes no sense to talk about $\Theta(1)$ or $\Theta(N^{\frac{d-2}{4}})$ regimes at finite N . Figure 2 also depicts the subspace alignments achieved with the maximum likelihood estimator computed with Algorithm 1 on $\mathcal{J} = \sqrt{\omega}\mathcal{P}_\circ + \frac{1}{\sqrt{N}}\mathbf{N}$. Although $\|\mathcal{P}_\circ\|_{\text{F}} = \sqrt{\sigma_N} = \Theta(N^{\frac{d-2}{4}})$, HOOI does not achieve perfect recovery of \mathcal{P} as one might expect from Theorem 14 (even if several iterations were performed here). In fact, at finite N , perfect recovery is not feasible. But, as N grows, the dashed line would approach 1 above the (computational) phase transition determined by the truncated MLSVD and stay close to 0 below.

5 Concluding Remarks

The analysis presented in this work yield theoretical and practical insight into the estimation of a low-rank signal from an observation $\mathcal{J} = \mathcal{P} + \frac{1}{\sqrt{N}}\mathbf{N}$ following the most general spiked tensor model. While Zhang and Xia (2018) gave a general overview of the different regimes governing the estimation of \mathcal{P} with a low-multilinear-rank approximation of \mathcal{J} — thereby confirming the existence of a computational-to-statistical gap —, our results shed light on the non-trivial aspects at stake around the $\Theta(N^{\frac{d-2}{4}})$ computational threshold. This is of particular importance as practical applications lie in this non-trivial regime where signal and noise have the same magnitude and must be decoupled. In particular, truncated MLSVD and HOOI are very standard and efficient algorithms to compute low-multilinear-rank approximations. Performances of the latter rely strongly on the quality of its initialization, which is usually performed with a truncated MLSVD in the absence of prior information. This approach allows the detection of the underlying signal as early as the computational threshold contrary to other methods such as AMP or tensor power iteration, which are efficient above a $\Theta(N^{\frac{d-1}{2}})$ algorithmic threshold (Montanari and Richard, 2014).

Relying on standard tools and methods from the theory of large random matrices, we have characterized the spectral behavior of the unfoldings of \mathcal{J} in the large N limit. Specifically, our first main result shows that, when properly centered and scaled, the eigenvalues of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ asymptotically follow a semicircle distribution. The rescaling exhibits their mean $\mu_N^{(\ell)} = \Theta(N^{d-2})$ and a quantity $\sigma_N = \Theta(N^{\frac{d-2}{2}})$ governing their spread. From our denois-

ing perspective, σ_N indicates the *strength* of the noise. Indeed, while the global behavior of the eigenvalues is controlled by the noise, the addition of a low-rank signal causes the presence of a finite number of eigenvalues outside the limiting semicircle distribution with corresponding eigenvectors aligned with the singular subspaces of the sought signal \mathcal{P} . Yet, the existence of these outlier eigenvalues hinges on the values of the signal-to-noise ratios $\rho_{q_\ell}^{(\ell)} = s_{q_\ell}^2(\mathbf{P}^{(\ell)})/\sigma_N$, manifesting a BBP phase transition phenomenon. When they exist, the positions of these isolated eigenvalues and the quality of the corresponding alignments are completely determined by $\rho_{q_\ell}^{(\ell)}$. These results justify the use of a truncated MLSVD to estimate \mathcal{P} from the observation \mathcal{J} and allow the precise characterization of the achievable reconstruction performances in the non-trivial regime, i.e., close to the computational threshold. In particular, we have seen that each singular value of $\mathbf{P}^{(\ell)}$ determines the position of a phase transition corresponding to the detectability of the corresponding principal direction.

Although truncated MLSVD does not yield the best low-multilinear-rank approximation — i.e., a maximum likelihood solution —, it serves as an excellent initialization for the HOOI algorithm, which converges to such an estimator if it is initialized *sufficiently close* to it (Xu, 2018). In fact, we precise this last assertion by showing that, as long as the initialization preserves the singular subspaces of \mathcal{P} in a sense precised in Theorem 14, HOOI converges to a maximum likelihood solution in a number of iterations which tends to 1 as $N \rightarrow +\infty$. Hence, when it is initialized with a truncated MLSVD, it shares the same phase transition, whose position depends on the singular values $s_{q_\ell}^{(\ell)}(\mathbf{P}^{(\ell)})$ of the unfoldings of \mathcal{P} . Yet, given prior information, HOOI can still reconstruct the maximum likelihood solution below the computational threshold $\|\mathcal{P}\|_F = \Theta(N^{\frac{d-2}{4}})$, where its success depends entirely on the quality of its initialization.

This work gives a comprehensive understanding of the low-multilinear-rank approximation problem near the computational threshold, which has both practical and theoretical implications. Besides, from a theoretical perspective, the behavior of the maximum likelihood estimator is still unclear near the statistical threshold — that is, in the regime where $\|\mathcal{P}\|_F = \Theta(1)$. Several works have studied the rank-1 symmetric case (Ben Arous et al., 2019; Jagannath et al., 2020) and the approach developed by Seddik et al. (2022) in their analysis of the rank-1 *asymmetric* case may be an attractive direction to consider. Relying solely on classical tools from random matrix theory, they bring the study of the best rank-1 tensor approximation down to that of a structured matrix defined from the contractions of the data tensor on its dominant singular vectors. Extending this procedure to our general spiked tensor model (2) presents no conceptual difficulty, despite being computationally cumbersome due to the multiple dimensions of the singular subspaces. It is an interesting line of investigation to refine our understanding of the statistical limits of spiked tensor models.

Acknowledgments and Disclosure of Funding

This work is supported by the MIAI LargeDATA Chair at Université Grenoble Alpes.

Appendix A. Proof of Theorem 5

The resolvent of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ is defined for all $z \in \mathbb{C} \setminus \text{Sp}\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ as

$$\mathbf{Q}^{(\ell)}(z) \stackrel{\text{def}}{=} \left(\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - z\mathbf{I}_{n_\ell} \right)^{-1}.$$

We will often drop the dependence in z to simplify notations.

Since $\mathbf{Q}^{(\ell)-1}\mathbf{Q}^{(\ell)} = \mathbf{I}_{n_\ell}$ and $\mathbf{T}^{(\ell)} = \mathbf{P}^{(\ell)} + \frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}$, we have,

$$\mathbf{P}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} + \frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} - z\mathbf{Q}^{(\ell)} = \mathbf{I}_{n_\ell}. \quad (9)$$

A.1 Expressions with Stein's lemma

Using Stein's lemma (Lemma 4), we find the following expressions.

$$\mathbb{E} \left[\mathbf{P}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right] = \mathbb{E} \left[\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right] - \frac{1}{N}\mathbb{E} \left[\mathbf{P}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \text{Tr}\mathbf{Q}^{(\ell)} + \mathbf{P}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)2} \right], \quad (10)$$

$$\mathbb{E} \left[\mathbf{N}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right] = \frac{\prod_{\ell' \neq \ell} n_{\ell'}}{\sqrt{N}}\mathbb{E} \left[\mathbf{Q}^{(\ell)} \right] - \frac{1}{\sqrt{N}}\mathbb{E} \left[(n_\ell + 1)\mathbf{Q}^{(\ell)} - z \left(\mathbf{Q}^{(\ell)2} + \mathbf{Q}^{(\ell)} \text{Tr}\mathbf{Q}^{(\ell)} \right) \right]. \quad (11)$$

Derivative of $\mathbf{Q}^{(\ell)}$ Firstly, we need to show that

$$\frac{\partial \mathbf{Q}_{a,b}^{(\ell)}}{\partial N_{c,d}^{(\ell)}} = -\frac{1}{\sqrt{N}} \left(\mathbf{Q}_{a,c}^{(\ell)} \left[\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right]_{d,b} + \left[\mathbf{Q}^{(\ell)}\mathbf{T}^{(\ell)} \right]_{a,d} \mathbf{Q}_{c,b}^{(\ell)} \right). \quad (12)$$

Indeed, using the fact that $\partial \mathbf{Q}^{(\ell)} = -\mathbf{Q}^{(\ell)}\partial(\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top})\mathbf{Q}^{(\ell)}$, we have,

$$\frac{\partial \mathbf{Q}_{a,b}^{(\ell)}}{\partial N_{c,d}^{(\ell)}} = - \left[\mathbf{Q}^{(\ell)} \frac{\partial \mathbf{T}^{(\ell)}}{\partial N_{c,d}^{(\ell)}} \mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right]_{a,b} - \left[\mathbf{Q}^{(\ell)}\mathbf{T}^{(\ell)} \frac{\partial \mathbf{T}^{(\ell)\top}}{\partial N_{c,d}^{(\ell)}} \mathbf{Q}^{(\ell)} \right]_{a,b}$$

and, since $\mathbf{T}^{(\ell)} = \mathbf{P}^{(\ell)} + \frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}$,

$$\begin{aligned} \left[\mathbf{Q}^{(\ell)} \frac{\partial \mathbf{T}^{(\ell)}}{\partial N_{c,d}^{(\ell)}} \mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right]_{a,b} &= \frac{1}{\sqrt{N}} \mathbf{Q}_{a,c}^{(\ell)} \left[\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)} \right]_{d,b}, \\ \left[\mathbf{Q}^{(\ell)}\mathbf{T}^{(\ell)} \frac{\partial \mathbf{T}^{(\ell)\top}}{\partial N_{c,d}^{(\ell)}} \mathbf{Q}^{(\ell)} \right]_{a,b} &= \frac{1}{\sqrt{N}} \left[\mathbf{Q}^{(\ell)}\mathbf{T}^{(\ell)} \right]_{a,d} \mathbf{Q}_{c,b}^{(\ell)}. \end{aligned}$$

Proof of Equation (10) Since $\mathbf{T}^{(\ell)} = \mathbf{P}^{(\ell)} + \frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}$, we have,

$$\mathbb{E}[\mathbf{P}^{(\ell)}\mathbf{T}^{(\ell)\top}\mathbf{Q}^{(\ell)}] = \mathbb{E}[\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}\mathbf{Q}^{(\ell)}] + \frac{1}{\sqrt{N}}\mathbb{E}[\mathbf{P}^{(\ell)}\mathbf{N}^{(\ell)\top}\mathbf{Q}^{(\ell)}].$$

To deal with the rightmost term, we successively use Stein's lemma (Lemma 4) and Equation (12).

$$\begin{aligned}
 \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{N}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]_{i,j} &= \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[P_{i,k}^{(\ell)} N_{l,k}^{(\ell)} Q_{l,j}^{(\ell)} \right] \\
 &= \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[P_{i,k}^{(\ell)} \frac{\partial Q_{l,j}^{(\ell)}}{\partial N_{l,k}^{(\ell)}} \right] \\
 &= -\frac{1}{\sqrt{N}} \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[P_{i,k}^{(\ell)} Q_{l,l}^{(\ell)} \left[\mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]_{k,j} \right] \\
 &\quad - \frac{1}{\sqrt{N}} \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[P_{i,k}^{(\ell)} \left[\mathbf{Q}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{l,k} Q_{l,j}^{(\ell)} \right] \\
 &= -\frac{1}{\sqrt{N}} \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \text{Tr} \mathbf{Q}^{(\ell)} + \mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)2} \right]_{i,j}.
 \end{aligned}$$

Proof of Equation (11) We proceed similarly with Stein's lemma (Lemma 4) and Equation (12).

$$\begin{aligned}
 \mathbb{E} \left[\mathbf{N}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]_{i,j} &= \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[N_{i,k}^{(\ell)} T_{l,k}^{(\ell)} Q_{l,j}^{(\ell)} \right] \\
 &= \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[\frac{\partial T_{l,k}^{(\ell)}}{\partial N_{i,k}^{(\ell)}} Q_{l,j}^{(\ell)} + T_{l,k}^{(\ell)} \frac{\partial Q_{l,j}^{(\ell)}}{\partial N_{i,k}^{(\ell)}} \right] \\
 &= \frac{\prod_{\ell' \neq \ell} n_{\ell'}}{\sqrt{N}} \mathbb{E} \left[\mathbf{Q}^{(\ell)} \right]_{i,j} - \frac{1}{\sqrt{N}} \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[T_{l,k}^{(\ell)} Q_{l,i}^{(\ell)} \left[\mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]_{k,j} \right] \\
 &\quad - \frac{1}{\sqrt{N}} \sum_{k=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \sum_{l=1}^{n_{\ell}} \mathbb{E} \left[T_{l,k}^{(\ell)} \left[\mathbf{Q}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{l,k} Q_{i,j}^{(\ell)} \right] \\
 &= \frac{\prod_{\ell' \neq \ell} n_{\ell'}}{\sqrt{N}} \mathbb{E} \left[\mathbf{Q}^{(\ell)} \right]_{i,j} - \frac{1}{\sqrt{N}} \mathbb{E} \left[\mathbf{Q}^{(\ell)} \mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} + \mathbf{Q}^{(\ell)} \text{Tr} \mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]_{i,j}.
 \end{aligned}$$

Since $\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} - z \mathbf{Q}^{(\ell)} = \mathbf{I}_{n_{\ell}}$, we find the result stated in Equation (11).

A.2 Deterministic Equivalent

Taking the expectation of Equation (9) and injecting Equation (11) yields

$$\begin{aligned} \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right] + \frac{\prod_{\ell' \neq \ell} n_{\ell'}}{N} \mathbb{E} \left[\mathbf{Q}^{(\ell)} \right] \\ - \frac{1}{N} \mathbb{E} \left[(n_{\ell} + 1) \mathbf{Q}^{(\ell)} - z \left(\mathbf{Q}^{(\ell)2} + \mathbf{Q}^{(\ell)} \text{Tr} \mathbf{Q}^{(\ell)} \right) \right] - z \mathbb{E} \left[\mathbf{Q}^{(\ell)} \right] = \mathbf{I}_{n_{\ell}}. \end{aligned}$$

We rearrange this expression into the more convenient following form,

$$\begin{aligned} z \frac{n_{\ell}}{N} \mathbb{E} \left[\mathbf{Q}^{(\ell)} \frac{\text{Tr} \mathbf{Q}^{(\ell)}}{n_{\ell}} \right] + \left(z + \frac{n_{\ell} - \prod_{\ell' \neq \ell} n_{\ell'}}{N} \right) \mathbb{E} \left[\mathbf{Q}^{(\ell)} \right] + \mathbf{I}_{n_{\ell}} \\ = -\frac{1}{N} \mathbb{E} \left[\mathbf{Q}^{(\ell)} - z \mathbf{Q}^{(\ell)2} \right] + \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right]. \end{aligned}$$

Here, the divergence of the spectrum of $\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top}$ becomes problematic: its resolvent $\mathbf{Q}^{(\ell)}$ vanish asymptotically, allowing the presence of the diverging coefficient $\frac{1}{N} \prod_{\ell' \neq \ell} n_{\ell'}$ in the previous equation. To bypass this difficulty, we proceed to a rescaling,

$$\begin{aligned} \tilde{z} \stackrel{\text{def}}{=} \frac{z - \mu_N^{(\ell)}}{\sigma_N}, \quad \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \stackrel{\text{def}}{=} \left(\frac{\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)} \mathbf{I}_{n_{\ell}}}{\sigma_N} - \tilde{z} \mathbf{I}_{n_{\ell}} \right)^{-1} = \sigma_N \mathbf{Q}^{(\ell)}(z), \\ \text{with} \quad \mu_N^{(\ell)} = \frac{1}{N} \prod_{\ell' \neq \ell} n_{\ell'}, \quad \sigma_N = \frac{1}{N} \sqrt{\prod_{\ell \in [d]} n_{\ell}}. \end{aligned}$$

This changes our equation into

$$\begin{aligned} \frac{\mu_N^{(\ell)} + \sigma_N \tilde{z} n_{\ell}}{\sigma_N^2} \frac{n_{\ell}}{N} \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_{\ell}} \right] + \left(\frac{\mu_N^{(\ell)}}{\sigma_N} + \tilde{z} + \frac{n_{\ell} - \prod_{\ell' \neq \ell} n_{\ell'}}{\sigma_N N} \right) \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] + \mathbf{I}_{n_{\ell}} \\ = -\frac{1}{\sigma_N N} \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} - \left(\frac{\mu_N^{(\ell)}}{\sigma_N} + \tilde{z} \right) \tilde{\mathbf{Q}}^{(\ell)2} \right] + \frac{1}{\sigma_N} \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \right]. \end{aligned}$$

Replacing the vanishing terms by their asymptotic order of magnitude, we have,

$$\begin{aligned} \left(\frac{\mu_N^{(\ell)}}{\sigma_N^2} + \Theta(N^{-\frac{d-2}{2}}) \right) \frac{n_{\ell}}{N} \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_{\ell}} \right] + \left(\tilde{z} + \Theta(N^{-\frac{d-2}{2}}) \right) \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] + \mathbf{I}_{n_{\ell}} \\ = -\Theta(N^{-\frac{d}{2}}) \times \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] + \Theta(N^{-1}) \times \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)2} \right] + \frac{1}{\sigma_N} \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \right]. \end{aligned}$$

We see that there is no longer any diverging coefficient, which justifies our choice of rescaling. Thus, keeping only the non-vanishing terms, we can now proceed to the following matrix equivalent formula²⁰

$$\tilde{\mathbf{Q}}^{(\ell)} \frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_{\ell}} + \tilde{z} \tilde{\mathbf{Q}}^{(\ell)} + \mathbf{I}_{n_{\ell}} \longleftrightarrow \frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \quad (13)$$

20. This is justified by the use of standard concentration arguments such as Poincaré-Nash inequality (Chen, 1982; Ledoux, 2001) and Borel-Cantelli lemma (Billingsley, 2012, equation 5.31) — see Appendix B for details. This reasoning is applied whenever we move from an equality with expectations to a matrix equivalent formula.

where we have used the fact that $\frac{\mu_N^{(\ell)} n_\ell}{\sigma_N^2 N} \xrightarrow{N \rightarrow +\infty} 1$.

At this point, we still need to handle the term on the right-hand-side, which has been purposely ignored until now. From Equation (10), we have,

$$\mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \left(\mathbf{Q}^{(\ell)} + \frac{n_\ell}{N} \mathbf{Q}^{(\ell)} \frac{\text{Tr} \mathbf{Q}^{(\ell)}}{n_\ell} + \frac{1}{N} \mathbf{Q}^{(\ell)2} \right) \right] = \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \mathbf{Q}^{(\ell)} \right].$$

Applying the same rescaling, this becomes

$$\mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \left(\tilde{\mathbf{Q}}^{(\ell)} + \Theta(N^{-\frac{d-2}{2}}) \times \tilde{\mathbf{Q}}^{(\ell)} \frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_\ell} + \Theta(N^{-\frac{d}{2}}) \times \tilde{\mathbf{Q}}^{(\ell)2} \right) \right] = \mathbb{E} \left[\mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \right].$$

Therefore,

$$\mathbf{P}^{(\ell)} \mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \longleftrightarrow \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)}$$

and Equation (13) becomes

$$\tilde{\mathbf{Q}}^{(\ell)} \frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_\ell} + \tilde{z} \tilde{\mathbf{Q}}^{(\ell)} + \mathbf{I}_{n_\ell} \longleftrightarrow \frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)}.$$

Thus, under the assumption that $\|\mathcal{P}\|_{\mathbb{F}}^2/\sigma_N$ stays bounded asymptotically, taking the normalized trace of the previous formula yields a relation for the Stieltjes transform of the limiting spectral distribution,

$$\frac{\text{Tr} \tilde{\mathbf{Q}}^{(\ell)}}{n_\ell} \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} \tilde{m}(\tilde{z}) \quad \text{with} \quad \tilde{m}^2(\tilde{z}) + \tilde{z} \tilde{m}(\tilde{z}) + 1 = 0.$$

Indeed, $\frac{1}{n_\ell} \frac{1}{\sigma_N} \text{Tr} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)} \rightarrow 0$ almost surely as $N \rightarrow +\infty$ because $\mathbf{P}^{(\ell)}$ is a low-rank matrix. The indeterminacy in the solution to the quadratic equation is raised by choosing $\tilde{m}(\tilde{z})$ such that $\tilde{z} \mapsto \tilde{m}(\tilde{z})$ follows the properties of a Stieltjes transform (in particular $\Im[\tilde{z}] \Im[\tilde{m}(\tilde{z})] > 0$ for all $\tilde{z} \in \mathbb{C} \setminus \mathbb{R}$, see Couillet and Liao (2022, §2.1.2)).

Hence, we have $\tilde{m}(\tilde{z}) \tilde{\mathbf{Q}}^{(\ell)} + \tilde{z} \tilde{\mathbf{Q}}^{(\ell)} + \mathbf{I}_{n_\ell} \leftrightarrow \frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)}$. Let us then define $\bar{\mathbf{Q}}^{(\ell)}$, the *deterministic* matrix such that

$$\begin{aligned} \tilde{m}(\tilde{z}) \bar{\mathbf{Q}}^{(\ell)} + \tilde{z} \bar{\mathbf{Q}}^{(\ell)} + \mathbf{I}_{n_\ell} &= \frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} \bar{\mathbf{Q}}^{(\ell)} \\ \iff \bar{\mathbf{Q}}^{(\ell)} &= \left(\frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} - (\tilde{m}(\tilde{z}) + \tilde{z}) \mathbf{I}_{n_\ell} \right)^{-1}. \end{aligned}$$

The evaluation of traces and bilinear forms on $\tilde{\mathbf{Q}}^{(\ell)}$ and $\bar{\mathbf{Q}}^{(\ell)}$ are equal almost surely as $N \rightarrow +\infty$, therefore we conclude that $\bar{\mathbf{Q}}^{(\ell)}$ is a *deterministic equivalent* of $\tilde{\mathbf{Q}}^{(\ell)}$ according to Definition 3. Since $\tilde{m}^2(\tilde{z}) + \tilde{z} \tilde{m}(\tilde{z}) + 1 = 0$, we have $\tilde{m}(\tilde{z}) + \tilde{z} = -\frac{1}{\tilde{m}(\tilde{z})}$ and we find the definition of $\bar{\mathbf{Q}}^{(\ell)}$ given in Theorem 5.

Appendix B. Concentration Around the Mean

B.1 Overview

In the proof of Theorem 5, the “jumps” from an equality between expectations to a matrix equivalent formula (such as Equation (13)) is justified by the concentration of the resolvent $\tilde{\mathbf{Q}}^{(\ell)}$ around its mean. This is derived from the two following lemmas.

Lemma 16 (Poincaré-Nash inequality) *Let $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and $f : \mathbb{R}^p \rightarrow \mathbb{R}$ be a differentiable function with polynomially bounded partial derivatives $\partial_1 f, \dots, \partial_p f$. Then,*

$$\text{Var}(f(\mathbf{z})) \leq \mathbb{E} \left[\|\nabla f(\mathbf{z})\|^2 \right]$$

where $\nabla = [\partial_1 \ \dots \ \partial_p]^\top$.

Proof See Pastur and Shcherbina (2011, Proposition 2.1.6). ■

Lemma 17 *Let $(X_n)_{n \geq 0}$ be a sequence of random variables on \mathbb{R} . If $\mathbb{E}[X_n] \rightarrow \bar{X} \in \mathbb{R}$ as $n \rightarrow +\infty$ and there exists an integer $k \geq 2$ such that $\sum_{n \geq 0} \mathbb{E}[|X_n - \mathbb{E}[X_n]|^k] < +\infty$, then $X_n \rightarrow \bar{X}$ almost surely as $n \rightarrow +\infty$.*

Proof Let $\varepsilon > 0$. By Markov’s inequality (Billingsley, 2012, equation 5.31),

$$\sum_{n \geq 0} \mathbb{P}(|X_n - \mathbb{E}[X_n]| \geq \varepsilon) \leq \frac{1}{\varepsilon^k} \sum_{n \geq 0} \mathbb{E}[|X_n - \mathbb{E}[X_n]|^k] < +\infty.$$

Thus, by the first Borel-Cantelli lemma²¹ (Billingsley, 2012, Theorem 4.3),

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} |X_n - \mathbb{E}[X_n]| \geq \varepsilon \right) = 0.$$

This implies the almost sure convergence of $(X_n)_{n \geq 0}$ to \bar{X} :

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} |X_n - \bar{X}| \geq \varepsilon \right) \leq \mathbb{P} \left(\limsup_{n \rightarrow +\infty} [|X_n - \mathbb{E}[X_n]| + |\mathbb{E}[X_n] - \bar{X}|] \geq \varepsilon \right) = 0.$$
■

Relying on Lemma 16 and Equation (12), we find in Appendix B.2 that

$$\begin{aligned} \mathbb{E} \left[\left| \text{Tr} \mathbf{A} \left(\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \right] \right) \right|^2 \right] &= \mathcal{O}(N^{-2}), \\ \mathbb{E} \left[\left| \mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \right] \right) \mathbf{b} \right|^4 \right] &= \mathcal{O}(N^{-2}) \end{aligned}$$

21. Given a sequence of events $(A_n)_{n \geq 0}$ in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, if $\sum_{n \geq 0} \mathbb{P}(A_n)$ converges then $\mathbb{P}(\limsup_{n \geq 0} A_n) = 0$.

for all deterministic matrix $\mathbf{A} \in \mathbb{R}^{n_\ell \times n_\ell}$ and vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n_\ell}$ of bounded norms (spectral and Euclidean norms respectively). By Lemma 17, this implies that

$$\mathrm{Tr} \mathbf{A} \left(\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \right] \right) \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} 0 \quad \text{and} \quad \mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \right] \right) \mathbf{b} \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} 0.$$

Hence, $\mathbb{E}[\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z})]$ is a deterministic equivalent of $\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z})$ according to Definition 3.

This concentration around the mean justifies our ‘‘jumps’’ because it preserves the evaluation of traces and bilinear forms (this is the only meaning of the relation \leftrightarrow).

B.2 Proof of the Concentration

B.2.1 CONCENTRATION OF TRACES

Let $\mathbf{A} \in \mathbb{R}^{n_\ell \times n_\ell}$ be a deterministic matrix of bounded norm. The Poincaré-Nash inequality (Lemma 16) yields

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{n_\ell} \mathrm{Tr} \mathbf{A} \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \right)^2 \right] &= \mathrm{Var} \left(\frac{1}{n_\ell} \mathrm{Tr} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)} \right) \\ &\leq \frac{1}{n_\ell^2} \sum_{i=1}^{n_\ell} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[\left(\frac{\partial \mathrm{Tr} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)}}{\partial N_{i,j}^{(\ell)}} \right)^2 \right]. \end{aligned}$$

With Equation (12), we have,

$$\begin{aligned} \frac{\partial \mathrm{Tr} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)}}{\partial N_{i,j}^{(\ell)}} &= \sum_{k=1}^{n_\ell} \sum_{l=1}^{n_\ell} A_{k,l} \frac{\partial \tilde{Q}_{l,k}^{(\ell)}}{\partial N_{i,j}^{(\ell)}} \\ &= -\frac{1}{\sigma_N \sqrt{N}} \sum_{k=1}^{n_\ell} \sum_{l=1}^{n_\ell} A_{k,l} \left[\tilde{Q}_{l,i}^{(\ell)} [\mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)}]_{j,k} + [\tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{l,j} \tilde{Q}_{i,k}^{(\ell)} \right] \\ &= -\frac{1}{\sigma_N \sqrt{N}} \left[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{A}^\top \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} + \tilde{\mathbf{Q}}^{(\ell)} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{i,j}. \end{aligned}$$

Thus, using the fact that $(a+b)^2 \leq 2(a^2+b^2)$, we have,

$$\begin{aligned} &\frac{1}{n_\ell^2} \sum_{i=1}^{n_\ell} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[\left(\frac{\partial \mathrm{Tr} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)}}{\partial N_{i,j}^{(\ell)}} \right)^2 \right] \\ &\leq \frac{2}{\sigma_N^2 N n_\ell^2} \sum_{i=1}^{n_\ell} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{A}^\top \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{i,j}^2 + [\tilde{\mathbf{Q}}^{(\ell)} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{i,j}^2 \right] \\ &= \frac{2}{\sigma_N^2 N n_\ell^2} \mathbb{E} \left[\left\| \tilde{\mathbf{Q}}^{(\ell)} \mathbf{A}^\top \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right\|_{\mathrm{F}}^2 + \left\| \tilde{\mathbf{Q}}^{(\ell)} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right\|_{\mathrm{F}}^2 \right] \\ &\leq \frac{4}{\sigma_N^2 N n_\ell^2} \|\mathbf{A}\|_{\mathrm{F}}^2 \mathbb{E} \left[\left\| \tilde{\mathbf{Q}}^{(\ell)} \right\|^4 \left\| \mathbf{T}^{(\ell)} \right\|^2 \right] \end{aligned}$$

where the last inequality is obtained using the property $\|\mathbf{A}\mathbf{B}\|_{\text{F}} \leq \|\mathbf{A}\| \|\mathbf{B}\|_{\text{F}}$ and the sub-multiplicative property $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$. Since $\|\mathbf{A}\|_{\text{F}}^2 \leq n_{\ell} \|\mathbf{A}\|^2 = \mathcal{O}(N)$ (because $n_{\ell} = \Theta(N)$ and $\|\mathbf{A}\|$ is bounded), $\|\tilde{\mathbf{Q}}^{(\ell)}\| = \mathcal{O}(1)$ and $\frac{1}{\sigma_N^2} \|\mathbf{T}^{(\ell)}\|^2 = \mathcal{O}(1)$, we find that $\text{Var}(\frac{1}{n_{\ell}} \text{Tr} \mathbf{A} \tilde{\mathbf{Q}}^{(\ell)}) = \mathcal{O}(N^{-2})$; Q.E.D.

B.2.2 CONCENTRATION OF BILINEAR FORMS

Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n_{\ell}}$ be two deterministic vectors with bounded norms. The Poincaré-Nash inequality (Lemma 16) yields

$$\mathbb{E} \left[\left(\mathbf{a}^{\top} \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \right] = \text{Var} \left(\mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \right) \leq \sum_{i=1}^{n_{\ell}} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[\left(\frac{\partial \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b}}{\partial N_{i,j}^{(\ell)}} \right)^2 \right].$$

With Equation (12), we have,

$$\begin{aligned} \frac{\partial \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b}}{\partial N_{i,j}^{(\ell)}} &= \sum_{k=1}^{n_{\ell}} \sum_{l=1}^{n_{\ell}} a_k \frac{\partial \tilde{Q}_{k,l}^{(\ell)}}{\partial N_{i,j}^{(\ell)}} b_l \\ &= -\frac{1}{\sigma_N \sqrt{N}} \sum_{k=1}^{n_{\ell}} \sum_{l=1}^{n_{\ell}} a_k \left[\tilde{Q}_{k,i}^{(\ell)} [\mathbf{T}^{(\ell)\top} \tilde{\mathbf{Q}}^{(\ell)}]_{j,l} + [\tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{k,j} \tilde{Q}_{i,l}^{(\ell)} \right] b_l \\ &= -\frac{1}{\sigma_N \sqrt{N}} \left[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{a} \mathbf{b}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} + \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{i,j}. \end{aligned}$$

Thus, using the fact that $(a+b)^2 \leq 2(a^2+b^2)$, we have,

$$\begin{aligned} &\sum_{i=1}^{n_{\ell}} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[\left(\frac{\partial \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b}}{\partial N_{i,j}^{(\ell)}} \right)^2 \right] \\ &\leq \frac{2}{\sigma_N^2 N} \sum_{i=1}^{n_{\ell}} \sum_{j=1}^{\prod_{\ell' \neq \ell} n_{\ell'}} \mathbb{E} \left[[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{a} \mathbf{b}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{i,j}^2 + [\tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)}]_{i,j}^2 \right] \\ &= \frac{2}{\sigma_N^2 N} \mathbb{E} \left[\left\| \tilde{\mathbf{Q}}^{(\ell)} \mathbf{a} \mathbf{b}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right\|_{\text{F}}^2 + \left\| \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right\|_{\text{F}}^2 \right] \\ &\leq \frac{2}{\sigma_N^2 N} \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \mathbb{E} \left[\left\| \tilde{\mathbf{Q}}^{(\ell)} \right\|^2 \left\| \mathbf{T}^{(\ell)} \right\|^2 \right] = \mathcal{O}(N^{-1}) \end{aligned}$$

where the last line is obtained similarly to the previous case. Hence, $\text{Var}(\mathbf{a}^{\top} \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b}) = \mathcal{O}(N^{-1})$, which is not enough to apply Lemma 17. We must therefore evaluate the moment of order 4,

$$\begin{aligned} \mathbb{E} \left[\left(\mathbf{a}^{\top} \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^4 \right] &= \text{Var} \left(\left(\mathbf{a}^{\top} \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \right) \\ &\quad + \mathbb{E} \left[\left(\mathbf{a}^{\top} \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \right]^2. \end{aligned}$$

Since we already known that the rightmost term is $\mathcal{O}(N^{-2})$, we use the Poincaré-Nash inequality (Lemma 16) on the variance of $(\mathbf{a}^\top (\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E}[\tilde{\mathbf{Q}}^{(\ell)}]) \mathbf{b})^2$,

$$\begin{aligned} & \text{Var} \left(\left(\mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \right) \\ & \leq \sum_{i,j} \mathbb{E} \left[\left(\mathbf{a}^\top \frac{\partial}{\partial N_{i,j}^{(\ell)}} \left[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)} - \tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^\top \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \mathbf{b} \mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)} \right] \mathbf{b} \right)^2 \right]. \end{aligned}$$

Similarly, injecting the expression of $\partial \tilde{\mathbf{Q}}^{(\ell)} / \partial N_{i,j}^{(\ell)}$, straightforward computations yield

$$\begin{aligned} & \text{Var} \left(\left(\mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \right) \\ & \leq \sum_{i,j} \mathbb{E} \left[\left(-\frac{2}{\sigma_N \sqrt{N}} \left(\mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right) \left[\tilde{\mathbf{Q}}^{(\ell)} \left(\mathbf{a} \mathbf{b}^\top + \mathbf{b} \mathbf{a}^\top \right) \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{i,j} \right)^2 \right] \\ & \leq \frac{8}{\sigma_N^2 N} \sum_{i,j} \mathbb{E} \left[\left(\mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \left(\left[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{a} \mathbf{b}^\top \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{i,j}^2 + \left[\tilde{\mathbf{Q}}^{(\ell)} \mathbf{b} \mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)} \mathbf{T}^{(\ell)} \right]_{i,j}^2 \right) \right] \\ & \leq \frac{16}{\sigma_N^2 N} \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \mathbb{E} \left[\left(\mathbf{a}^\top \left(\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E} \left[\tilde{\mathbf{Q}}^{(\ell)} \right] \right) \mathbf{b} \right)^2 \left\| \tilde{\mathbf{Q}}^{(\ell)} \right\|^4 \left\| \mathbf{T}^{(\ell)} \right\|^2 \right]. \end{aligned}$$

Since we have shown that $\mathbb{E}[(\mathbf{a}^\top (\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E}[\tilde{\mathbf{Q}}^{(\ell)}]) \mathbf{b})^2] = \mathcal{O}(N^{-1})$, we have $\text{Var}((\mathbf{a}^\top (\tilde{\mathbf{Q}}^{(\ell)} - \mathbb{E}[\tilde{\mathbf{Q}}^{(\ell)}]) \mathbf{b})^2) = \mathcal{O}(N^{-2})$; Q.E.D.

Appendix C. Proof of Theorem 8

C.1 Isolated Eigenvalues

We seek eigenvalues of $\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)}]$ which stay outside the support of the semicircle distribution $[-2, +2]$. The resolvent $\tilde{\mathbf{Q}}^{(\ell)}$ is defined on $\mathbb{C} \setminus \text{Sp} \left(\frac{1}{\sigma_N} [\mathbf{T}^{(\ell)} \mathbf{T}^{(\ell)\top} - \mu_N^{(\ell)}] \right)$. Therefore, we seek the singular points of $\tilde{\mathbf{Q}}^{(\ell)}$ outside the support of the limiting spectral distribution. Because $\tilde{\mathbf{Q}}^{(\ell)}$ and $\bar{\mathbf{Q}}^{(\ell)}$ are almost surely asymptotically equal when evaluated on bilinear forms and traces (Definition 3), they must have the same singular points as $N \rightarrow +\infty$. Hence, the asymptotic positions of isolated eigenvalues are singular points of the deterministic equivalent $\bar{\mathbf{Q}}^{(\ell)} : \tilde{z} \mapsto \left(\frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} + \frac{1}{\tilde{m}(\tilde{z})} \mathbf{I}_{n_\ell} \right)^{-1}$. Thus, we seek $\tilde{\xi}_{q_\ell}^{(\ell)} \in \mathbb{R} \setminus [-2, +2]$ such that

$$\frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} + \frac{1}{\tilde{m}(\tilde{\xi}_{q_\ell}^{(\ell)})} = 0, \quad q_\ell \in [r_\ell].$$

Injecting the expression $\tilde{m}(\tilde{\xi}_{q_\ell}^{(\ell)}) = -\frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}$ into the equation $\tilde{m}^2(\tilde{\xi}_{q_\ell}^{(\ell)}) + \tilde{\xi}_{q_\ell}^{(\ell)} \tilde{m}(\tilde{\xi}_{q_\ell}^{(\ell)}) + 1 = 0$ yields

$$\frac{\sigma_N^2}{s_{q_\ell}^4(\mathbf{P}^{(\ell)})} - \tilde{\xi}_{q_\ell}^{(\ell)} \frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})} + 1 = 0 \iff \tilde{\xi}_{q_\ell}^{(\ell)} = \frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} + \frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}.$$

As $\tilde{\xi}_{q_\ell}^{(\ell)} > 0$ by definition, to be a singular point of $\bar{\mathbf{Q}}^{(\ell)}(\tilde{z})$, it must be strictly greater than 2 (the right edge of the semicircle). This is true only if $\rho_{q_\ell}^{(\ell)} \stackrel{\text{def}}{=} \frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} > 1$.

C.2 Eigenvector Alignments

Let $\hat{\mathbf{u}}_{i_\ell}^{(\ell)}$, $i_\ell \in [n_\ell]$, denote the i_ℓ -th left singular vector of $\mathbf{T}^{(\ell)}$ (sorted in non-increasing order of its corresponding singular value). From the spectral decomposition $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} = \sum_{i_\ell=1}^{n_\ell} s_{i_\ell}^2(\mathbf{T}^{(\ell)})\hat{\mathbf{u}}_{i_\ell}^{(\ell)}\hat{\mathbf{u}}_{i_\ell}^{(\ell)\top}$, we have,

$$\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) = \sum_{i_\ell=1}^{n_\ell} \frac{\hat{\mathbf{u}}_{q_\ell}^{(\ell)}\hat{\mathbf{u}}_{q_\ell}^{(\ell)\top}}{\frac{1}{\sigma_N} \left[s_{i_\ell}^2(\mathbf{T}^{(\ell)}) - \mu_N^{(\ell)} \right] - \tilde{z}}.$$

If $\rho_{q_\ell}^{(\ell)} > 1$, $q_\ell \in [r_\ell]$, then $s_{q_\ell}^2(\mathbf{T}^{(\ell)})$ is an isolated eigenvalue in the spectrum of $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ and $\frac{1}{\sigma_N} \left[s_{q_\ell}^2(\mathbf{T}^{(\ell)}) - \mu_N^{(\ell)} \right] \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} \tilde{\xi}_{q_\ell}^{(\ell)}$. Hence, for any positively-oriented complex contour $\gamma_{q_\ell}^{(\ell)}$ circling around $\tilde{\xi}_{q_\ell}^{(\ell)}$, leaving all the other $\tilde{\xi}_{q'_\ell}^{(\ell)}$, $q'_\ell \neq q_\ell$, outside and not crossing $[-2, +2]$, Cauchy's integral formula yields, for N large enough and any $\mathbf{a} \in \mathbb{R}^{n_\ell}$,

$$\left| \mathbf{a}^\top \hat{\mathbf{u}}_{q_\ell}^{(\ell)} \right|^2 = -\frac{1}{2i\pi} \oint_{\gamma_{q_\ell}^{(\ell)}} \mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a} \, d\tilde{z} \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} -\frac{1}{2i\pi} \oint_{\gamma_{q_\ell}^{(\ell)}} \mathbf{a}^\top \bar{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a} \, d\tilde{z}$$

by the dominated convergence theorem since, for all $\tilde{z} \in \gamma_{q_\ell}^{(\ell)}$, $\mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a} \rightarrow \mathbf{a}^\top \bar{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a}$ almost surely as $N \rightarrow +\infty$ by the definition of deterministic equivalent (Definition 3) and $\tilde{z} \mapsto |\mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a}|$ is almost surely bounded on $\gamma_{q_\ell}^{(\ell)}$ for N large enough because, according to Theorem 8, the eigenvalue is almost surely isolated thus we can choose $\gamma_{q_\ell}^{(\ell)}$ such that $\text{Dist}(\tilde{\xi}_{q_\ell}^{(\ell)}, \gamma_{q_\ell}^{(\ell)}) \geq \varepsilon > 0$ and therefore $|\mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a}| \leq \|\mathbf{a}\|^2 \|\tilde{\mathbf{Q}}^{(\ell)}(\tilde{z})\| \leq \|\mathbf{a}\|^2 / \varepsilon$ almost surely.

Using residue calculus, we can compute,

$$-\frac{1}{2i\pi} \oint_{\gamma_{q_\ell}^{(\ell)}} \mathbf{a}^\top \tilde{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a} \, d\tilde{z} = -\lim_{\tilde{z} \rightarrow \tilde{\xi}_{q_\ell}^{(\ell)}} \left(\tilde{z} - \tilde{\xi}_{q_\ell}^{(\ell)} \right) \mathbf{a}^\top \left(\frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} + \frac{1}{\tilde{m}(\tilde{z})} \mathbf{I}_{n_\ell} \right)^{-1} \mathbf{a}.$$

Note that $\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top} = \mathbf{X}^{(\ell)}\mathbf{H}^{(\ell)}\mathbf{H}^{(\ell)\top}\mathbf{X}^{(\ell)\top}$ and there exist an $r_\ell \times r_\ell$ orthogonal matrix $\mathbf{O}^{(\ell)}$ such that $\mathbf{H}^{(\ell)}\mathbf{H}^{(\ell)\top} = \mathbf{O}^{(\ell)}\mathbf{L}^{(\ell)}\mathbf{O}^{(\ell)\top}$ with $\mathbf{L}^{(\ell)} = \text{Diag}(s_1^2(\mathbf{P}^{(\ell)}), \dots, s_{r_\ell}^2(\mathbf{P}^{(\ell)}))$. Hence,

$$\mathbf{a}^\top \left(\frac{1}{\sigma_N} \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)\top} + \frac{1}{\tilde{m}(\tilde{z})} \mathbf{I}_{n_\ell} \right)^{-1} \mathbf{a} = \mathbf{a}^\top \mathbf{X}^{(\ell)} \mathbf{O}^{(\ell)} \left(\frac{1}{\sigma_N} \mathbf{L}^{(\ell)} + \frac{1}{\tilde{m}(\tilde{z})} \mathbf{I}_{n_\ell} \right)^{-1} \mathbf{O}^{(\ell)\top} \mathbf{X}^{(\ell)\top} \mathbf{a}.$$

Let us therefore compute the following quantity,

$$-\lim_{\tilde{z} \rightarrow \tilde{\xi}_{q_\ell}^{(\ell)}} \left(\tilde{z} - \tilde{\xi}_{q_\ell}^{(\ell)} \right) \left(\frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} + \frac{1}{\tilde{m}(\tilde{z})} \right)^{-1} = \begin{cases} 0 & \text{if } q'_\ell \neq q_\ell \\ \zeta_{q_\ell}^{(\ell)} & \text{if } q'_\ell = q_\ell \end{cases}, \quad q'_\ell \in [r_\ell],$$

where we have used the fact that $\tilde{m}(\tilde{\xi}_{q_\ell}^{(\ell)}) = -\frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}$. The case $q'_\ell = q_\ell$ can be handled with L'Hôpital's rule,

$$\begin{aligned}\zeta_{q_\ell}^{(\ell)} &= -\left(\frac{d}{d\tilde{z}}\left[\frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} + \frac{1}{\tilde{m}(\tilde{z})}\right]_{\tilde{z}=\tilde{\xi}_{q_\ell}^{(\ell)}}\right)^{-1} \\ &= \frac{\tilde{m}^2(\tilde{\xi}_{q_\ell}^{(\ell)})}{\tilde{m}'(\tilde{\xi}_{q_\ell}^{(\ell)})} \\ &= \frac{\sigma_N^2}{s_{q_\ell}^4(\mathbf{P}^{(\ell)})\tilde{m}'(\tilde{\xi}_{q_\ell}^{(\ell)})}.\end{aligned}$$

In order to compute $\tilde{m}'(\tilde{\xi}_{q_\ell}^{(\ell)})$, let us differentiate the relation $\tilde{m}^2(\tilde{z}) + \tilde{z}\tilde{m}(\tilde{z}) + 1 = 0$,

$$\begin{aligned}2\tilde{m}'(\tilde{z})\tilde{m}(\tilde{z}) + \tilde{m}(\tilde{z}) + \tilde{z}\tilde{m}'(\tilde{z}) &= 0, \\ \tilde{m}'(\tilde{z}) &= -\frac{\tilde{m}(\tilde{z})}{2\tilde{m}(\tilde{z}) + \tilde{z}}.\end{aligned}$$

Hence,

$$\begin{aligned}\tilde{m}'(\tilde{\xi}_{q_\ell}^{(\ell)}) &= -\frac{-\frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}}{-2\frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})} + \frac{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}{\sigma_N} + \frac{\sigma_N}{s_{q_\ell}^2(\mathbf{P}^{(\ell)})}} \\ &= \frac{1}{\frac{s_{q_\ell}^4(\mathbf{P}^{(\ell)})}{\sigma_N^2} - 1}.\end{aligned}$$

Back to our previous expression of $\zeta_{q_\ell}^{(\ell)}$, we now have,

$$\zeta_{q_\ell}^{(\ell)} = 1 - \frac{\sigma_N^2}{s_{q_\ell}^4(\mathbf{P}^{(\ell)})}.$$

Therefore, for all $\mathbf{a} \in \mathbb{R}^{n_\ell}$,

$$-\frac{1}{2i\pi} \oint_{\gamma_{q_\ell}^{(\ell)}} \mathbf{a}^\top \bar{\mathbf{Q}}^{(\ell)}(\tilde{z}) \mathbf{a} \, d\tilde{z} = \mathbf{a}^\top \mathbf{X}^{(\ell)} \mathbf{O}^{(\ell)} \mathbf{Z}_{q_\ell}^{(\ell)} \mathbf{O}^{(\ell)\top} \mathbf{X}^{(\ell)\top} \mathbf{a}$$

where $\mathbf{Z}_{q_\ell}^{(\ell)}$ is an $r_\ell \times r_\ell$ matrix with all its entries equal to 0 except $[\mathbf{Z}_{q_\ell}^{(\ell)}]_{q_\ell, q_\ell} = \zeta_{q_\ell}^{(\ell)}$. Thus, summing the alignments of $\hat{\mathbf{u}}_{q_\ell}^{(\ell)}$ with each column of $\mathbf{X}^{(\ell)}$ yields

$$\left\| \mathbf{X}^{(\ell)\top} \hat{\mathbf{u}}_{q_\ell}^{(\ell)} \right\|^2 \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} \zeta_{q_\ell}^{(\ell)} \sum_{q'_\ell=1}^{r_\ell} O_{q'_\ell, q_\ell}^{(\ell)2} = \zeta_{q_\ell}^{(\ell)}.$$

Appendix D. Proof of Lemma 13

Our proof of Lemma 13 uses the notion of ε -covering. An ε -covering of a *compact* set \mathcal{K} for the norm $\|\cdot\|$ is a *finite* set $\mathcal{C} \subset \mathcal{K}$ such that for all $x \in \mathcal{K}$, there exists $\bar{x} \in \mathcal{C}$ such that $\|x - \bar{x}\| \leq \varepsilon$. We also define the covering number $N(\varepsilon, \mathcal{K}, \|\cdot\|)$ as the smallest possible number of elements in \mathcal{C} .

Moreover, we recall the definition of the Gamma function $\Gamma(s) = \int_0^{+\infty} t^{s-1} e^{-t} dt$ and the (upper) incomplete Gamma function $\Gamma(s, x) = \int_x^{+\infty} t^{s-1} e^{-t} dt$ for $s > 0$ and $x \geq 0$.

For our proof, we need to introduce a few preliminary results which are stated and proven below (except Lemma 20 for which a reference is given).

Lemma 18 *For $\ell \in [d]$ and $\varepsilon > 0$, let $\Delta^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$ be such that $\|\Delta^{(\ell)}\| \leq \varepsilon$ and $\mathbf{V}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$ be the matrix of its left singular vectors. For all $\mathbf{A}^{(\ell')} \in \mathbb{R}^{n_{\ell'} \times r_{\ell'}}$, $\ell' \neq \ell$,*

$$\left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \Delta^{(\ell)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}} \leq \varepsilon \left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{V}^{(\ell)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}}.$$

Proof Let $\mathbf{V}^{(\ell)} \Sigma^{(\ell)} \mathbf{W}^{(\ell)\top}$ be the singular value decomposition of $\Delta^{(\ell)}$. We have,

$$\begin{aligned} \left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \Delta^{(\ell)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}} &= \left\| \Delta^{(\ell)\top} \mathcal{N}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{A}^{(\ell')} \right\|_{\mathbb{F}} \\ &= \left\| \mathbf{W}^{(\ell)\top} \Sigma^{(\ell)} \mathbf{V}^{(\ell)\top} \mathcal{N}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{A}^{(\ell')} \right\|_{\mathbb{F}} \leq \underbrace{\left\| \mathbf{W}^{(\ell)\top} \Sigma^{(\ell)} \right\|}_{=\varepsilon} \underbrace{\left\| \mathbf{V}^{(\ell)\top} \mathcal{N}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{A}^{(\ell')} \right\|_{\mathbb{F}}}_{=\left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{V}^{(\ell)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}}} \end{aligned}$$

using the fact that $\|\mathbf{AB}\|_{\mathbb{F}} \leq \|\mathbf{A}\| \|\mathbf{B}\|_{\mathbb{F}}$. ■

Lemma 19 *Given $\mathbf{A}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$, $\ell \in [d]$,*

$$\left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}}^2 \sim \chi^2 \left(\prod_{\ell \in [d]} r_\ell \right).$$

Proof Firstly, observe that, for all $(q_1, \dots, q_d) \in \times_{\ell \in [d]} [r_\ell]$,

$$[\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})]_{q_1, \dots, q_d} = \sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} \mathcal{N}_{i_1, \dots, i_d} A_{i_1, q_1}^{(1)} \dots A_{i_d, q_d}^{(d)} \sim \mathcal{N}(0, 1)$$

since $\sum_{i_\ell=1}^{n_\ell} A_{i_\ell, q_\ell}^{(\ell)2} = 1$ for all $\ell \in [d]$. Then, we show that all the entries of $\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})$ are independent because their covariance is identity,

$$\begin{aligned} & \mathbb{E} \left[[\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})]_{q_1, \dots, q_d} [\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})]_{q'_1, \dots, q'_d} \right] \\ &= \sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} \sum_{i'_1, \dots, i'_d=1}^{n_1, \dots, n_d} \mathbb{E} \left[\mathcal{N}_{i_1, \dots, i_d} \mathcal{N}_{i'_1, \dots, i'_d} \right] A_{i_1, q_1}^{(1)} A_{i'_1, q'_1}^{(1)} \dots A_{i_d, q_d}^{(d)} A_{i'_d, q'_d}^{(d)} \\ &= \sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} A_{i_1, q_1}^{(1)} A_{i_1, q'_1}^{(1)} \dots A_{i_d, q_d}^{(d)} A_{i_d, q'_d}^{(d)} \\ &= \begin{cases} 1 & \text{if } (q_1, \dots, q_d) = (q'_1, \dots, q'_d) \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

Hence, the result follows from the fact that $\|\mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})\|_F^2$ is the sum of $\prod_{\ell \in [d]} r_\ell$ squared independent $\mathcal{N}(0, 1)$ variables. \blacksquare

Lemma 20 (Hinrichs et al., 2017, Lemma 4.1) *For $0 < \varepsilon < 1$, we have the following upper bound on the ε -covering number of the Stiefel manifold $V_r(\mathbb{R}^n)$ for the spectral norm $\|\cdot\|$,*

$$N(\varepsilon, V_r(\mathbb{R}^n), \|\cdot\|) \leq \left[\frac{C}{\varepsilon} \right]^{r(n - \frac{r+1}{2})} .$$

where $C > 0$ is a universal constant.

Lemma 21 $\Gamma(s, x) \leq \max(1, e^{s-1})\Gamma(s)e^{-x/2}$ for all $x \geq 0$ and $s > 0$.

Proof Given $s > 0$, consider the function $f : x \in [0, +\infty[\mapsto \frac{\Gamma(s, x)}{C e^{-x/2}}$ with $C > 0$. Our goal is to show that $0 < f \leq 1$ when C is well chosen. f is continuously differentiable on $[0, +\infty[$ and

$$f'(x) = \frac{1}{C e^{-x/2}} \left(-x^{s-1} e^{-x} + \frac{1}{2} \Gamma(s, x) \right) \geq 0 \iff \Gamma(s, x) - 2x^{s-1} e^{-x} \geq 0.$$

Consider the function $g : x \in [0, +\infty[\mapsto \Gamma(s, x) - 2x^{s-1} e^{-x}$. g is also continuously differentiable on $[0, +\infty[$ and

$$g'(x) = x^{s-1} e^{-x} - 2(s-1)x^{s-2} e^{-x} \geq 0 \iff x \geq 2(s-1).$$

We distinguish two cases.

1. If $0 < s \leq 1$, then g increases monotonically on $[0, +\infty[$. Since $\lim_{x \rightarrow +\infty} g(x) = 0$, we necessarily have $g(x) \leq 0$ for all $x \in [0, +\infty[$. Hence, $f(x) \leq f(0) = \frac{\Gamma(s)}{C}$ and we can choose $C = \Gamma(s)$.
2. If $s > 1$, our conclusion stems from the following table.

x	0	$2(s-1)$	$+\infty$
$g'(x)$		-	+
$g(x)$	$\Gamma(s)$	$g(2(s-1))$	0

Since g is strictly increasing on $[2(s-1), +\infty[$ and $\lim_{x \rightarrow +\infty} g(x) = 0$, we necessarily have $g(2(s-1)) < 0$. Hence, since $\Gamma(s) > 0$, the equation $g(x) = 0$ has a unique solution on $[0, +\infty[$ and it lies between 0 and $2(s-1)$. Let $x_0(s) \in [0, 2(s-1)]$ be this unique solution. Then,

$$\sup_{[0, +\infty[} f = f(x_0(s)) = \frac{\Gamma(s, x_0(s))}{C e^{-x_0(s)/2}}$$

and we can choose $C = \Gamma(s, x_0(s)) e^{x_0(s)/2}$. The final result follows from $\Gamma(s, x_0(s)) \leq \Gamma(s)$ and $e^{x_0(s)/2} \leq e^{s-1}$. ■

We are now ready to prove Lemma 13.

Let $\varepsilon > 0$ and \mathcal{C}_ℓ be an ε -covering of $V_{r_\ell}(\mathbb{R}^{n_\ell})$ for the spectral norm $\|\cdot\|$, $\ell \in [d]$. Since $\times_{\ell \in [d]} V_{r_\ell}(\mathbb{R}^{n_\ell})$ is compact, it contains an element $(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)})$ such that

$$\sup_{\mathbf{A}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell}), \ell \in [d]} \left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}} = \left\| \mathcal{N}(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)}) \right\|_{\mathbb{F}}.$$

Let $\bar{\mathbf{A}}^{(\ell)} \in \mathcal{C}_\ell$ be such that $\mathbf{A}_\star^{(\ell)} = \bar{\mathbf{A}}^{(\ell)} + \mathbf{\Delta}^{(\ell)}$ with $\|\mathbf{\Delta}^{(\ell)}\| \leq \varepsilon$. Then, using the triangle inequality, Lemma 18 and the optimality of $(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)})$, we have,

$$\left\| \mathcal{N}(\bar{\mathbf{A}}^{(1)} + \mathbf{\Delta}^{(1)}, \dots, \bar{\mathbf{A}}^{(d)} + \mathbf{\Delta}^{(d)}) \right\|_{\mathbb{F}} \leq \left\| \mathcal{N}(\bar{\mathbf{A}}^{(1)}, \dots, \bar{\mathbf{A}}^{(d)}) \right\|_{\mathbb{F}} + S \left\| \mathcal{N}(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)}) \right\|_{\mathbb{F}}$$

with $S \stackrel{\text{def}}{=} \sum_{k=1}^d \binom{d}{k} \varepsilon^k \leq \sum_{k=1}^d \varepsilon^k \frac{d^k}{k!} \leq e^{\varepsilon d} - 1$. Hence, choosing $\varepsilon = \frac{1}{d} \log \frac{3}{2}$, we get,

$$\left\| \mathcal{N}(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)}) \right\|_{\mathbb{F}} \leq 2 \left\| \mathcal{N}(\bar{\mathbf{A}}^{(1)}, \dots, \bar{\mathbf{A}}^{(d)}) \right\|_{\mathbb{F}}$$

and, from the union bound, for any $t \geq 0$,

$$\begin{aligned} \mathbb{P} \left(\left\| \mathcal{N}(\mathbf{A}_\star^{(1)}, \dots, \mathbf{A}_\star^{(d)}) \right\|_{\mathbb{F}} \geq t \right) &\leq \mathbb{P} \left(\bigcup_{\mathbf{A}^{(\ell)} \in \mathcal{C}_\ell, \ell \in [d]} \left\{ \left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}} \geq \frac{t}{2} \right\} \right) \\ &\leq \sum_{\mathbf{A}^{(\ell)} \in \mathcal{C}_\ell, \ell \in [d]} \mathbb{P} \left(\left\| \mathcal{N}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \right\|_{\mathbb{F}} \geq \frac{t}{2} \right). \end{aligned}$$

Thus, combining Lemma 19 and 20, we have,

$$\mathbb{P}\left(\left\|\mathcal{N}(\mathbf{A}_*^{(1)}, \dots, \mathbf{A}_*^{(d)})\right\|_{\mathbb{F}} \geq t\right) \leq \left[\frac{Cd}{\log \frac{3}{2}}\right]^{\sum_{\ell=1}^d r_{\ell} \binom{n_{\ell} - \frac{r_{\ell} + 1}{2}}}{\mathbb{P}\left(X \geq \frac{t^2}{4}\right)}$$

where X is a random variable following a $\chi^2(\prod_{\ell \in [d]} r_{\ell})$ distribution. Eventually, the probability on the right-hand side can be bounded using Lemma 21,

$$\mathbb{P}\left(X \geq \frac{t^2}{4}\right) = \frac{\Gamma\left(\frac{1}{2} \prod_{\ell \in [d]} r_{\ell}, \frac{t^2}{8}\right)}{\Gamma\left(\frac{1}{2} \prod_{\ell \in [d]} r_{\ell}\right)} \leq \max(1, e^{\frac{1}{2} \prod_{\ell \in [d]} r_{\ell} - 1}) e^{-t^2/16}.$$

We get the result stated in Lemma 13 with

$$t^2 = 16 \left[\left(\sum_{\ell=1}^d r_{\ell} \binom{n_{\ell} - \frac{r_{\ell} + 1}{2}} \right) \log \frac{Cd}{\log \frac{3}{2}} + \log \left(\frac{1}{\delta} \max \left(1, e^{\frac{1}{2} \prod_{\ell=1}^d r_{\ell} - 1} \right) \right) \right].$$

Appendix E. Proof of Theorem 14

Recall the decomposition $\mathcal{P} = \llbracket \mathcal{H}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)} \rrbracket$. We use the following lemma whenever we state that $\|\mathcal{P}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)})\|_{\mathbb{F}} = \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}})$ as $N \rightarrow +\infty$.

Lemma 22 *For all $(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)}) \in \times_{\ell \in [d]} V_{r_{\ell}}(\mathbb{R}^{n_{\ell}})$,*

$$\left\|\mathcal{P}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})\right\|_{\mathbb{F}} \leq \|\mathcal{P}\|_{\mathbb{F}} \prod_{\ell \in [d]} \left\|\mathbf{X}^{(\ell)\top} \mathbf{A}^{(\ell)}\right\|.$$

Proof The proof relies on the property $\|\mathbf{AB}\|_{\mathbb{F}} \leq \|\mathbf{A}\| \|\mathbf{B}\|_{\mathbb{F}}$.

$$\begin{aligned} \left\|\mathcal{P}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(d)})\right\|_{\mathbb{F}} &= \left\|\mathbf{A}^{(1)\top} \mathbf{X}^{(1)} \mathbf{H}^{(1)} \boxtimes_{\ell=2}^d \mathbf{X}^{(\ell)\top} \mathbf{A}^{(\ell)}\right\|_{\mathbb{F}} \\ &\leq \left\|\mathbf{A}^{(1)\top} \mathbf{X}^{(1)}\right\| \left\|\mathbf{H}^{(1)} \boxtimes_{\ell=2}^d \mathbf{X}^{(\ell)\top} \mathbf{A}^{(\ell)}\right\|_{\mathbb{F}} \\ &= \left\|\mathbf{A}^{(1)\top} \mathbf{X}^{(1)}\right\| \left\|\mathbf{A}^{(2)\top} \mathbf{X}^{(2)} \mathbf{H}^{(2)} \left(\mathbf{I}_{r_1} \boxtimes \boxtimes_{\ell=3}^d \mathbf{X}^{(\ell)\top} \mathbf{A}^{(\ell)}\right)\right\|_{\mathbb{F}} \\ &\leq \left\|\mathbf{A}^{(1)\top} \mathbf{X}^{(1)}\right\| \left\|\mathbf{A}^{(2)\top} \mathbf{X}^{(2)}\right\| \left\|\mathbf{H}^{(2)} \left(\mathbf{I}_{r_1} \boxtimes \boxtimes_{\ell=3}^d \mathbf{X}^{(\ell)\top} \mathbf{A}^{(\ell)}\right)\right\|_{\mathbb{F}} \\ &\dots \\ &\leq \left(\prod_{\ell=1}^d \left\|\mathbf{A}^{(\ell)\top} \mathbf{X}^{(\ell)}\right\|\right) \underbrace{\left\|\mathbf{H}^{(d)} \boxtimes_{\ell=1}^d \mathbf{I}_{r_{\ell}}\right\|_{\mathbb{F}}}_{=\|\mathcal{H}\|_{\mathbb{F}} = \|\mathcal{P}\|_{\mathbb{F}}}. \end{aligned}$$

■

Given $\ell \in [d]$, $\mathbf{U}_1^{(\ell)}$ gathers the r_ℓ dominant left singular vectors of $\mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}$, i.e., it is solution to

$$\max_{\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})} \left\| \mathbf{U}^{(\ell)\top} \mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2. \quad (14)$$

Consider also a solution $\tilde{\mathbf{U}}_1^{(\ell)}$ to the following related problem

$$\max_{\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})} \left\| \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2. \quad (15)$$

Observe that, using the property $\|\mathbf{AB}\|_{\mathbb{F}} \leq \|\mathbf{A}\| \|\mathbf{B}\|_{\mathbb{F}}$, we have,

$$\begin{aligned} \left\| \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 &= \left\| \mathbf{U}^{(\ell)\top} \mathbf{X}^{(\ell)} \mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &\leq \left\| \mathbf{U}^{(\ell)\top} \mathbf{X}^{(\ell)} \right\|^2 \left\| \mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \leq \left\| \mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \end{aligned}$$

and this upper bound is only reached with $\mathbf{U}^{(\ell)} = \mathbf{X}^{(\ell)} \mathbf{O}^{(\ell)}$, for any $r_\ell \times r_\ell$ orthogonal matrix $\mathbf{O}^{(\ell)}$. Hence, $\tilde{\mathbf{U}}_1^{(\ell)} = \mathbf{X}^{(\ell)} \mathbf{O}^{(\ell)}$. The strategy of our proof is to show that, as $N \rightarrow +\infty$, Problem (14) has the same solutions as Problem (15), which are known.

With the decomposition $\mathcal{J} = \mathcal{P} + \frac{1}{\sqrt{N}} \mathcal{N}$, we have,

$$\begin{aligned} \left\| \mathbf{U}^{(\ell)\top} \mathbf{T}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 &= \left\| \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 + \frac{1}{N} \left\| \mathbf{U}^{(\ell)\top} \mathbf{N}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &\quad + \frac{2}{\sqrt{N}} \left\langle \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}, \mathbf{U}^{(\ell)\top} \mathbf{N}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\rangle_{\mathbb{F}}. \end{aligned}$$

From Lemma 13, $\frac{1}{N} \|\mathbf{U}^{(\ell)\top} \mathbf{N}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2 = \mathcal{O}(1)$ almost surely and

$$\begin{aligned} \frac{2}{\sqrt{N}} \left| \left\langle \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}, \mathbf{U}^{(\ell)\top} \mathbf{N}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\rangle_{\mathbb{F}} \right| \\ \leq \frac{2}{\sqrt{N}} \underbrace{\left\| \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}}_{=\mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}})} \underbrace{\left\| \mathbf{U}^{(\ell)\top} \mathbf{N}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}}_{=\mathcal{O}(\sqrt{N})}. \end{aligned}$$

Therefore, for all $\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$,

$$\left\| \mathbf{U}^{(\ell)\top} \mathbf{T}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 = \left\| \mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 + \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}}) \quad \text{almost surely.} \quad (16)$$

In particular,

$$\begin{aligned} \left\| \mathbf{U}_1^{(\ell)\top} \mathbf{T}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 &= \left\| \mathbf{U}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 + \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}}) \quad \text{almost surely,} \\ \left\| \mathbf{U}_1^{(\ell)\top} \mathbf{T}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 &= \left\| \tilde{\mathbf{U}}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 + \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}}) \quad \text{almost surely,} \end{aligned}$$

where the first equation is simply Equation (16) with $\mathbf{U}^{(\ell)} = \mathbf{U}_1^{(\ell)}$ and the second equation stems from the maximum over $\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$ of both sides of Equation (16)²². Hence,

$$\left\| \mathbf{U}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 = \left\| \tilde{\mathbf{U}}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 + \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}}) \quad \text{almost surely.} \quad (17)$$

Then, consider the singular value decomposition $\mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)} = \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^{(\ell)} \mathbf{v}_{q_\ell}^{(\ell)} \mathbf{w}_{q_\ell}^{(\ell)\top}$.

$$\begin{aligned} \left\| \mathbf{U}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 &= \left\| \mathbf{U}_1^{(\ell)\top} \mathbf{X}^{(\ell)} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &= \left\| \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^{(\ell)} \mathbf{w}_{q_\ell}^{(\ell)} \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &= \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^{(\ell)2} \left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &= \sum_{q_\ell=1}^{r_\ell} \left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 - \sum_{q_\ell=1}^{r_\ell} \left(1 - s_{q_\ell}^{(\ell)2}\right) \left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 \\ &= \left\| \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 - \sum_{q_\ell=1}^{r_\ell} \left(1 - s_{q_\ell}^{(\ell)2}\right) \left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2. \end{aligned}$$

22. $\max\{\|\mathbf{U}^{(\ell)\top} \mathbf{T}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2\} \leq \max\{\|\mathbf{U}^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2\} + \max\{\mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}})\}$ where each max is over $\mathbf{U}^{(\ell)} \in V_{r_\ell}(\mathbb{R}^{n_\ell})$. Thus, $\|\mathbf{U}_1^{(\ell)\top} \mathbf{T}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2 - \|\tilde{\mathbf{U}}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \bigotimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2 = \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}})$ almost surely.

Therefore, because $\|\mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2 = \|\tilde{\mathbf{U}}_1^{(\ell)\top} \mathbf{P}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{U}_0^{(\ell')}\|_{\mathbb{F}}^2$, Equation (17) yields,

$$\sum_{q_\ell=1}^{r_\ell} \left(1 - s_{q_\ell}^{(\ell)2}\right) \left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 = \mathcal{O}(\|\mathcal{P}\|_{\mathbb{F}}) \quad \text{almost surely.}$$

Using the decomposition $\mathbf{v}_{q_\ell}^{(\ell)} = \sum_{q'_\ell=1}^{r_\ell} [\mathbf{v}_{q_\ell}^{(\ell)}]_{q'_\ell} \mathbf{X}^{(\ell)\top} \mathbf{x}_{q'_\ell}^{(\ell)}$, we can see that

$$\left\| \mathbf{v}_{q_\ell}^{(\ell)\top} \mathbf{H}^{(\ell)} \boxtimes_{\ell' \neq \ell} \mathbf{X}^{(\ell')\top} \mathbf{U}_0^{(\ell')} \right\|_{\mathbb{F}}^2 = \sum_{q'_\ell=1}^{r_\ell} [\mathbf{v}_{q_\ell}^{(\ell)}]_{q'_\ell}^2 \left\| \mathcal{P}(\mathbf{U}_0^{(1)}, \dots, \mathbf{x}_{q'_\ell}^{(\ell)}, \dots, \mathbf{U}_0^{(d)}) \right\|_{\mathbb{F}}^2 = \Theta(L_N^2).$$

Hence,

$$\sum_{q_\ell=1}^{r_\ell} \left(1 - s_{q_\ell}^{(\ell)2}\right) = \mathcal{O}\left(\frac{\|\mathcal{P}\|_{\mathbb{F}}}{L_N^2}\right) \quad \text{almost surely,}$$

which is the result stated in Theorem 14:

$$\frac{1}{r_\ell} \left\| \mathbf{X}^{(\ell)\top} \mathbf{U}_1^{(\ell)} \right\|_{\mathbb{F}}^2 = \frac{1}{r_\ell} \sum_{q_\ell=1}^{r_\ell} s_{q_\ell}^{(\ell)2} = 1 + \mathcal{O}\left(\frac{\|\mathcal{P}\|_{\mathbb{F}}}{L_N^2}\right) \quad \text{almost surely.}$$

References

- Karim M. Abadir and Jan R. Magnus. *Matrix Algebra*. Econometric Exercises. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-53746-9. doi: 10.1017/CBO9780511810800. URL <https://www.cambridge.org/core/books/matrix-algebra/BCE8FD2D62006D4061F88E02615B5622>.
- Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, April 2009. ISBN 978-1-4008-3024-4. doi: 10.1515/9781400830244. URL <https://www.degruyter.com/document/doi/10.1515/9781400830244/html>. Publication Title: Optimization Algorithms on Matrix Manifolds.
- Evrin Acar, Canan Aykut-Bingol, Haluk Bingol, Rasmus Bro, and Bülent Yener. Multiway analysis of epilepsy tensors. *Bioinformatics (Oxford, England)*, 23(13):i10–18, July 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btm210.
- Animashree Anandkumar, Rong Ge, Daniel Hsu, and Sham Kakade. A Tensor Spectral Approach to Learning Mixed Membership Community Models. In *Proceedings of the 26th Annual Conference on Learning Theory*, pages 867–881. PMLR, June 2013. URL <https://proceedings.mlr.press/v30/Anandkumar13.html>. ISSN: 1938-7228.
- Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor Decompositions for Learning Latent Variable Models. *Journal of Machine Learning Research*, 15(80):2773–2832, 2014. ISSN 1533-7928. URL <http://jmlr.org/papers/v15/anandkumar14b.html>.

Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

Zhidong Bai and Yanqing Yin. Convergence to the Semicircle Law. *The Annals of Probability*, 16(2):863–875, April 1988. ISSN 0091-1798, 2168-894X. doi: 10.1214/aop/1176991792. URL [Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33\(5\):1643–1697, September 2005. ISSN 0091-1798, 2168-894X. doi: 10.1214/009117905000000233. URL \[Afonso S. Bandeira, Amelia Perry, and Alexander S. Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *Portugaliae Mathematica*, 75\\(2\\):159–186, December 2018. ISSN 0032-5155. doi: 10.4171/pm/2014. URL <https://ems.press/journals/pm/articles/15908>.\]\(https://projecteuclid.org/journals/annals-of-probability/volume-33/issue-5/Phase-transition-Publisher: Institute of Mathematical Statistics.</p>
</div>
<div data-bbox=\)](https://projecteuclid.org/journals/annals-of-probability/volume-16/issue-2/Convergence-to-Publisher: Institute of Mathematical Statistics.</p>
</div>
<div data-bbox=)

Marwa Banna, Florence Merlevède, and Magda Peligrad. On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Processes and their Applications*, 125(7):2700–2726, July 2015. ISSN 0304-4149. doi: 10.1016/j.spa.2015.01.010. URL <https://www.sciencedirect.com/science/article/pii/S0304414915000290>.

Gérard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The Landscape of the Spiked Tensor Model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019. ISSN 1097-0312. doi: 10.1002/cpa.21861. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21861>.

Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor PCA. *The Annals of Probability*, 48(4):2052–2087, July 2020. ISSN 0091-1798, 2168-894X. doi: 10.1214/19-AOP1415. URL [Gérard Ben Arous, Daniel Zhengyu Huang, and Jiaoyang Huang. Long random matrices and tensor unfolding. *The Annals of Applied Probability*, 33\(6B\):5753–5780, December 2023. ISSN 1050-5164, 2168-8737. doi: 10.1214/23-AAP1958. URL \[Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227\\(1\\):494–521, 2011. ISSN 0001-8708. doi: 10.1016/j.aim.2011.02.007. URL <https://www.sciencedirect.com/science/article/pii/S0001870811000570>.\]\(https://projecteuclid.org/journals/annals-of-applied-probability/volume-33/issue-6B/Long-random-matrices-and-tensor-unfolding-Publisher: Institute of Mathematical Statistics.</p>
</div>
<div data-bbox=\)](https://projecteuclid.org/journals/annals-of-probability/volume-48/issue-4/Algorithmic-thresholds-for-tensor-PCA-Publisher: Institute of Mathematical Statistics.</p>
</div>
<div data-bbox=)

Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, October 2012. ISSN 0047-259X. doi: 10.1016/j.jmva.2012.04.019. URL <https://www.sciencedirect.com/science/article/pii/S0047259X12001108>.

Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2012.

Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973. ISSN 0025-5718, 1088-6842. doi: 10.1090/S0025-5718-1973-0348991-3. URL <https://www.ams.org/mcom/1973-27-123/S0025-5718-1973-0348991-3/>.

Rasmus Bro and Claus A. Andersson. Improving the speed of multiway algorithms: Part II: Compression. *Chemometrics and Intelligent Laboratory Systems*, 42(1):105–113, August 1998. ISSN 0169-7439. doi: 10.1016/S0169-7439(98)00011-2. URL <https://www.sciencedirect.com/science/article/pii/S0169743998000112>.

Hongbin Chen, Jean-Christophe Mourrat, and Jiaming Xia. Statistical inference of finite-rank tensors. *Annales Henri Lebesgue*, 5:1161–1189, 2022. ISSN 2644-9463. doi: 10.5802/ahl.146. URL <https://ahl.centre-mersenne.org/articles/10.5802/ahl.146/>.

Louis H. Y. Chen. An inequality for the multivariate normal distribution. *Journal of Multivariate Analysis*, 12(2):306–315, June 1982. ISSN 0047-259X. doi: 10.1016/0047-259X(82)90022-7. URL <https://www.sciencedirect.com/science/article/pii/0047259X82900227>.

Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman. Phase transition in random tensors with multiple independent spikes. *The Annals of Applied Probability*, 31(4):1868–1913, August 2021. ISSN 1050-5164, 2168-8737. doi: 10.1214/20-AAP1636. URL <https://projecteuclid.org/journals/annals-of-applied-probability/volume-31/issue-4/Phase-t>
Publisher: Institute of Mathematical Statistics.

Antoine Chevreuil and Philippe Loubaton. On The Non-Detectability of Spiked Large Random Tensors. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 443–447, June 2018. doi: 10.1109/SSP.2018.8450752. URL <https://ieeexplore.ieee.org/document/8450752>.

Yasuko Chikuse. *Statistics on Special Manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer, New York, NY, 2003. ISBN 978-0-387-00160-9 978-0-387-21540-2. doi: 10.1007/978-0-387-21540-2. URL <http://link.springer.com/10.1007/978-0-387-21540-2>.

Andrzej Cichocki, Danilo P. Mandic, Anh Huy Phan, Cesar F. Caiafa, Guoxu Zhou, Qibin Zhao, and Lieven De Lathauwer. Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, March 2015. ISSN 1558-0792. doi: 10.1109/MSP.2013.2297439. URL <https://ieeexplore.ieee.org/document/7038247>. Conference Name: IEEE Signal Processing Magazine.

Pierre Comon. Tensors versus Matrices, usefulness and unexpected properties. In IEEE, editor, *IEEE Workshop on Statistical Signal Processing*, pages 780–788, Cardiff, United Kingdom, September 2009. IEEE. URL <https://hal.archives-ouvertes.fr/hal-00417258>.

Pierre Comon. Tensors: a Brief Introduction. *IEEE Signal Processing Magazine*, 31(3):44–53, May 2014. doi: 10.1109/MSP.2014.2298533. URL <https://hal.archives-ouvertes.fr/hal-00923279>. Publisher: Institute of Electrical and Electronics Engineers.

Romain Couillet and Zhenyu Liao. *Random Matrix Methods for Machine Learning*. Cambridge University Press, Cambridge, 2022. ISBN 978-1-00-912323-5. doi: 10.1017/9781009128490. URL <https://www.cambridge.org/core/books/random-matrix-methods-for-machine-learning/6B681EB69E>.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the Best Rank-1 and Rank-(R1, R2, ..., RN) Approximation of Higher-Order Tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, January 2000a. ISSN 0895-4798. doi: 10.1137/S0895479898346995. URL <https://epubs.siam.org/doi/10.1137/S0895479898346995>. Publisher: Society for Industrial and Applied Mathematics.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A Multilinear Singular Value Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, January 2000b. ISSN 0895-4798. doi: 10.1137/S0895479896305696. URL <https://epubs.siam.org/doi/10.1137/S0895479896305696>. Publisher: Society for Industrial and Applied Mathematics.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, September 1936. ISSN 1860-0980. doi: 10.1007/BF02288367. URL <https://doi.org/10.1007/BF02288367>.

Samuel F. Edwards and Raymund C. Jones. The eigenvalue spectrum of a large symmetric random matrix. *Journal of Physics A: Mathematical and General*, 9(10):1595, October 1976. ISSN 0305-4470. doi: 10.1088/0305-4470/9/10/011. URL <https://dx.doi.org/10.1088/0305-4470/9/10/011>.

Hadi Fanaee-T and João Gama. EigenEvent: An Algorithm for Event Detection from Complex Data Streams in Syndromic Surveillance. *Intelligent Data Analysis*, 19(3):597–616, June 2015. ISSN 1088467X, 15714128. doi: 10.3233/IDA-150734. URL <http://arxiv.org/abs/1406.3496>. arXiv:1406.3496 [cs, stat].

Michael J. Feldman. Spiked singular values and vectors under extreme aspect ratios. *Journal of Multivariate Analysis*, 196:105187, July 2023. ISSN 0047-259X. doi: 10.1016/j.jmva.2023.105187. URL <https://www.sciencedirect.com/science/article/pii/S0047259X23000337>.

Michael J. Feldman and David L. Donoho. Sharp Recovery Thresholds of Tensor PCA Spectral Algorithms. *Advances in Neural Information Processing Systems*, 36:56628–56640, December 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/b14d76c7266be21b338527cd25deac45.

Evgeny Frolov and Ivan Oseledets. Tensor methods and recommender systems. *WIREs Data Mining and Knowledge Discovery*, 7(3):e1201, 2017. ISSN 1942-4795. doi: 10.1002/widm.1201. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1201>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1201>.

Delphine Féral and Sandrine Péché. The Largest Eigenvalue of Rank One Deformation of Large Wigner Matrices. *Communications in Mathematical Physics*, 272(1):185–228, May 2007. ISSN 1432-0916. doi: 10.1007/s00220-007-0209-3. URL <https://doi.org/10.1007/s00220-007-0209-3>.

Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, September 1981. ISSN 1439-6912. doi: 10.1007/BF02579329. URL <https://doi.org/10.1007/BF02579329>.

José Henrique de M. Goulart, Romain Couillet, and Pierre Comon. A Random Matrix Perspective on Random Tensors. *Journal of Machine Learning Research*, 23(264):1–36, 2022. ISSN 1533-7928. URL <http://jmlr.org/papers/v23/21-1038.html>.

Lars Grasedyck, Daniel Kressner, and Christine Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013. ISSN 1522-2608. doi: 10.1002/gamm.201310004. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/gamm.201310004>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/gamm.201310004>.

Razvan Gurau. Universality for random tensors. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(4):1474–1525, November 2014. ISSN 0246-0203. doi: 10.1214/13-AIHP567. URL <https://projecteuclid.org/journals/Annales-de-linstitut-henri-poincare-probabilites-et-statistique/Publisher: Institut Henri Poincaré>.

Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics. Springer, 2012. URL <https://link.springer.com/book/10.1007/978-3-642-28027-6>.

Christopher J. Hillar and Lek-Heng Lim. Most Tensor Problems Are NP-Hard. *Journal of the ACM*, 60(6):45:1–45:39, November 2013. ISSN 0004-5411. doi: 10.1145/2512329. URL <https://doi.org/10.1145/2512329>.

Aicke Hinrichs, Joscha Prochno, and Jan Vybíral. Entropy numbers of embeddings of Schatten classes. *Journal of Functional Analysis*, 273(10):3241–3261, November 2017. ISSN 0022-1236. doi: 10.1016/j.jfa.2017.08.008. URL <https://www.sciencedirect.com/science/article/pii/S0022123617303221>.

Frank L. Hitchcock. The Expression of a Tensor or a Polyadic as a Sum of Products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927. ISSN 1467-9590. doi: 10.1002/sapm192761164. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm192761164>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sapm192761164>.

- Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*, pages 956–1006. PMLR, June 2015. URL <https://proceedings.mlr.press/v40/Hopkins15.html>. ISSN: 1938-7228.
- Samuel B. Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, STOC '16, pages 178–191, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4132-5. doi: 10.1145/2897518.2897529. URL <https://doi.org/10.1145/2897518.2897529>.
- Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The Power of Sum-of-Squares for Detecting Hidden Structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE Computer Society, October 2017. ISBN 978-1-5386-3464-6. doi: 10.1109/FOCS.2017.72. URL <https://www.computer.org/csdl/proceedings-article/focs/2017/3464a720/120mNxFsmCD>. ISSN: 0272-5428.
- Jiaoyang Huang, Daniel Z. Huang, Qing Yang, and Guang Cheng. Power Iteration for Tensor PCA. *Journal of Machine Learning Research*, 23(128):1–47, 2022. ISSN 1533-7928. URL <http://jmlr.org/papers/v23/21-1290.html>.
- Borbála Hunyadi, Patrick Dupont, Wim Van Paesschen, and Sabine Van Huffel. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *WIREs Data Mining and Knowledge Discovery*, 7(1):e1197, 2017. ISSN 1942-4795. doi: 10.1002/widm.1197. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1197>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1197>.
- Aukosh Jagannath, Patrick Lopatto, and Léo Miolane. Statistical thresholds for tensor PCA. *The Annals of Applied Probability*, 30(4):1910–1933, August 2020. ISSN 1050-5164, 2168-8737. doi: 10.1214/19-AAP1547. URL <https://projecteuclid.org/journals/annals-of-applied-probability/volume-30/issue-4/Statist>. Publisher: Institute of Mathematical Statistics.
- Jonathan Kadmon and Surya Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://papers.nips.cc/paper_files/paper/2018/hash/b3848d61bbbc6207c6668a8a9e2730ed-Abstract.
- Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma. Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517, December 2018. ISSN 1941-0476. doi: 10.1109/TSP.2018.2876362. URL <https://ieeexplore.ieee.org/abstract/document/8494792>. Conference Name: IEEE Transactions on Signal Processing.

- Arie Kapteyn, Heinz Neudecker, and Tom Wansbeek. An approach ton-mode components analysis. *Psychometrika*, 51(2):269–275, June 1986. ISSN 1860-0980. doi: 10.1007/BF02293984. URL <https://doi.org/10.1007/BF02293984>.
- Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. Multi-verse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems, RecSys '10*, pages 79–86, New York, NY, USA, September 2010. Association for Computing Machinery. ISBN 978-1-60558-906-0. doi: 10.1145/1864708.1864727. URL <https://doi.org/10.1145/1864708.1864727>.
- Chiheon Kim, Afonso S. Bandeira, and Michel X. Goemans. Community detection in hypergraphs, spiked tensor models, and Sum-of-Squares. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages 124–128, July 2017. doi: 10.1109/SAMPTA.2017.8024470. URL <https://ieeexplore.ieee.org/document/8024470>.
- Tamara G. Kolda. A Counterexample to the Possibility of an Extension of the Eckart–Young Low-Rank Approximation Theorem for the Orthogonal Rank Tensor Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 24(3):762–767, January 2003. ISSN 0895-4798. doi: 10.1137/S0895479801394465. URL <https://epubs.siam.org/doi/10.1137/S0895479801394465>. Publisher: Society for Industrial and Applied Mathematics.
- Tamara G. Kolda and Brett W. Bader. Tensor Decompositions and Applications. *SIAM Review*, August 2009. doi: 10.1137/07070111X. URL <https://epubs.siam.org/doi/10.1137/07070111X>. Publisher: Society for Industrial and Applied Mathematics.
- Pieter M. Kroonenberg and Jan de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, March 1980. ISSN 1860-0980. doi: 10.1007/BF02293599. URL <https://doi.org/10.1007/BF02293599>.
- Joseph M. Landsberg. *Tensors: Geometry and Applications*, volume 128 of *Graduate Studies in Mathematics*. American Mathematical Society, December 2011. ISBN 978-0-8218-6907-9 978-0-8218-8481-2 978-0-8218-8483-6 978-1-4704-0923-4. doi: 10.1090/gsm/128. URL <http://www.ams.org/gsm/128>. ISSN: 1065-7339.
- Michel Ledoux. *The Concentration of Measure Phenomenon*. Number 89 in *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.
- Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515, June 2017. doi: 10.1109/ISIT.2017.8006580. URL <http://arxiv.org/abs/1701.08010>. arXiv:1701.08010 [cond-mat, stat].

Nan Li and Baoxin Li. Tensor completion for on-board compression of hyperspectral images. In *2010 IEEE International Conference on Image Processing*, pages 517–520, September 2010. doi: 10.1109/ICIP.2010.5651225. ISSN: 2381-8549.

Tianqi Liu, Ming Yuan, and Hongyu Zhao. Characterizing Spatiotemporal Transcriptome of the Human Brain Via Low-Rank Tensor Decomposition. *Statistics in Biosciences*, 14(3):485–513, December 2022. ISSN 1867-1772. doi: 10.1007/s12561-021-09331-5. URL <https://doi.org/10.1007/s12561-021-09331-5>.

Anna Lytova and Leonid Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *The Annals of Probability*, 37(5):1778–1840, 2009. Publisher: Institute of Mathematical Statistics.

Vladimir A. Marčenko and Leonid A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967. Publisher: IOP Publishing.

Florence Merlevède, Costel Peligrad, and Magda Peligrad. On the universality of spectral limit for random matrices with martingale differences entries. *Random Matrices: Theory and Applications*, 04(01):1550003, January 2015. ISSN 2010-3263. doi: 10.1142/S2010326315500033. URL <https://www.worldscientific.com/doi/abs/10.1142/S2010326315500033>. Publisher: World Scientific Publishing Co.

Leonid Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, January 1960. ISSN 0033-5606. doi: 10.1093/qmath/h/11.1.50. URL <https://doi.org/10.1093/qmath/11.1.50>.

Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL <https://proceedings.neurips.cc/paper/2014/hash/b5488aef42889188d03c9895255cecc-Abstract.html>.

Chaitanya Muralidhara, Andrew M. Gross, Robin R. Gutell, and Orly Alter. Tensor Decomposition Reveals Concurrent Evolutionary Convergences and Divergences and Correlations with Structural Motifs in Ribosomal RNA. *PLOS ONE*, 6(4):e18768, 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0018768. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018768>. Publisher: Public Library of Science.

Larsson Omberg, Gene H. Golub, and Orly Alter. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proceedings of the National Academy of Sciences*, 104(47):18371–18376, November 2007. doi: 10.1073/pnas.0709146104. URL <https://www.pnas.org/doi/full/10.1073/pnas.0709146104>. Publisher: Proceedings of the National Academy of Sciences.

Larsson Omberg, Joel R. Meyerson, Kayta Kobayashi, Lucy S. Drury, John F. X. Diffley, and Orly Alter. Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. *Molecular Systems Biology*, 5

- (1):312, January 2009. ISSN 1744-4292. doi: 10.1038/msb.2009.70. URL <https://www.embopress.org/doi/full/10.1038/msb.2009.70>. Publisher: John Wiley & Sons, Ltd.
- Leonid Andreevich Pastur and Mariya Shcherbina. *Eigenvalue Distribution of Large Random Matrices*. Number 171 in Mathematical Surveys and Monographs. American Mathematical Society, 2011.
- Amelia Perry, Alexander S. Wein, and Afonso S. Bandeira. Statistical limits of spiked tensor models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):230–264, February 2020. ISSN 0246-0203. doi: 10.1214/19-AIHP960. URL <https://projecteuclid.org/journals/Annales-de-l'institut-henri-poincare-probabilites-et-statistiques>. Publisher: Institut Henri Poincaré.
- Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory*. Cambridge University Press, 2020.
- Sandrine Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, January 2006. ISSN 1432-2064. doi: 10.1007/s00440-005-0466-z. URL <https://doi.org/10.1007/s00440-005-0466-z>.
- Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. Introduction to Tensor Decompositions and their Applications in Machine Learning. *arXiv:1711.10781 [cs, stat]*, November 2017. URL <http://arxiv.org/abs/1711.10781>. arXiv: 1711.10781.
- Neil C. Rabinowitz, Robbe L. Goris, Marlene Cohen, and Eero P. Simoncelli. Attention stabilizes the shared gain of V4 populations. *eLife*, 4:e08998, November 2015. ISSN 2050-084X. doi: 10.7554/eLife.08998. URL <https://doi.org/10.7554/eLife.08998>. Publisher: eLife Sciences Publications, Ltd.
- Steffen Rendle and Lars Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining, WSDM '10*, pages 81–90, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 978-1-60558-889-6. doi: 10.1145/1718487.1718498. URL <https://doi.org/10.1145/1718487.1718498>.
- Valentina Ros, Gerard Ben Arous, Giulio Biroli, and Chiara Cammarota. Complex Energy Landscapes in Spiked-Tensor and Simple Glassy Models: Ruggedness, Arrangements of Local Minima, and Phase Transitions. *Physical Review X*, 9(1):011003, January 2019. doi: 10.1103/PhysRevX.9.011003. URL <https://link.aps.org/doi/10.1103/PhysRevX.9.011003>. Publisher: American Physical Society.
- Berkant Savas and Lars Eldén. Handwritten digit classification using higher order singular value decomposition. *Pattern Recognition*, 40(3):993–1003, March 2007. ISSN 0031-3203. doi: 10.1016/j.patcog.2006.08.004. URL <https://www.sciencedirect.com/science/article/pii/S0031320306003542>.

Mohamed El Amine Seddik, Maxime Guillaud, and Romain Couillet. When Random Tensors meet Random Matrices, November 2022. URL <http://arxiv.org/abs/2112.12348>. arXiv:2112.12348 [math, stat].

Mohamed El Amine Seddik, Malik Tiomoko, Alexis Decurninge, Maxim Panov, and Maxime Gauillaud. Learning from Low Rank Tensor Data: A Random Tensor Theory Perspective. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 1858–1867. PMLR, July 2023. URL <https://proceedings.mlr.press/v216/seddik23a.html>. ISSN: 2640-3498.

Jeffrey S. Seely, Matthew T. Kaufman, Stephen I. Ryu, Krishna V. Shenoy, John P. Cunningham, and Mark M. Churchland. Tensor Analysis Reveals Distinct Population Structure that Parallels the Different Computational Roles of Areas M1 and V1. *PLOS Computational Biology*, 12(11):e1005164, November 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005164. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005164>. Publisher: Public Library of Science.

Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. Tensor Decomposition for Signal Processing and Machine Learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, July 2017. ISSN 1941-0476. doi: 10.1109/TSP.2017.2690524. Conference Name: IEEE Transactions on Signal Processing.

Charles M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9(6):1135–1151, November 1981. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176345632. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-6/Estimation-of-the> Publisher: Institute of Mathematical Statistics.

Gilbert Wright Stewart and Ji-guang Sun. *Matrix perturbation theory*. Computer science and scientific computing. Academic Press, Boston, 1990. ISBN 978-0-12-670230-9. URL <http://catdir.loc.gov/catdir/toc/els031/90033378.html>. OCLC: 21227976.

Will Wei Sun, Botao Hao, and Lexin Li. Tensors in Modern Statistical Learning. In *Wiley StatsRef: Statistics Reference Online*, pages 1–25. John Wiley & Sons, Ltd, 2021. ISBN 978-1-118-44511-2. doi: 10.1002/9781118445112.stat08319. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08319>.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat08319>.

Terence Tao. *Topics in Random Matrix Theory*. Number 132 in Graduate Studies in Mathematics. American Mathematical Society, 2012.

Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors, July 2014. URL <http://arxiv.org/abs/1407.1870>. arXiv:1407.1870 [math, stat].

Ledyard R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, September 1966. ISSN 1860-0980. doi: 10.1007/BF02289464. URL <https://doi.org/10.1007/BF02289464>.

- M. Alex O. Vasilescu. Human motion signatures: analysis, synthesis, recognition. In *2002 International Conference on Pattern Recognition*, volume 3, pages 456–460 vol.3, August 2002. doi: 10.1109/ICPR.2002.1047975. ISSN: 1051-4651.
- M. Alex O. Vasilescu and Demetri Terzopoulos. Multilinear subspace analysis of image ensembles. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–93, June 2003. doi: 10.1109/CVPR.2003.1211457. ISSN: 1063-6919.
- Nico Vervliet, Otto Debals, Laurent Sorber, and Lieven De Lathauwer. Breaking the Curse of Dimensionality Using Decompositions of Incomplete Tensors: Tensor-based scientific computing in big data analysis. *IEEE Signal Processing Magazine*, 31(5):71–79, September 2014. ISSN 1558-0792. doi: 10.1109/MSP.2014.2329429. Conference Name: IEEE Signal Processing Magazine.
- Nico Vervliet, Otto Debals, Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Tensorlab 3.0, March 2016. URL <https://www.tensorlab.net/>.
- Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi Hierarchy and Tensor PCA. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE Computer Society, November 2019. ISBN 978-1-72814-952-3. doi: 10.1109/FOCS.2019.000-2. URL <https://www.computer.org/csdl/proceedings-article/focs/2019/495200b446/1grNBDzixFe>.
- Eugene P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. ISSN 0003-486X. doi: 10.2307/1970079. URL <https://www.jstor.org/stable/1970079>. Publisher: Annals of Mathematics.
- Eugene P. Wigner. On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics*, 67(2):325–327, 1958. ISSN 0003-486X. doi: 10.2307/1970008. URL <https://www.jstor.org/stable/1970008>. Publisher: Annals of Mathematics.
- Alex H. Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I. Ryu, Krishna V. Shenoy, Mark Schnitzer, Tamara G. Kolda, and Surya Ganguli. Un-supervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron*, 98(6):1099–1115.e8, June 2018. ISSN 0896-6273. doi: 10.1016/j.neuron.2018.05.015. URL <https://www.sciencedirect.com/science/article/pii/S0896627318303878>.
- Yangyang Xu. On the convergence of higher-order orthogonal iteration. *Linear and Multilinear Algebra*, 66(11):2247–2265, November 2018. ISSN 0308-1087. doi: 10.1080/03081087.2017.1391743. URL <https://doi.org/10.1080/03081087.2017.1391743>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/03081087.2017.1391743>.
- Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, September 2016. ISSN 0001-8732, 1460-6976. doi: 10.1080/00018732.2016.1211393. URL <http://arxiv.org/abs/1511.02476>. arXiv:1511.02476 [cond-mat, stat].

Anru Zhang and Dong Xia. Tensor SVD: Statistical and Computational Limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, November 2018. ISSN 1557-9654. doi: 10.1109/TIT.2018.2841377. URL <https://ieeexplore.ieee.org/document/8368145>. Conference Name: IEEE Transactions on Information Theory.

Liangpei Zhang, Lefei Zhang, Dacheng Tao, and Xin Huang. Tensor Discriminative Locality Alignment for Hyperspectral Image Spectral–Spatial Feature Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):242–256, January 2013. ISSN 1558-0644. doi: 10.1109/TGRS.2012.2197860. URL <https://ieeexplore.ieee.org/abstract/document/6213108>. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Hua Zhou, Lexin Li, and Hongtu Zhu. Tensor Regression with Applications in Neuroimaging Data Analysis. *Journal of the American Statistical Association*, 108(502): 540–552, June 2013. ISSN 0162-1459. doi: 10.1080/01621459.2013.776499. URL <https://doi.org/10.1080/01621459.2013.776499>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01621459.2013.776499>.