The Annals of Statistics
2022, Vol. 50, No. 2, 1220–1245
https://doi.org/10.1214/21-AOS2146

INFERENCE FOR LOW-RANK TENSORS—NO NEED TO DEBIAS

BY DONG XIA[1,a], ANRU R. ZHANG[2,b] AND YUCHEN ZHOU[3,c]

[1]*Department of Mathematics, Hong Kong University of Science and Technology,* [a]*madxia@ust.hk*

[2]*Departments of Biostatistics & Bioinformatics, Computer Science, Mathematics, and Statistical Science, Duke University,* [b]*anru.zhang@duke.edu*

[3]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania,* [c]*yczhou@wharton.upenn.edu*

In this paper, we consider the statistical inference for several low-rank tensor models. Specifically, in the Tucker low-rank tensor PCA or regression model, provided with any estimates achieving some attainable error rate, we develop the data-driven confidence regions for the singular subspace of the parameter tensor based on the asymptotic distribution of an updated estimate by two-iteration alternating minimization. The asymptotic distributions are established under some essential conditions on the signal-to-noise ratio (in PCA model) or sample size (in regression model). If the parameter tensor is further orthogonally decomposable, we develop the methods and nonasymptotic theory for inference on each individual singular vector. For the rank-one tensor PCA model, we establish the asymptotic distribution for general linear forms of principal components and confidence interval for each entry of the parameter tensor. Finally, numerical simulations are presented to corroborate our theoretical discoveries.

In all of these models, we observe that different from many matrix/vector settings in existing work, debiasing is not required to establish the asymptotic distribution of estimates or to make statistical inference on low-rank tensors. In fact, due to the widely observed statistical-computational-gap for low-rank tensor estimation, one usually requires stronger conditions than the statistical (or information-theoretic) limit to ensure the computationally feasible estimation is achievable. Surprisingly, such conditions "incidentally" render a feasible low-rank tensor inference without debiasing.

1. Introduction. An mth order tensor is a multiway array along m directions. Recent years have witnessed a fast growing demand for the collection, processing and analysis of data in the form of tensors. These tensor data commonly arise, to name a few, when features are collected from different domains, or when multiple data copies are provided by various agents or sources. For instances, the worldwide food trading flows [25, 38] produce a fourth-order tensor (countries × countries × food × years); the online click-through data [32, 60] in e-commerce form a third-order tensor (users × categories × periods); Berkeley human mortality data [67, 76] yield a third-order tensor (ages × years × countries). In addition, the applications of tensor also include collaborative filtering [39, 57], recommender system design [9], computational imaging [79] and neuroimaging [84]. Researchers have made tremendous efforts to innovate effective methods for the analysis of tensor data.

Low-rank models have rendered fundamental toolkits to analyze tensor data. A tensor $\mathcal{T} \in \mathbb{R}^{p_1 \times \cdots \times p_m}$ has low Tucker rank (or multilinear rank) if all fibers[1] of $\mathcal{T}$ along different ways lie in rank-reduced subspaces of high dimension, say $\{U_j\}_{j=1}^m$, respectively [62]. The core assumption of low-rank tensor models is that the observed data is driven by an *unknown*

Received October 2020; revised June 2021.
MSC2020 subject classifications. 62H10, 62H25.
Key words and phrases. Asymptotic distribution, confidence region, tensor principal component analysis, tensor regression, statistical inference.

[1]Here, the tensor fibers are the counterpart of matrix columns and rows for tensors; see [42] for a review.

low-rank tensor $\mathcal{T}$, while the Tucker low-rank conditions can significantly reduce the model complexity. Consequently, the analysis of tensor data often boils down to the estimation and inference of the low-rank tensor $\mathcal{T}$ or its principal components based on the given data sets.

In the literature, a rich list of methods have been developed for the *estimation* of low-rank tensor $\mathcal{T}$ and the associated subspace U_j, such as alternating minimization [1], convex regularization [61, 74], power iterations [1], orthogonal iteration [23, 77], vanilla gradient descent with spectral initialization [11], projected gradient descent [18], simultaneous gradient descent [32], etc. However, in many practical scenarios, to enable more reliable decision making and prediction, it is important to quantify the estimation error in addition to point estimations. This task, referred to as *uncertainty quantification* or *statistical inference*, usually involves the construction of confidence intervals/regions for the unknown parameters through the development of the (approximate) distributions of the estimators. The statistical inference or uncertainty quantification for low-rank tensor models remains largely unexplored. In this paper, we aim to make an attempt to this fundamental and challenging problem. Our focus is on two basic yet important settings: low-rank tensor PCA and tensor regression, which we briefly summarize as follows.

Tensor principal component analysis (PCA) is among the most basic problem of unsupervised inference for low-rank tensors. We consider the tensor PCA model [1, 20, 47, 52, 55, 77], which assumes

$$\mathcal{A} = \mathcal{T} + \mathcal{Z}, \tag{1.1}$$

where the signal $\mathcal{T}$ admits a low-rank decomposition (2.1) and the noise $\mathcal{Z}$ contains i.i.d. entries with mean zero and variance σ^2. A central goal of tensor PCA is on the estimation and inference of $\mathcal{T}$ and/or $\{U_j\}_j$, that is, the low-rank structure from $\mathcal{A}$. Tensor PCA has been proven effective for learning hidden components in Gaussian mixture models [1], where $\{U_j\}_j$ represent the hidden components. By constructing confidence regions of $\{U_j\}_j$, we are able to make uncertainty quantification for the hidden components of Gaussian mixture models. In addition, confidence regions of $\{U_j\}_j$ can be useful for the inference of spatial and temporal patterns of gene regulation during brain development [47]. When applying the tensor PCA model to community detection in hypergraph networks [40] or multilayer networks [38], U_j is directly related to the estimated community structures and the confidence region of U_j is an important tool to quantify the uncertainty of community detection. This also applies to the uncertainty quantification for tensor/high-order clustering [31, 48].

Low-rank tensor regression can be seen as one of the most basic setting of *supervised inference* for low-rank tensors. Specifically, suppose we observe a set of random pairs $\{\mathcal{X}_i, Y_i\}_{i=1}^n$ associated as

$$Y_i = \langle \mathcal{T}, \mathcal{X}_i \rangle + \xi_i. \tag{1.2}$$

Here, the main point of interest is $\mathcal{T}$, a low-rank tensor that characterizes the association between response Y and covariate $\mathcal{X}$, and ξ_i is the noise term. When the tensor order is $m = 2$, this problem is reduced to the widely studied *trace matrix regression model* in the literature [14, 15, 18, 28, 43, 44, 53, 54, 61]. This model can also be used as the prototype of many problems in high-dimensional statistics and machine learning, including phase retrieval [16] and blind deconvolution [45]. When $m \geq 3$, this problem has been studied under the scenario of high-order interaction pursuit [33] and large-scale linear system from partial differential equations [50]. In applications of tensor regression to neuroimaging analysis, the principal components of $\mathcal{T}$ are useful in the understanding of the association between disease outcomes and brain image patterns [84]. In addition, the principal components determine the cluster memberships of neuroimaging data [59]. Confidence regions of $\{U_j\}_j$ in the aforementioned

applications allow us to make significance test for the detected regions of interest, and to make uncertainty quantification for clustering outcomes, respectively.

In addition to tensor PCA and regression, there is a broad range of low-rank tensor models, such as tensor completion [51, 71, 74, 75], generalized tensor estimation [32] and tensor high-order clustering [22, 29, 31, 48, 59, 68]. A common goal of these problems is to accurately estimate and make inference on some type of low-rank structures.

1.1. *Summary of the main results*. In this paper, we aim to develop the methods and nonasymptotic theory for statistical inference under the low-rank tensor PCA and regression models. First, suppose the target tensor $\mathcal{T}$ is Tucker low-rank with singular subspace U_j as the point of interest. Given any estimator $\hat{U}_j^{(0)}$ that achieves some reasonable estimation error, we introduce a straightforward two-iteration alternating minimization scheme (Algorithms 1 and 2 in Section 3.1) and obtain $\hat{U}_j$. Surprisingly, we are able to derive an asymptotic distribution of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ (definition of sin-theta distance is postponed to Section 2) even though $\hat{U}_j$ is from nonconvex iterations. Under the tensor PCA model with some essential conditions on SNR, we prove that

$$\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p_j \to \infty. \tag{1.3}$$

Here, Λ_j is the diagonal matrix containing all nonzero singular values of the jth matricization of $\mathcal{G}$ (see definition of matricization in Section 2). Under the tensor regression model with some essential conditions on sample size and SNR, we prove that

$$\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j n^{-1}\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}n^{-1}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p_j \to \infty. \tag{1.4}$$

We also develop the nonasymptotic Berry–Essen-type bounds for the limiting distributions in (1.3) and (1.4).

Then we consider a special class of *orthogonally decomposable tensors* $\mathcal{T}$ in the sense that $\mathcal{T} = \sum_{j=1}^r \lambda_j \cdot u_j \otimes v_j \otimes w_j \in \mathbb{R}^{p_1\times p_2\times p_3}$ for orthonormal vectors $\{u_j\}_j$, $\{v_j\}_j$, and $\{w_j\}_j$. The orthogonally decomposable tensor has been widely studied as a benchmark setting for tensor decomposition in the literature [4, 7, 19, 41, 56]. In addition, the (near-)orthogonally decomposable tensors have been used in various applications of statistics and machine learning, such as latent variable model [1], hidden Markov models [3], etc. Under the tensor PCA model, we prove that

$$\frac{\langle\hat{u}_j, u_j\rangle^2 - (1 - p_j\sigma^2\lambda_j^{-2})}{\sqrt{2p_j}\sigma^2\lambda_j^{-2}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p_1 \to \infty \tag{1.5}$$

for $j = 1, \ldots, r$ when some essential SNR condition holds. Here, $\{\hat{u}_j, \hat{v}_j, \hat{w}_j\}_j$ are the estimates of $\{u_j, v_j, w_j\}_j$ (up to some permutation of index j) based on a two-step power iteration (Algorithm 3). Similar results can also be obtained for $\langle\hat{v}_j, v_j\rangle^2$ and $\langle\hat{w}_j, w_j\rangle^2$.

Next, we propose the estimates of Λ_j, λ_j, σ^2 that are involved in the asymptotic distributions of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ in (1.3), (1.4) and $\langle\hat{u}_j, u_j\rangle^2$ in (1.5). We prove that the asymptotic normality in (1.3), (1.4) and (1.5) still hold after plugging in these estimates. These results immediately yield the data-driven confidence regions for U_j (Tucker low-rank settings) or $\{u_j\}_j$ (orthogonally decomposable settings).

If $\mathcal{A}$ is a rank-1 tensor, the low-rank tensor PCA model reduces to the widely studied *rank-1 tensor PCA* (see a literature survey in Section 1.2). Under this model, we establish the

asymptotic normality of any linear functionals for the power iteration estimators $\hat{u}$, $\hat{v}$, $\hat{w}$: for all unit vectors $q_i \in \mathbb{R}^{p_i}$, under regularity conditions, we have

$$\left(\frac{\langle q_1, \hat{u}-u\rangle + \frac{p_1\langle q_1,u\rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_1\langle q_1,u\rangle^2}{2(\lambda/\sigma)^4} + \frac{1-\langle q_1,u\rangle^2}{(\lambda/\sigma)^2}}}, \frac{\langle q_2, \hat{v}-v\rangle + \frac{p_2\langle q_2,v\rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_2\langle q_2,v\rangle^2}{2(\lambda/\sigma)^4} + \frac{1-\langle q_2,v\rangle^2}{(\lambda/\sigma)^2}}}, \frac{\langle q_3, \hat{w}-w\rangle + \frac{p_3\langle q_3,w\rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_3\langle q_3,w\rangle^2}{2(\lambda/\sigma)^4} + \frac{1-\langle q_3,w\rangle^2}{(\lambda/\sigma)^2}}}\right)^\top \xrightarrow{\text{d.}} N(0, I_3)$$

as $p_1, p_2, p_3 \to \infty$. We further derive the entrywise asymptotic distribution for each entry of the estimator $\hat{\mathcal{T}}$, and propose a thresholding procedure to construct the asymptotic $1-\alpha$ entrywise confidence interval for $\mathcal{T}$, which is the first of such work to our best knowledge.

Our theoretical results reveal a key message: under the tensor PCA and regression model, *the inference of principal components can be efficiently done when a computationally feasible optimal estimate is achievable.* In recent literature, it is widely observed in many low-rank tensor models (see 1.2 for a review of literature) that in order to achieve an accurate estimation in polynomial time, one often requires a more stringent condition than what is needed in the statistical (or information-theoretic) limit. Such a statistical and computational gap becomes a "blessing" to the statistical inference of low-rank tensor models, as debiasing can become unnecessary if those strong but essential conditions for computational feasibility are met.

1.2. *Related prior work.* This paper is related to a broad range of literature in high-dimensional statistics and matrix/tensor analysis. First, a variety of methods have been proposed for tensor PCA in the literature. A nonexhaustive list include high-order orthogonal iteration [23]; sequential-HOSVD [63], inference for low-rank matrix completion [21, 30], (truncated) power iteration [2, 47, 60], STAT-SVD [76]. In addition, the computational hardness was widely considered for tensor PCA. Particularly in the worse case scenario, the best low-rank approximation of tensors can be NP hard [24, 34]. The average-case computational complexity for tensor PCA model has also been widely studied under various computational models, including the sum-of-squares [35], optimization landscape [8], average-case reduction [10, 48, 49, 77] and statistical query [26]. It has now been widely justified that the SNR condition $\lambda_{\min}/\sigma \geq Cp^{3/4}$ is essential to ensure tensor PCA is solvable in polynomial time.

Regression of low-rank tensor has attracted enormous attention recently. Various methods, such as the (regularized) alternating minimization [46, 58, 84], convex regularization [53, 61], projected gradient descent [18, 54] and importance sketching [78] were studied. Recently, [32] proved that a gradient descent algorithm can recover a low-rank third-order tensor $\mathcal{T}$ with statistically optimal convergence rate when the sample size n is much greater than the tensor dimension $p^{3/2}$. It was widely conjectured that $n \geq Cp^{3/2}$ is essential for the problem being solvable in polynomial time (see [6] for the evidence).

While the statistical inference for low-rank tensor models remain largely unexplored, there have been several recent results demystifying the statistical inference for low-rank *matrix* models. For matrix PCA, [70] introduced an explicit representation formula for $\hat{U}_j\hat{U}_j^\top$. A more precise characterization of the distribution of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ was established in [5] by random matrix theory. On the other hand, the estimators of tensor PCA are often calculated from iterative optimization algorithms (e.g., power iterations or gradient descent) in existing literature, while the estimator of matrix PCA is based on noniterative schemes. Due to the complex statistical dependence involved in iterative optimization algorithms, it is significantly more challenging to analyze the asymptotic distribution of the estimator in tensor PCA than the one in matrix PCA. We also note that, when studying the asymptotic distributions of individual eigenvectors, an eigengap condition is often crucial for matrix PCA but not required for tensor PCA.

The inference and uncertainty quantification were also considered for low-rank matrix regression. For example, [17] introduced a debiased estimator based on the nuclear norm penalized low-rank estimator. [13] introduced another debiasing technique and characterize the entrywise distribution of the debiased estimator under the restricted isometry property. [69] studied a debiased estimator for matrix regression under the isotropic Gaussian design and established the distribution of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ under nearly optimal sample size conditions. All of these approaches rely on suitable debiasing of certain initial estimates. In addition to low-rank estimation, an appropriate debias was found crucial for high-dimensional sparse regression [80], and various debiasing schemes were introduced [37, 64, 81]. Interestingly, as will be shown in Section 3, our estimating and inference procedure for low-rank tensor regression does not involve debiasing.

Statistical inference for low-rank models are particularly challenging for tensor problems. In a concurrent work, [36] studied the statistical inference and power iteration for tensor PCA. Recently, [12] studied the entrywise statistical inference for noisy low-rank tensor completion based on a nearly unbiased estimator and an incoherence condition on U_js, that is, all the rows of U_j have comparable magnitudes. In comparison, our results do not require further conditions on U_js or debiasing.

1.3. *Organizations*. The rest of the paper is organized as follows. After an introduction on notation and preliminaries in Section 2, we discuss the inference for principal components under the Tucker low-rank models in Section 3. Specifically, a general two-iteration alternating minimization procedure, inference for tensor PCA, inference for tensor regression, and a proof sketch are given in Sections 3.1, 3.2, 3.3 and 3.4, respectively. In Section 4, we focus on the inference for individual singular vectors of orthogonally decomposable tensors. The asymptotic distribution and entrywise confidence interval are discussed for the rank-1 tensor PCA model in Section 5. In the Supplementary Material [73], Section A includes some algorithms for tensor PCA and regression in the literature. All proofs of the main technical results are collected in Section B.

2. Notation and preliminaries. Let $\{a_k\}$ and $\{b_k\}$ be two sequences of nonnegative numbers. We denote $a_k \ll b_k$ if $\lim_{k\to\infty} a_k/b_k = 0$ and $a_k \gg b_k$ if $\lim_{k\to\infty} a_k/b_k = \infty$. We use calligraphic letters $\mathcal{T}, \mathcal{G}$ to denote tensors, uppercase letters U, W to denote matrices and lowercase letters u, w to denote vectors or scalars. For a random variable X and $\alpha > 0$, the Orlicz ψ_α-norm of X is defined as

$$\|X\|_{\psi_\alpha} = \inf\bigl\{K > 0 : \mathbb{E}\bigl\{\exp\bigl(|X|/K\bigr)^\alpha\bigr\} \le 2\bigr\}.$$

Specifically, a random variable with finite ψ_2-norm or ψ_1-norm is called the sub-Gaussian or subexponential random variable, respectively. Let e_j denote the jth canonical basis vector whose dimension varies at different places. Let $\operatorname{rank}(\mathcal{T})$ be the Tucker rank of $\mathcal{T}$ and write $(a_1, \ldots, a_m) \le (b_1, \ldots, b_m)$ if $a_j \le b_j$ for all $j \in [m]$. We use $\|\cdot\|_{\mathrm{F}}$ for Frobenius norm, $\|\cdot\|$ for matrix spectral norm and $\|\cdot\|_q$ for vector ℓ_q-norm. Denote $\mathbb{S}^{p-1} = \{v \in \mathbb{R}^p : \|v\|_2 \le 1\}$ as the set of p-dimensional unit vectors. Define $\mathbb{O}_{p,r} = \{U \in \mathbb{R}^{p\times r} : U^\top U = I_r\}$ as the set of all p-by-r matrices with orthonormal columns. In particular, $\mathbb{O}_r$ is the set of all $r \times r$ orthogonal matrices.

We denote $\times_j$ the jth multilinear product between a tensor and matrix. For instance, if $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ and $V_1 \in \mathbb{R}^{p_1\times r_1}$, then

$$\mathcal{G} \times_1 V_1 = \Biggl(\sum_{j_1=1}^{r_1} \mathcal{G}(j_1, i_2, i_3) V(i_1, j_1)\Biggr)_{i_1\in[p_1], i_2\in[r_2], i_3\in[r_3]}.$$

We write $(U_1, \ldots, U_m) \cdot \mathcal{G}$ in short for $\mathcal{G} \times_1 U_1 \times_2 \cdots \times_m U_m$. Let $\mathcal{M}_j$ be the jth tensor matricization that rearranges each mode-j fiber of $\mathcal{T} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$ to a column of $\mathcal{M}_j(\mathcal{T}) \in \mathbb{R}^{p_j \times (p_1 \cdots p_d / p_j)}$.

We say $\mathcal{T}$ has Tucker rank $(r_1, \ldots, r_m)$ if it admits a Tucker decomposition

$$\mathcal{T} = (U_1, \ldots, U_m) \cdot \mathcal{G}, \tag{2.1}$$

where $\mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_m}$ and $U_i \in \mathbb{O}_{p_i, r_i}$ for $i \in [m]$. The Tucker decomposition (2.1) can be roughly seen as a generalization of matrix singular value decomposition (SVD) to higher-order tensors, where U_j can be viewed as principal components of the jth matricization of $\mathcal{T}$, and $\mathcal{G}$ contains the singular values. In the case that $r_1 = \cdots = r_m = r$ and $\mathcal{G}$ is diagonalizable, we say $\mathcal{T}$ is *orthogonally decomposable*. If $\mathcal{T}$ satisfies Tucker decomposition (2.1), one has

$$\mathcal{M}_j(\mathcal{T}) = U_j \mathcal{M}_j(\mathcal{G})(U_1 \otimes \cdots \otimes U_{j-1} \otimes U_{j+1} \otimes \cdots \otimes U_m)^\top \in \mathbb{R}^{p_j \times (p_1 \cdots p_m / p_j)}.$$

Here $\otimes$ stands for Kronecker product so that $U \otimes W \in \mathbb{R}^{(p_1 p_2) \times (r_1 r_2)}$ if $U \in \mathbb{R}^{p_1 \times r_1}$ and $W \in \mathbb{R}^{p_2 \times r_2}$. The readers are referred to [42] for a comprehensive survey on tensor algebra.

Let $\sigma_r(\cdot)$ be the rth largest singular value of a matrix. If $\mathcal{T}$ has Tucker ranks $(r_1, \ldots, r_m)$, the signal strength of $\mathcal{T}$ is defined by

$$\lambda_{\min} := \lambda_{\min}(\mathcal{T}) = \min\{\sigma_{r_1}(\mathcal{M}_1(\mathcal{T})), \sigma_{r_2}(\mathcal{M}_2(\mathcal{T})), \ldots, \sigma_{r_m}(\mathcal{M}_m(\mathcal{T}))\},$$

that is, the smallest positive singular value of all matricizations. Similarly, define $\lambda_{\max} := \lambda_{\max}(\mathcal{T}) = \max_j \sigma_1(\mathcal{M}_j(\mathcal{T}))$. The condition number of $\mathcal{T}$ is defined by $\kappa(\mathcal{T}) := \lambda_{\max}(\mathcal{T}) \times \lambda_{\min}^{-1}(\mathcal{T})$. We let Λ_j be the $r_j \times r_j$ diagonal matrix containing the singular values of $\mathcal{M}_j(\mathcal{G})$ (or equivalently the singular values of $\mathcal{M}_j(\mathcal{T})$). Note that Λ_js are not necessarily equal for different j, although $\|\Lambda_1\|_{\mathrm{F}} = \cdots = \|\Lambda_m\|_{\mathrm{F}} = \|\mathcal{T}\|_{\mathrm{F}}$.

We define the principle angles between $U, \widehat{U} \in \mathbb{O}_{p,r}$ as an r-by-r diagonal matrix: $\Theta(U, \widehat{U}) = \operatorname{diag}(\arccos(\sigma_1), \ldots, \arccos(\sigma_r))$, where $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$ are the singular values of $U^\top \widehat{U}$. Then the $\sin\Theta$ distances between $\hat{U}$ and U are defined as

$$\|\sin\Theta(U, \widehat{U})\| = \|\operatorname{diag}(\sin(\arccos(\sigma_1)), \ldots, \sin(\arccos(\sigma_r)))\| = \sqrt{1 - \sigma_r^2},$$

$$\|\sin\Theta(U, \widehat{U})\|_{\mathrm{F}} = \left(\sum_{i=1}^{r} \sin^2(\arccos(\sigma_i))\right)^{1/2} = \left(r - \sum_{i=1}^{r} \sigma_i^2\right)^{1/2}.$$

3. Inference for principal components of Tucker low-rank tensor. For notational simplicity, we focus on the inference for third-order tensors, that is, $m = 3$, while the results for general mth order tensor essentially follows and will be briefly discussed in Section 7.

3.1. *Estimating procedure.* An accurate estimation is often the starting point for statistical inference and uncertainty quantification. In this section, we briefly discuss the estimation procedure for both tensor regression and PCA models. First, we summarize both models as follows:

$$Y_i = \langle \mathcal{X}_i, \mathcal{T} \rangle + \xi_i, \quad i = 1, \ldots, n.$$

Here, $\mathcal{X}_i$ can be the covariate in tensor regression; $n = p_1 p_2 p_3$, $Y_i = \mathcal{A}(j_1, j_2, j_3)$ and $\mathcal{X}_i = (e_{j_1}, e_{j_2}, e_{j_3}) \cdot 1$ with $i = (j_1 - 1)p_2 p_3 + (j_2 - 1)p_3 + j_3$, $j_1 \in [p_1]$, $j_2 \in [p_2]$, $j_3 \in [p_3]$ in tensor PCA. Let $l_n(\mathcal{T}) = \sum_{i=1}^n (Y_i - \langle \mathcal{X}_i, \mathcal{T} \rangle)^2$ be the loss function in both settings. Then a straightforward solution to both problems is via the following Tucker rank constrained least squares estimator:

$$\begin{aligned} &\min_{\operatorname{rank}(\mathcal{T}) \leq (r_1, r_2, r_3)} \ell_n(\mathcal{T}) := \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle \mathcal{X}_i, \mathcal{T} \rangle)^2, \\ &\text{or equivalently} \quad (\hat{\mathcal{G}}, \hat{U}_1, \hat{U}_2, \hat{U}_3) := \operatorname*{arg\,min}_{\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3},\ U_j \in \mathbb{O}_{p_j, r_j}} \ell_n((U_1, U_2, U_3) \cdot \mathcal{G}). \end{aligned} \tag{3.1}$$

Algorithm 1: Power Iteration for Tensor PCA

Input: $\ell_n(\cdot)$: Objective function (3.1); Initializations $(\hat{U}_1^{(0)}, \hat{U}_2^{(0)}, \hat{U}_3^{(0)})$;
for $t = 0, 1$ **do**
  $\hat{U}_1^{(t+1)} =$ leading r_1 left singular vectors of $\mathcal{M}_1(\mathcal{A} \times_2 \hat{U}_2^{(t)\top} \times_3 \hat{U}_3^{(t)\top})$;
  $\hat{U}_2^{(t+1)} =$ leading r_2 left singular vectors of $\mathcal{M}_2(\mathcal{A} \times_1 \hat{U}_1^{(t)\top} \times_3 \hat{U}_3^{(t)\top})$;
  $\hat{U}_3^{(t+1)} =$ leading r_3 left singular vectors of $\mathcal{M}_3(\mathcal{A} \times_1 \hat{U}_1^{(t)\top} \times_2 \hat{U}_2^{(t)\top})$;
end
Output: Test statistic $\hat{U}_1 := \hat{U}_1^{(2)}$, $\hat{U}_2 := \hat{U}_2^{(2)}$, $\hat{U}_3 := \hat{U}_3^{(2)}$, and $\hat{\mathcal{G}} = (\hat{U}_1^{(2)\top}, \hat{U}_2^{(2)\top}, \hat{U}_3^{(2)\top}) \cdot \mathcal{A}$.

Since the objective function (3.1) is highly nonconvex, an efficient algorithm with provable guarantees is crucial for both tensor PCA and regression. As discussed earlier, various computationally feasible procedures have been proposed in the literature. For tensor regression, [32] recently introduced a simultaneous gradient descent algorithm and proved their proposed procedure achieves the minimax optimal estimation error; for tensor PCA, a simpler and more direct approach, higher-order orthogonal iteration (HOOI), was introduced by [23]. The implementation details of both algorithms are provided in Section A in the Supplementary Material.

Moreover, the primary interest of this paper is on the statistical inference for $\mathcal{T}$ or U_j, far beyond deriving estimators achieving optimal estimation error. In general, even estimators achieving minimax optimal estimation error rate may not enjoy a proper asymptotic distribution. For example, the true parameter $\mathcal{T}$ or U_j plus a small enough perturbation can achieve optimal estimation error but does not satisfy any tractable distribution.

To this end, we introduce a *two-iteration alternating minimization* algorithm for both the Tucker low-rank tensor PCA and tensor regression in Algorithms 1 and 2, respectively. Our theory in later this section reveals a surprising fact: if any estimator $\tilde{\mathcal{T}} = (\hat{U}_1^{(0)}, \hat{U}_2^{(0)}, \hat{U}_3^{(0)}) \cdot \hat{\mathcal{G}}^{(0)}$ achieving some attainable estimation error is provided as the input, the two-iteration

Algorithm 2: Alternating Minimization for Tensor Regression

Input: $\ell_n(\cdot)$: Objective function (3.1); Initializations $(\hat{U}_1^{(0)}, \hat{U}_2^{(0)}, \hat{U}_3^{(0)})$, and $\hat{\mathcal{G}}^{(0)}$ is the solution of $\operatorname{argmin}_{\mathcal{G}} \ell_n((\hat{U}_1^{(0)}, \hat{U}_2^{(0)}, \hat{U}_3^{(0)}) \cdot \mathcal{G})$ for tensor regression model;
for $t = 0, 1$ **do**
  Solve $\nabla_{U_1} \ell_n((\hat{U}_1^{(t+0.5)}, \hat{U}_2^{(t)}, \hat{U}_3^{(t)}) \cdot \hat{\mathcal{G}}^{(t)}) = 0$ to obtain $\hat{U}_1^{(t+0.5)}$;
  Update by $\hat{U}_1^{(t+1)} = \mathrm{SVD}_{r_1}(\hat{U}_1^{(t+0.5)})$;
  Solve $\nabla_{U_2} \ell_n((\hat{U}_1^{(t)}, \hat{U}_2^{(t+0.5)}, \hat{U}_3^{(t)}) \cdot \hat{\mathcal{G}}^{(t)}) = 0$ to obtain $\hat{U}_2^{(t+0.5)}$;
  Update by $\hat{U}_2^{(t+1)} = \mathrm{SVD}_{r_2}(\hat{U}_2^{(t+0.5)})$;
  Solve $\nabla_{U_3} \ell_n((\hat{U}_1^{(t)}, \hat{U}_2^{(t)}, \hat{U}_3^{(t+0.5)}) \cdot \hat{\mathcal{G}}^{(t)}) = 0$ to obtain $\hat{U}_3^{(t+0.5)}$;
  Update by $\hat{U}_3^{(t+1)} = \mathrm{SVD}_{r_3}(\hat{U}_3^{(t+0.5)})$;
  Solve $\nabla_{\mathcal{G}} \ell_n((\hat{U}_1^{(t+1)}, \hat{U}_2^{(t+1)}, \hat{U}_3^{(t+1)}) \cdot \hat{\mathcal{G}}^{(t+1)}) = 0$ to obtain $\hat{\mathcal{G}}^{(t+1)}$;
end
Output: Test statistic $\hat{U}_1 := \hat{U}_1^{(2)}$, $\hat{U}_2 := \hat{U}_2^{(2)}$, $\hat{U}_3 := \hat{U}_3^{(2)}$, and $\hat{\mathcal{G}} := \hat{\mathcal{G}}^{(2)}$.

alternating minimization in Algorithms 1 and 2 will provide an estimator enjoying asymptotic normality and being ready to use for confidence region construction.

REMARK 1 (Interpretation of alternating minimization update in tensor PCA). A key observation by [23], Theorems 4.1, 4.2, shows minimizing $\min_{\operatorname{rank}(\mathcal{T})\le(r_1,r_2,r_3)} \|\mathcal{T}-\mathcal{A}\|_{\mathrm{F}}^2$ is equivalent to maximizing $\max_{U_j\in\mathbb{O}_{p_j,r_j}} \|(U_1^\top, U_2^\top, U_3^\top)\cdot\mathcal{A}\|_{\mathrm{F}}^2$. Therefore, the optimization in tensor PCA is equivalent to

$$(\hat{U}_1,\hat{U}_2,\hat{U}_3) := \mathop{\arg\min}_{U_j\in\mathbb{O}_{p_j,r_j}} \ell_n\big((U_1,U_2,U_3)\cdot\mathcal{G}\big) := \mathop{\arg\max}_{U_j\in\mathbb{O}_{p_j,r_j}} \big\|\big(U_1^\top,U_2^\top,U_3^\top\big)\cdot\mathcal{A}\big\|_{\mathrm{F}}^2$$
$$= \mathop{\arg\max}_{U_j\in\mathbb{O}_{p_j,r_j}} \|U_j\mathcal{M}_j(\mathcal{A}\times_{j+1}U_{j+1}\times_{j+2}U_{j+2})\|_{\mathrm{F}}^2.$$

Here, for convenience of notation, $U_4=U_1$, $U_5=U_2$, $r_4=r_1$, $r_5=r_2$. Note that, given fixed $\hat{U}_{j+1}^{(t)}$ and $\hat{U}_{j+2}^{(t)}$, the Eckart–Young–Mirsky theorem [27] implies the optimal solution to $\max_{U_j\in\mathbb{O}_{p_j,r_j}} \|(U_j^\top,\hat{U}_{j+1}^{(t)\top},\hat{U}_{j+2}^{(t)\top})\cdot\mathcal{A}\|_{\mathrm{F}}^2$ is attainable via singular value decomposition:

$$\hat{U}_j^{(t+1)} = \text{leading } r_j \text{ left singular vectors of } \mathcal{M}_j\big(\mathcal{A}\times_{j+1}\hat{U}_{j+1}^{(t)\top}\times_{j+2}\hat{U}_{j+2}^{(t)\top}\big).$$

This explains the alternating minimization update steps for tensor PCA in Algorithm 1.

Hereinafter, we denote $\hat{U}_j$ the output of Algorithms 1 and 2, $p=\max\{p_1,p_2,p_3\}$ and $r_{\max}=\max\{r_1,r_2,r_3\}$. Next, we establish the asymptotic distribution and develop the inference procedure for $\|\sin\Theta(\hat{U}_j,U_j)\|_{\mathrm{F}}^2$ in tensor PCA and tensor regression models when $\mathcal{T}$ admits the Tucker decomposition (2.1).

3.2. *Inference for the Tucker low-rank tensor PCA.* We assume the following condition on initialization $(\hat{U}_1^{(0)},\hat{U}_2^{(0)},\hat{U}_3^{(0)})$ of Algorithm 1 holds.

ASSUMPTION 1. Under tensor PCA model (1.1) with $\mathcal{Z}_{i_1,i_2,i_3}\overset{\text{i.i.d.}}{\sim}N(0,\sigma^2)$, there is an event $\mathcal{E}_0$ with $\mathbb{P}(\mathcal{E}_0)\ge 1-C_1e^{-c_1p}$ for some absolute constants $c_1, C_1>0$ so that, under $\mathcal{E}_0$, the initialization $(\hat{U}_1^{(0)},\hat{U}_2^{(0)},\hat{U}_3^{(0)})$ satisfy $\max_{j=1,2,3}\|\sin\Theta(\hat{U}_j^{(0)},U_j)\|\le C_2\sqrt{p}\sigma/\lambda_{\min}$ for some absolute constant $C_2>0$.

The claimed error rates in Assumption 1 are attainable by the algorithm HOOI under the SNR condition $\lambda_{\min}/\sigma\ge Cp^{3/4}$ [77], Theorem 1. Such the SNR condition is essential to ensure a consistent estimator is achievable in polynomial time as illustrated by the literature reviewed in Section 1.2. Note that [77], Theorem 1, presented an expectation error bound $\mathbb{E}\|\sin\Theta(\hat{U}_j^{(0)},U_j)\|$, while its proof indeed involved a desired probabilistic bound as claimed by Assumption 1. If a given initialization estimation error upper bound is in a metric other than the $\sin\Theta$ distance described in Assumption 1, we may apply Lemma 4 in the Supplementary Material to "translate" the upper bound in another metric to the desired $\sin\Theta$ distance.

Suppose $\hat{U}_j$ is the output of Algorithm 1. Built on Assumption 1, we characterize the distribution of $\|\sin\Theta(\hat{U}_j,U_j)\|_{\mathrm{F}}^2$ by the following theorem.

THEOREM 1 (Asymptotic normality of principal components in tensor PCA). *Suppose Assumption 1 holds for tensor PCA model* (1.1), $\mathcal{Z}(i_1,i_2,i_3)\overset{\text{i.i.d.}}{\sim}N(0,\sigma^2)$, $p_j\asymp p$ *for* $j=$

1, 2, 3, *and* $\kappa(\mathcal{T}) \le \kappa_0$. *Let* $\hat{U}_j$*s be the output of Algorithm* 1 *for tensor PCA model. There exist absolute constants* $c_1, C_0, C_1, C_2, C_3 > 0$ *such that if* $\lambda_{\min}/\sigma \ge C_0(p^{3/4} + \kappa_0^2 p^{1/2})$, *then*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \le x\right) - \Phi(x)\right|$$
$$\le C_1 e^{-c_1 p} + C_2\left(\frac{\kappa_0^6(pr_{\max})^{3/2}}{(\lambda_{\min}/\sigma)^2} + \frac{\kappa_0^2(p\log p)^{1/2}}{\lambda_{\min}/\sigma}\right) + C_3\frac{r_{\max}^{3/2}}{\sqrt{p}},$$

where $\Lambda_j = \mathrm{diag}(\lambda_1^{(j)}, \ldots, \lambda_{r_j}^{(j)})$ *is the diagonal matrix containing the singular values of* $\mathcal{M}_j(\mathcal{G})$, *and* $\Phi(x)$ *is the cumulative distribution function of* $N(0,1)$.

If the condition number $\kappa_0 = O(1)$, $(pr_{\max})^{3/4}(\lambda_{\min}/\sigma)^{-1} \to 0$ and $r_{\max}^3/p \to 0$ as $p \to \infty$, Theorem 1 yields

$$\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p \to \infty.$$

By the proof of Theorem 1, we can further establish the following joint distribution of all U_js:

$$\left(\frac{\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 - p_1\sigma^2\|\Lambda_1^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_1}\sigma^2\|\Lambda_1^{-2}\|_{\mathrm{F}}}, \frac{\|\sin\Theta(\hat{U}_2, U_2)\|_{\mathrm{F}}^2 - p_2\sigma^2\|\Lambda_2^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_2}\sigma^2\|\Lambda_2^{-2}\|_{\mathrm{F}}},\right.$$
$$\left.\frac{\|\sin\Theta(\hat{U}_3, U_3)\|_{\mathrm{F}}^2 - p_3\sigma^2\|\Lambda_3^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_3}\sigma^2\|\Lambda_3^{-2}\|_{\mathrm{F}}}\right) \xrightarrow{\text{d.}} N(0, I_3) \quad \text{as } p \to \infty.$$

REMARK 2. We briefly compare Theorem 1 with the existing results in the literature. The asymptotic normality of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ in Theorem 1 requires SNR condition $\lambda_{\min}/\sigma \gg (r_{\max}p)^{3/4}$, which is slightly stronger than the optimal SNR condition $\lambda_{\min}/\sigma \ge C_0 p^{3/4}$ for achieving the consistent estimation in [77], Theorem 1 (if $r \ge 1$), matches the condition in [83], Theorem 1, (if $r = 1$), and weaker than the condition in [55], Theorem 4, (if $r = 1$). Second, note that Theorem 1 implies $\mathbb{E}\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 = (1 + o(1))p_j\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2$. To the best of our knowledge, this is the first result with a precise constant characterization of the estimation error in tensor PCA.

Compared with the conditions for consistent estimation in [77], our Theorem 1 is for valid statistical inference, which requires the additional $\kappa_0^2 p^{1/2}$ term in SNR and a rank condition $r_{\max}^3/p \to 0$. These terms emerge from technical issues, in particular from the way we bound the higher-order terms in the empirical spectral projector $\hat{U}_j\hat{U}_j^\top$ to better cope with the dependence across iterations. We note [20] proves that a rank-one planted tensor is distinguishable from the pure noise tensor if SNR $\lambda_{\min}/\sigma \ge C_0 p^{1/2}$ holds for a certain positive constant threshold C_0. In comparison, our Theorem 1 requires a stronger condition $\lambda_{\min}/\sigma \gg p^{3/4}$. In fact, the gap between $p^{1/2}$ and $p^{3/4}$ is fundamental. Without considering the computational feasibility, the SNR threshold $p^{1/2}$ is sufficient not only for detection but also for estimation (see [77], Theorem 2). However, when SNR falls below the threshold $p^{3/4}$, various pieces of evidence were established in the literature, as described in the first paragraph of Section 1.2, that show no polynomial time algorithms can reliably estimate the principal components.

While Theorem 1 characterizes the asymptotic distribution of $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ for tensor PCA model, the result is not immediately applicable to uncertainty quantification of $\hat{U}_j$ since

$\|\Lambda_j^{-1}\|_{\mathrm{F}}^2$, $\|\Lambda_j^{-2}\|_{\mathrm{F}}$ and σ^2 are often unknown in practice. We thus propose an estimate for Λ_j, σ:

$$\begin{aligned}\hat{\Lambda}_j = {} & \text{ diagonal matrix with the top } r_j \text{ singular values of} \\ & \mathcal{M}_j(\mathcal{A} \times_{j+1} \hat{U}_{j+1}^\top \times_{j+2} \hat{U}_{j+2}^\top), \\ \hat{\sigma} = {} & \|\mathcal{A} - \mathcal{A} \times_1 \hat{U}_1\hat{U}_1^\top \times_2 \hat{U}_2\hat{U}_2^\top \times_3 \hat{U}_3\hat{U}_3^\top\|_{\mathrm{F}}/\sqrt{p_1p_2p_3}.\end{aligned} \tag{3.2}$$

We can prove a deviation bound for $\hat{\sigma}$ and the normal approximation for $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2$ with the proposed plug-in estimators.

LEMMA 1. *Under conditions of Theorem 1, there exist two constants $C_1, C_2 > 0$ such that*

$$\mathbb{P}\{|\hat{\sigma}^2/\sigma^2 - 1| \le C_2(\kappa_0\sqrt{r_{\max}}p^{-1} + p^{-3/4}\sqrt{\log(p)})\} \ge 1 - C_1p^{-3}.$$

THEOREM 2 (Inference for Tucker low-rank tensor PCA). *Suppose the conditions in Theorem 1 hold. Let $\hat{\Lambda}_1 \in \mathbb{R}^{r_1\times r_1}$ and $\hat{\sigma}$ be defined as (3.2). There exist absolute constants $c_1, C_0, C_1, C_2, C_3 > 0$ such that if $\lambda_{\min}/\sigma \ge C_0(p^{3/4} + \kappa_0^2 p^{1/2})$, then for $j = 1, 2, 3$,*

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j\hat{\sigma}^2\|\hat{\Lambda}_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}\hat{\sigma}^2\|\hat{\Lambda}_j^{-2}\|_{\mathrm{F}}} \le x\right) - \Phi(x)\right| \\ &\quad\le C_1e^{-c_1p} + C_2\left(\frac{r_{\max}^{3/2}\kappa_0^6p^{3/2}}{(\lambda_{\min}/\sigma)^2} + \frac{\kappa_0^3\sqrt{pr_{\max}(r_{\max}^2 + \log p)}}{\lambda_{\min}/\sigma} + \frac{\sqrt{\log(p)}}{p^{1/4}} + \frac{\kappa_0\sqrt{r_{\max}}}{\sqrt{p}}\right) \\ &\qquad + C_3\frac{r_{\max}^{3/2}}{\sqrt{p}}.\end{aligned}$$

When the condition number $\kappa_0 = O(1)$, $(pr_{\max})^{3/4}(\lambda_{\min}/\sigma)^{-1} \to 0$, and $r_{\max}^3/p \to 0$ as $p \to \infty$, Theorem 2 implies

$$\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j\hat{\sigma}^2\|\hat{\Lambda}_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}\hat{\sigma}^2\|\hat{\Lambda}_j^{-2}\|_{\mathrm{F}}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p \to \infty. \tag{3.3}$$

Equation (3.3) is readily applicable to statistical inference for U_j. After getting $\hat{U}_j$ by Algorithm 1, we propose a $(1-\alpha)$-level confidence region for U_j as

$$\mathrm{CR}_\alpha(\hat{U}_j) := \{V \in \mathbb{O}_{p_j,r_j} : \|\sin\Theta(\hat{U}_j, V)\|_{\mathrm{F}}^2 \le p_j\hat{\sigma}^2\|\hat{\Lambda}_j^{-1}\|_{\mathrm{F}}^2 + z_\alpha\sqrt{2p_j}\hat{\sigma}^2\|\hat{\Lambda}_j^{-2}\|_{\mathrm{F}}\}, \tag{3.4}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$ is the $(1-\alpha)$ quantile of the standard normal distribution. The following corollary is an immediate result of Theorem 2, which confirms that the confidence region $\mathrm{CR}_\alpha(\hat{U}_1)$ is indeed asymptotically accurate.

COROLLARY 1 (Confidence region for tensor PCA). *Suppose the conditions of Theorem 2 hold and the confidence region $\mathrm{CR}_\alpha(\hat{U}_j)$ is defined in (3.4). If $\kappa_0^6(r_{\max}^{3/2}p^{3/2} + r_{\max}p\log p)(\lambda_{\min}/\sigma)^{-2} \to 0$ and $r_{\max}^3/p \to 0$ as $p \to \infty$, then*

$$\lim_{p\to\infty}\mathbb{P}(U_j \in \mathrm{CR}_\alpha(\hat{U}_j)) = 1 - \alpha.$$

We note that, through a more sophisticated analysis, Theorem 1 can be generalized to the setting with sub-Gaussian noise. For simplicity, we only prove the following rank-one case, which has been the focus of many papers on tensor PCA.

THEOREM 3 (Rank-one tensor PCA under sub-Gaussian noise). *Suppose Assumption 1 holds for tensor PCA model (1.1) with* $r = 1$, $p_j \asymp p$ *for* $j = 1, 2, 3$, $\mathcal{Z}$ *has i.i.d. entries with* $\mathbb{E}\mathcal{Z}(i_1, i_2, i_3) = 0$, $\mathbb{E}[\mathcal{Z}(i_1, i_2, i_3)^2] = \sigma^2$, $\mathbb{E}[\mathcal{Z}(i_1, i_2, i_3)^4]/\sigma^4 = \nu$ *and* $\|\mathcal{Z}(i_1, i_2, i_3)\|_{\psi_2} \le C\sigma$ *for some constant* $C > 0$. *There exist absolute constants* $c_1, C_0, C_1, C_2, C_3 > 0$ *such that if* $\lambda/\sigma \ge C_0 p^{3/4}$, *then*

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 - p_1\sigma^2\lambda^{-2}}{\sigma^2\lambda^{-2}\sqrt{p_1(2+(\nu-3)\|U_2\|_4^4\|U_3\|_4^4)}} \le x\right) - \Phi(x)\right| \\
&\quad\le \frac{C_2}{p^{1/2}(2+(\nu-3)\|U_2\|_4^4\|U_3\|_4^4)^{3/2}} \\
&\qquad + C_3\left(\frac{\sqrt{p\log p}}{\lambda/\sigma} + \frac{p^{3/2}}{(\lambda/\sigma)^2} + \frac{\log p}{\sqrt{p}}\right)\cdot\frac{1}{\sqrt{2+(\nu-3)\|U_2\|_4^4\|U_3\|_4^4}} + C_1 e^{-c_1 p}.
\end{aligned}$$

Similar results can be derived for $\|\sin\Theta(\hat{U}_2, U_2)\|_{\mathrm{F}}^2$ *and* $\|\sin\Theta(\hat{U}_3, U_3)\|_{\mathrm{F}}^2$.

If $\nu - 1 \ge c_0$ for some absolute constant $c_0 > 0$ and $\lambda^{-1}\sigma p^{3/4} \to 0$ as $p \to \infty$, Theorem 3 implies

$$\frac{\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 - p_1\sigma^2\lambda^{-2}}{\sigma^2\lambda^{-2}\sqrt{p_1(2+(\nu-3)\|U_2\|_4^4\|U_3\|_4^4)}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p \to \infty.$$

The SNR condition in Theorem 3 is the same as that in Theorem 1. Moreover, the asymptotic variance of $\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2$ includes the kurtosis of the noise distribution ν, which can be challenging to estimate. We leave the estimation of the kurtosis and the data-driven inference for $\hat{U}_k$ as future research.

3.3. *Inference for Tucker low-rank tensor regression.* This section is devoted to the asymptotic distribution and inference in low-rank tensor regression. We first introduce the following assumption on the initialization for Algorithm 2.

ASSUMPTION 2. Under tensor regression model (1.2) with $\mathcal{X}(i_1, i_2, i_3) \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\mathrm{Var}(\xi_i) = \sigma^2$ and $\|\xi_i\|_{\psi_2} \le C\sigma$ for some constant $C > 0$, there is an event $\mathcal{E}_0$ with $\mathbb{P}(\mathcal{E}_0) \ge 1 - C_1 e^{-c_1 p}$ for some absolute constants $c_1, C_1 > 0$ so that, under $\mathcal{E}_0$, the initialization $\tilde{\mathcal{T}} = (\hat{U}_1^{(0)}, \hat{U}_2^{(0)}, \hat{U}_3^{(0)})\cdot\hat{\mathcal{G}}^{(0)}$ satisfy $\|\tilde{\mathcal{T}} - \mathcal{T}\|_{\mathrm{F}}^2 \le C_2 p r_{\mathsf{max}}\sigma^2/n$ or $\max_j\|\sin\Theta(\hat{U}_j^{(0)}, U_j)\| \le C_2\sqrt{p/n}\sigma/\lambda_{\mathsf{min}}$ for some absolute constant $C_2 > 0$.

The claimed bound of $\|\tilde{\mathcal{T}} - \mathcal{T}\|_{\mathrm{F}}^2$ in Assumption 2 is attainable, for instance, by the gradient descent algorithm developed in [32] and the importance sketching algorithm developed in [78] under the SNR condition $n(\lambda_{\mathsf{min}}/\sigma)^2 \ge Cp^{3/2}$ and the sample size condition $n \ge Cp^{3/2}r_{\mathsf{max}}$. The theoretical guarantees for this claim can be found in [32], Theorem 4.2, and [78], Theorem 4.

Based on Assumption 2, we establish the following asymptotic results for tensor regression.

THEOREM 4. *Suppose Assumption 2 holds for tensor regression model (1.2),* $\mathcal{X}(i_1, i_2, i_3) \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\mathrm{Var}(\xi_i) = \sigma^2$ *and* $\|\xi_i\|_{\psi_2} \le C\sigma$ *for some constant* $C > 0$, $p_j \asymp p$ *for* $j = 1, 2, 3$, *and* $\kappa(\mathcal{T}) \le \kappa_0$. *Let* $\hat{U}_j$*s be the output of two-iteration alternating minimization (Algorithm 2). There exist absolute constants* $c_1, C_0, C_1, C_2, C_3, C_4 > 0$ *such that if* $n(\lambda_{\min}/\sigma)^2 \ge C_0(p^{3/2} \vee \kappa_0^4 p r_{\max}^2)$ *and* $n \ge C_2(p^{3/2} \vee \kappa_0^2 p r_{\max}^3)$, *then*

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j n^{-1}\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}n^{-1}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \le x\right) - \Phi(x)\right| \\
&\quad\le \frac{C_3\kappa_0^4 r_{\max}^{5/2}p^{3/2}}{n} + C_3\kappa_0^3\left(\frac{r_{\max}^5 p\log^2 n}{n}\right)^{1/2} + \frac{C_3 p^{3/2}}{n}\left(\frac{\kappa_0^5 r_{\max}^2}{\lambda_{\min}/\sigma} + \frac{\kappa_0^5 r_{\max}^{3/2}}{(\lambda_{\min}/\sigma)^2}\right) \\
&\qquad + C_3\kappa_0^4\left(\frac{p r_{\max}^3 + r_{\max} p\log p}{n(\lambda_{\min}/\sigma)^2}\right)^{1/2} + C_1 e^{-c_1 p} + C_4\frac{r_{\max}^{3/2}}{\sqrt{p}},
\end{aligned}$$

where Λ_j *is the* $r_j \times r_j$ *diagonal matrix containing the singular values of* $\mathcal{M}_j(\mathcal{T})$.

If the condition number $\kappa_0 = O(1)$, $r_{\max}^3/p \to 0$, $(r_{\max}^{5/2}p^{3/2} + r_{\max}^5 p\log^2 n)/n \to 0$ and $r_{\max}^{3/2}p^{3/2}/(n(\lambda_{\min}/\sigma)^2) \to 0$ as $p \to \infty$, Theorem 4 implies

$$\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j n^{-1}\sigma^2\|\Lambda_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}n^{-1}\sigma^2\|\Lambda_j^{-2}\|_{\mathrm{F}}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p \to \infty.$$

To make inference for tensor regression, we develop the following asymptotic normal distribution for $\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}$ with the plug-in estimates of Λ_j.

THEOREM 5 (Tensor regression). *Suppose the conditions in Theorem 4 hold. Let* $\hat{\Lambda}_j = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{r_j})$ *be a diagonal matrix containing the singular values of* $\mathcal{M}_1(\hat{\mathcal{G}})$, *where* $\hat{\mathcal{G}}$ *is the output of Algorithm 2. There exist absolute constants* $c_1, C_0, C_1, C_2, C_3, C_4 > 0$ *such that if* $n(\lambda_{\min}/\sigma)^2 \ge C_0(p^{3/2} \vee \kappa_0^6 p r_{\max}^2)$ *and* $n \ge C_2(p^{3/2} \vee \kappa_0^8 p r_{\max}^3)$, *then for* $j = 1, 2, 3$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\|\sin\Theta(\hat{U}_j, U_j)\|_{\mathrm{F}}^2 - p_j n^{-1}\sigma^2\|\hat{\Lambda}_j^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_j}n^{-1}\sigma^2\|\hat{\Lambda}_j^{-2}\|_{\mathrm{F}}} \le x\right) - \Phi(x)\right| \\
&\quad\le C_3\frac{\kappa_0^4 r_{\max}^{5/2}p^{3/2}}{n} + C_3\kappa_0^3\left(\frac{r_{\max}^5 p\log^2 n}{n}\right)^{1/2} + \frac{C_3 p^{3/2}}{n}\left(\frac{\kappa_0^5 r_{\max}^2}{\lambda_{\min}/\sigma} + \frac{\kappa_0^5 r_{\max}^{3/2}}{(\lambda_{\min}/\sigma)^2}\right) \\
&\qquad + C_3\kappa_0^4\left(\frac{p r_{\max}^3 + r_{\max} p\log p}{n(\lambda_{\min}/\sigma)^2}\right)^{1/2} + C_1 e^{-c_1 p} + C_4\frac{r_{\max}^{3/2}}{\sqrt{p}}.
\end{aligned}$$

We propose the following $(1-\alpha)$-level confidence region for U_j:

$$\widetilde{\mathrm{CR}}_\alpha(\hat{U}_j) := \left\{V \in \mathbb{O}_{p_j, r_j} : \|\sin\Theta(\hat{U}_j, V)\|_{\mathrm{F}}^2 \le \frac{p_j\sigma^2\|\hat{\Lambda}_j^{-1}\|_{\mathrm{F}}^2}{n} + z_\alpha\frac{\sqrt{2p_j}\sigma^2\|\hat{\Lambda}_j^{-2}\|_{\mathrm{F}}}{n}\right\}. \tag{3.5}$$

The following corollary establishes the coverage probability of the proposed confidence region.

COROLLARY 2 (Confidence region for tensor regression). *Suppose the conditions of Theorem 5 hold and the confidence region* $\widetilde{\mathrm{CR}}_\alpha(\hat{U}_j)$ *is defined by (3.5). If* $(\kappa_0^5 r_{\max}^{5/2}p^{3/2} +$

$\kappa_0^6 r_{\max}^5 p \log^2 n)/n \to 0$, $\kappa_0^5 r_{\max}^{3/2} p^{3/2}/(n(\lambda_{\min}/\sigma)^2) \to 0$ *and* $r_{\max}^3/p \to 0$ *as* $p \to \infty$, *then*

$$\lim_{p\to\infty} \mathbb{P}\big(U_j \in \widetilde{\mathrm{CR}}_\alpha(\hat{U}_j)\big) = 1-\alpha.$$

REMARK 3 (Selection of σ). When σ is unknown, we can estimate it by a sample splitting scheme as follows. First, we retain a part of sample $\{(\mathcal{X}_k, Y_k)\}_{k=1}^{\lceil p^{3/2}\rceil}$ and use the other samples to compute the estimator $\tilde{\mathcal{T}}$. Define

$$\hat{\sigma}^2 := \sum\nolimits_{k=1}^{\lceil p^{3/2}\rceil} \big(Y_k - \langle \tilde{\mathcal{T}}, \mathcal{X}_k\rangle\big)^2/\lceil p^{3/2}\rceil.$$

Under Assumption 2 and conditions of Theorem 4, we can show with probability at least $1-p^{-3}$, $|\hat{\sigma}^2/\sigma^2 - 1| = O(p^{-3/4}\sqrt{\log p} + r_{\max} p n^{-1})$. By plugging in $\hat{\sigma}$ to (3.5), we obtain a data-driven $(1-\alpha)$ asymptotic confidence region for U_j.

3.4. *Proof sketch.* In this section, we briefly explain the proof strategy for tensor PCA model, that is, Theorem 1. The proof for tensor regression model is more complicated but shares similar spirits. Without loss of generality, we assume $\sigma = 1$. First,

$$\begin{aligned} 2\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 &= \|\hat{U}_1\hat{U}_1^\top - U_1U_1^\top\|_{\mathrm{F}}^2 = 2r_1 - 2\langle \hat{U}_1\hat{U}_1^\top, U_1U_1^\top\rangle \\ &= -2\langle U_1U_1^\top, \hat{U}_1\hat{U}_1^\top - U_1U_1^\top\rangle. \end{aligned}$$

It thus suffices to investigate the distribution of $\langle U_1U_1^\top, \hat{U}_1\hat{U}_1^\top - U_1U_1^\top\rangle$. By Algorithm 1, $\hat{U}_1$ are the top-r_1 left singular vectors of $\mathcal{M}_1(\mathcal{A}\times_2 \hat{U}_2^{(1)\top} \times_3 \hat{U}_3^{(1)\top})$. As a result, $\hat{U}_1\hat{U}_1^\top$ is the spectral projector and can be decomposed as

$$\mathcal{M}_1(\mathcal{A})(\hat{U}_2^{(1)}\hat{U}_2^{(1)\top} \otimes \hat{U}_3^{(1)}\hat{U}_3^{(1)\top})\mathcal{M}_1^\top(\mathcal{A}) =: \mathcal{M}_1(\mathcal{T})\mathcal{M}_1^\top(\mathcal{T}) + D_1^{(1)}.$$

The high-level ideas of the proof include the following steps.

Step 1: We apply the spectral representation formula ([70]; also see the statement in Lemma 2 from the Supplementary Material) and expand

$$\hat{U}_1\hat{U}_1^\top = U_1U_1^\top + S_1(D_1^{(1)}) + S_2(D_1^{(1)}) + S_3(D_1^{(1)}) + \sum_{k\ge 4} S_k(D_1^{(1)}),$$

where $S_k(\cdot)$ denotes the kth order perturbation term:

$$S_k(D_1^{(1)}) = \sum_{s_1+\cdots+s_{k+1}=k} (-1)^{1+\tau(\mathbf{s})}\cdot B_1^{-s_1}D_1^{(1)}B_1^{-s_2}D_1^{(1)}B_1^{-s_3}\cdots B_1^{-s_k}D_1^{(1)}B_1^{-s_{k+1}},$$

where $B_1^{-k} = U_1\Lambda_1^{-2k}U_1^\top$ for each positive integer k, $B_1^0 := I_{p_1} - U_1U_1^\top$, $s_1,\ldots,s_{k+1}$ are nonnegative integers, and $\tau(\mathbf{s}) = \sum_{j=1}^{k+1}\mathbb{I}(s_j > 0)$.

Step 2: Since $\langle U_1U_1^\top, S_1(D_1^{(1)})\rangle = 0$ and $\|S_k(D_1^{(1)})\| \le (C_1\kappa_0^2\sqrt{p}/\lambda_{\min})^k$ with high probability, we can write

$$\langle \hat{U}_1\hat{U}_1^\top - U_1U_1^\top, U_1U_1^\top\rangle = \langle S_2(D_1^{(1)}), U_1U_1^\top\rangle + \langle S_3(D_1^{(1)}), U_1U_1^\top\rangle + O\Big(\frac{r_{\max}\kappa_0^8 p^2}{\lambda_{\min}^4}\Big).$$

In other words, the higher-order terms ($k \ge 4$) can be bounded with high probability, which becomes small-order terms.

Step 3: We show, with high probability, the third-order term can be bounded by

$$|\langle S_3(D_1^{(1)}), U_1U_1^\top\rangle| = O\Big(\frac{\kappa_0^3 p\sqrt{r_{\max}\log p}}{\lambda_{\min}^3} + \frac{\kappa_0^3 p^2 r_{\max}^{3/2}}{\lambda_{\min}^4}\Big)$$

and becomes a small-order term. Now, it suffices to only investigate the second-order term carefully.

Step 4: We decompose the second-order term $\langle S_2(D_1^{(1)}), U_1U_1^\top\langle$ into a leading term and remainder terms. Similar to *Step 2* and *Step 3*, we show that the remainder terms are, with high probability, bounded by $O(\kappa_0^3 p\sqrt{r_{\max}\log p}\lambda_{\min}^{-3} + \kappa_0^3 p^2 r_{\max}^{3/2}\lambda_{\min}^{-4})$.

Step 5: We prove that the leading term of $\langle S_2(D_1^{(1)}), U_1U_1^\top\rangle$ can be written as a sum of independent random variables, which yields a normal approximation by the Berry–Essen theorem. Finally, combining all these steps, we get the normal approximation for $\|\hat{U}_1\hat{U}_1^\top - U_1U_1^\top\|_{\mathrm{F}}^2$.

Among these steps, Steps 4 and 5 are the most technically involved. Throughout the proof, we apply the spectral representation formula at multiple stages to prove sharp upper bounds for higher-order terms, and establish central limit theorem for the second-order term.

The following lemmas are used in our proof and could be of independent interest. First, Lemma 2 is used to establish the concentration inequalities for the sum of random variables that have heavier tails than Gaussian.

LEMMA 2 (Orlicz ψ_α-norm for product of random variables). *Suppose $X_1, \ldots, X_n$ are n random variables (not necessarily independent) satisfying $\|X_i\|_{\psi_{\alpha_i}} \le K_i$. Define $\bar{\alpha} = (\sum_{i=1}^n \alpha_i^{-1})^{-1}$. Then*

$$\left\|\prod_{i=1}^n X_i\right\|_{\psi_{\bar{\alpha}}} \le \prod_{i=1}^n K_i.$$

Next, Lemma 3 provides a tight probabilistic upper bound for sum of third moments of Gaussian random matrices.

LEMMA 3. *Suppose $Z_1, \ldots, Z_n \in \mathbb{R}^{p\times r}$ are independent random matrices satisfying $Z_i(j,k) \overset{\text{i.i.d.}}{\sim} N(0,1)$. Then there exist two universal constants $C, C_1 > 0$ such that for fixed $M_1, \ldots, M_n \in \mathbb{R}^{p\times r}$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \|Z_i\|_{\mathrm{F}}^2\langle Z_i, M_i\rangle\right| \ge Cpr\left(\sum_{i=1}^n \|M_i\|_{\mathrm{F}}^2\right)^{1/2}\sqrt{\log(p)}\right) \le p^{-C_1}.$$

4. PCA for orthogonally decomposable tensors. In this section, we specifically focus on the tensor PCA model (1.1) with orthogonally decomposable signal tensor $\mathcal{T}$:

$$\mathcal{T} = \sum_{i=1}^r \lambda_i \cdot u_i \otimes v_i \otimes w_i, \tag{4.1}$$

where $U = (u_1, \ldots, u_r) \in \mathbb{O}_{p_1,r}$, $V = (v_1, \ldots, v_r) \in \mathbb{O}_{p_2,r}$, and $W = (w_1, \ldots, w_r) \in \mathbb{O}_{p_3,r}$ all have orthonormal columns; the singular values satisfy $\lambda_{\min} = \min\{\lambda_1, \ldots, \lambda_r\} > 0$. Here, for any $u \in \mathbb{R}^{p_1}$, $v \in \mathbb{R}^{p_2}$, $w \in \mathbb{R}^{p_3}$, $u \otimes v \otimes w$ is a $p_1 \times p_2 \times p_3$ tensor whose (i,j,k)th entry is $u(i)v(j)w(k)$.

Our goal is to make inference on the principal components based on a noisy observation $\mathcal{A} = \mathcal{T} + \mathcal{Z}$. Different from the inference for Tucker low-rank tensor discussed in Section 3, where an accurate estimation is hopeful only for the joint column space of U_j due to the nonidentifiability of Tucker decomposition, we can make inference for each individual vector $\{u_j, v_j, w_j\}$ if $\mathcal{T}$ is orthogonally decomposable as (4.1). Given some estimates $\{\hat{u}_j^{(0)}, \hat{v}_j^{(0)}, \hat{w}_j^{(0)}\}_{j=1}^r$, we propose to pass them to a post-processing step by two-iteration procedure in Algorithm 3 to obtain the test statistics $\{\hat{u}_j, \hat{v}_j, \hat{w}_j\}_{j=1}^r$.

Algorithm 3: Power Iterations for Orthogonally decomposable $\mathcal{T}$

Input: $\mathcal{A}$, initialization $\{\hat{u}_j^{(0)}, \hat{v}_j^{(0)}, \hat{w}_j^{(0)}\}_{j=1}^r$;
for $t = 0, 1$ **do**
  for $j = 1, 2, \ldots, r$ **do**
    Compute $\hat{u}_j^{(t+0.5)} = \mathcal{A} \times_2 \hat{v}_j^{(t)\top} \times_3 \hat{w}_j^{(t)\top}$; Update $\hat{u}_j^{(t+1)} = \hat{u}_j^{(t+0.5)} \|\hat{u}_j^{(t+0.5)}\|_2^{-1}$;
    Compute $\hat{v}_j^{(t+0.5)} = \mathcal{A} \times_1 \hat{u}_j^{(t)\top} \times_3 \hat{w}_j^{(t)\top}$; Update $\hat{v}_j^{(t+1)} = \hat{v}_j^{(t+0.5)} \|\hat{v}_j^{(t+0.5)}\|_2^{-1}$;
    Compute $\hat{w}_j^{(t+0.5)} = \mathcal{A} \times_1 \hat{u}_j^{(t)\top} \times_2 \hat{v}_j^{(t)\top}$; Update
    $\hat{w}_j^{(t+1)} = \hat{w}_j^{(t+0.5)} \|\hat{w}_j^{(t+0.5)}\|_2^{-1}$;
  end
end
Output: $\hat{u}_j = \hat{u}_j^{(2)}$, $\hat{v}_j = \hat{v}_j^{(2)}$ and $\hat{w}_j = \hat{w}_j^{(2)}$ for all $j = 1, \ldots, r$.

Since our primary interest is about the statistical inference for $\{u_j, v_j, w_j\}$, we assume that the initializations of Algorithm 3 satisfies the following Assumption 3. Such an assumption is achievable by the power iteration method with k-means initialization introduced in [1] along with the theoretical guarantees developed in [47] when $\lambda/\sigma \geq Cp^{3/4}$.

ASSUMPTION 3. Under the tensor PCA model (1.1) with $\mathcal{T}$ being orthogonally decomposable as (4.1), there is an event $\mathcal{E}_0$ with $\mathbb{P}(\mathcal{E}_0) \geq 1 - C_1 e^{-c_1 p}$ for some absolute constants $c_1, C_1 > 0$ such that, under $\mathcal{E}_0$, the initializations $\{\hat{u}_j^{(0)}, \hat{v}_j^{(0)}, \hat{w}_j^{(0)}\}_j$ satisfy $\max\{\|\hat{u}_{\pi(j)}^{(0)} - u_j\|_2, \|\hat{v}_{\pi(j)}^{(0)} - v_j\|_2, \|\hat{w}_{\pi(j)}^{(0)} - w_j\|_2\} \leq C_2\sigma\sqrt{p}/\lambda_j$ for some permutation $\pi : [r] \to [r]$, all $1 \leq j \leq r$, and some absolute constant $C_2 > 0$.

We establish the asymptotic normality for the outcome of Algorithm 3 as follows.

THEOREM 6 (PCA for orthogonally decomposable tensors). *Suppose Assumption 3 holds for tensor PCA model (1.1) with an orthogonally decomposable $\mathcal{T}$ as (4.1), $\mathcal{Z}(i_1, i_2, i_3) \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, $p_j \asymp p$ for $j = 1, 2, 3$ and $\kappa(\mathcal{T}) \leq \kappa_0$. Let $\{\hat{u}_j, \hat{v}_j, \hat{w}_j\}_{j=1}^r$ be the output of Algorithm 3. There exist absolute constants $c_1, C_0, C_1, C_2, C_3 > 0$ such that if $\lambda_{\min}/\sigma \geq C_0(p^{3/4} + \kappa_0^2 p^{1/2})$, then*

$$\begin{aligned} &\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\langle \hat{u}_{\pi(j)}, u_j\rangle^2 - (1 - p_j\sigma^2\lambda_j^{-2})}{\sqrt{2p_j}\sigma^2\lambda_j^{-2}} \leq x \right) - \Phi(x) \right| \\ &\qquad \leq C_1 e^{-c_1 p} + C_2\left(\frac{\kappa_0^6\sigma^2(pr)^{3/2}}{\lambda_{\min}^2} + \frac{\kappa_0^2\sigma(p\log p)^{1/2}}{\lambda_{\min}} \right) + C_3\frac{r^{3/2}}{\sqrt{p}} \end{aligned} \tag{4.2}$$

for all $j = 1, \ldots, r$. Here, $\pi(\cdot)$ is the permutation introduced in Assumption 3. Moreover, let $\hat{\lambda}_j = \|\mathcal{A} \times_2 \hat{v}_j^\top \times_3 \hat{w}_j^\top\|_2$. Then, (4.2) also holds if λ_j is replaced by $\hat{\lambda}_j$ and $\kappa_0^2\sigma(p\log p)^{1/2}\lambda_{\min}^{-1}$ is replaced by $\kappa_0^3\sigma\sqrt{pr(r^2 + \log p)}\lambda_{\min}^{-1}$. Similar results also hold for $\langle \hat{v}_{\pi(j)}, v_j\rangle^2$ and $\langle \hat{w}_{\pi(j)}, w_j\rangle^2$.

By Theorem 6, if $\lambda_{\min}/\sigma \gg \kappa_0^3(pr)^{3/4} + \kappa_0^2(p\log p)^{1/2}$ and $r \ll p^{1/3}$, then for each $j = 1, \ldots, r$,

$$\frac{\langle \hat{u}_{\pi(j)}, u_j\rangle^2 - (1 - p_j\sigma^2\lambda_j^{-2})}{\sqrt{2p_j}\sigma^2\lambda_j^{-2}} \xrightarrow{\text{d.}} N(0,1) \quad \text{as } p \to \infty.$$

Similar to Section 3.2, we plug in data-driven estimates of λ_j and σ^2 and construct a $(1-\alpha)$ confidence region for u_j as

$$\begin{aligned} \mathrm{CR}_\alpha(\hat{u}_{\pi(j)}) := \big\{ & v \in \mathbb{R}^{p_j} : \|v\|_2 = 1 \text{ and} \\ & \langle \hat{u}_{\pi(j)}, v\rangle^2 \geq \big(1 - p_j\hat{\sigma}^2\hat{\lambda}_{\pi(j)}^{-2}\big) - z_\alpha\sqrt{2p_j}\hat{\sigma}^2\hat{\lambda}_{\pi(j)}^{-2}\big\}. \end{aligned} \tag{4.3}$$

The confidence region for v_j, w_j can be constructed similarly.

5. Entrywise inference for rank-1 tensors. In this section, we consider the statistical inference for tensor PCA model with a rank-1 signal tensor:

$$\mathcal{A} = \mathcal{T} + \mathcal{Z}, \quad \mathcal{T} = \lambda \cdot u \otimes v \otimes w. \tag{5.1}$$

Here, $u \in \mathbb{S}^{p_1-1}$, $v \in \mathbb{S}^{p_2-1}$, $w \in \mathbb{S}^{p_3-1}$, the singular value $\lambda > 0$, and $\mathcal{Z} \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$. We specifically aim to study the inference for any linear form of u, v, w, that is, $\langle q_1, u\rangle$, $\langle q_2, v\rangle$ and $\langle q_3, w\rangle$, with arbitrary deterministic unit vectors $\{q_1, q_2, q_3\}$. We also aim to study the inference for each entry $\mathcal{T}_{ijk}, i \in [p_1],\ j \in [p_2],\ k \in [p_3]$. To this end, we first apply the rank-1 power iteration in Algorithm 4 [55, 82]. Algorithm 4 can be roughly seen as a rank-1 special case of Algorithm 3 for the Tucker low-rank tensor PCA and Algorithm 1 for the orthogonally decomposable tensor PCA.

Next, we establish the asymptotic normality for the output of Algorithm 4, $\hat{u}$, $\hat{v}$, $\hat{w}$, under the essential SNR condition that ensures tensor PCA is solvable in polynomial time. Without loss of generality, we assume that the signs of $\hat{u}$, $\hat{v}$, $\hat{w}$ satisfy $\langle \hat{u}, u\rangle \geq 0$, $\langle \hat{v}, v\rangle \geq 0$ and $\langle \hat{w}, w\rangle \geq 0$ (otherwise one can flip the sign of $\hat{u}$, $\hat{v}$, $\hat{w}$ without changing the problem essentially). With a slight abuse of notation, let u_i, v_j and w_k be the ith entry of u, the jth entry of v, and the kth entry of w, respectively.

THEOREM 7. *Consider the tensor PCA model* (1.1) *with Gaussian noise* $\mathcal{Z}(i_1,i_2,i_3) \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$ *and* $\mathrm{rank}(\mathcal{T}) = 1$, $p_j \asymp p$ *for* $j = 1,2,3$. *Let* $\hat{\lambda}$, $\hat{u}$, $\hat{v}$, $\hat{w}$, $\hat{\mathcal{T}}$ *be the outputs of Algorithm 4 with iteration number* $t_{\max} \geq C_1\log(p)$ *for constant* $C_1 > 0$. *Suppose* $\lambda/\sigma \gg p^{3/4}$.

Algorithm 4: Power iterations for rank-1 tensor $\mathcal{T}$

Input: $\mathcal{A}$

Initialize $\hat{u}^{(0)} = \mathrm{SVD}_1(\mathcal{M}_1(\mathcal{A}))$, $\hat{v}^{(0)} = \mathrm{SVD}_1(\mathcal{M}_2(\mathcal{A}))$, $\hat{w}^{(0)} = \mathrm{SVD}_1(\mathcal{M}_3(\mathcal{A}))$, $t = 1$;

while $t < t_{\max}$ **do**

  Compute $\hat{u}^{(t+0.5)} = \mathcal{A} \times_2 \hat{v}^{(t)\top} \times_3 \hat{w}^{(t)\top}$; Update $\hat{u}^{(t+1)} = \hat{u}^{(t+0.5)}\|\hat{u}^{(t+0.5)}\|_2^{-1}$;

  Compute $\hat{v}^{(t+0.5)} = \mathcal{A} \times_1 \hat{u}^{(t)\top} \times_3 \hat{w}^{(t)\top}$; Update $\hat{v}^{(t+1)} = \hat{v}^{(t+0.5)}\|\hat{v}^{(t+0.5)}\|_2^{-1}$;

  Compute $\hat{w}^{(t+0.5)} = \mathcal{A} \times_1 \hat{u}^{(t)\top} \times_2 \hat{v}^{(t)\top}$; Update $\hat{w}^{(t+1)} = \hat{w}^{(t+0.5)}\|\hat{w}^{(t+0.5)}\|_2^{-1}$;

  $t = t + 1$;

end

$\hat{\lambda} = \mathcal{A} \times_1 \hat{u}^{(t_{\max})\top} \times_2 \hat{v}^{(t_{\max})\top} \times_3 \hat{w}^{(t_{\max})\top}$;

$\hat{\mathcal{T}} = \hat{\lambda}(\hat{u}^{(t_{\max})} \otimes \hat{v}^{(t_{\max})} \otimes \hat{w}^{(t_{\max})})$;

Output: $\hat{u} = \hat{u}^{(t_{\max})}$, $\hat{v} = \hat{v}^{(t_{\max})}$ $\hat{w} = \hat{w}^{(t_{\max})}$, $\hat{\lambda}$ and $\hat{\mathcal{T}}$.

For any deterministic array $\{q_1^{(k)}, q_2^{(k)}, q_3^{(k)}\}_{k=1}^{\infty}$ *satisfying* $q_i^{(k)} \in \mathbb{S}^{k-1}$, *denote*

$$T_{q_1^{(p_1)}, q_2^{(p_2)}, q_3^{(p_3)}} = \Bigg(\frac{\langle q_1^{(p_1)}, \hat{u} - u \rangle + \frac{p_1 \langle q_1^{(p_1)}, u \rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_1 \langle q_1^{(p_1)}, u \rangle^2}{2(\lambda/\sigma)^4} + \frac{1 - \langle q_1^{(p_1)}, u \rangle^2}{(\lambda/\sigma)^2}}}, \frac{\langle q_2^{(p_2)}, \hat{v} - v \rangle + \frac{p_2 \langle q_2^{(p_2)}, v \rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_2 \langle q_2^{(p_2)}, v \rangle^2}{2(\lambda/\sigma)^4} + \frac{1 - \langle q_2^{(p_2)}, v \rangle^2}{(\lambda/\sigma)^2}}}, \frac{\langle q_3^{(p_3)}, \hat{w} - w \rangle + \frac{p_3 \langle q_3^{(p_3)}, w \rangle}{2(\lambda/\sigma)^2}}{\sqrt{\frac{p_3 \langle q_3^{(p_3)}, w \rangle^2}{2(\lambda/\sigma)^4} + \frac{1 - \langle q_3^{(p_3)}, w \rangle^2}{(\lambda/\sigma)^2}}} \Bigg)^{\top}.$$

Then

$$T_{q_1^{(p_1)}, q_2^{(p_2)}, q_3^{(p_3)}} \xrightarrow{\text{d.}} N(0, I_3) \quad \text{as } p \to \infty. \tag{5.2}$$

Specifically, if $|u_i|, |v_j|, |w_k| \ll \min\{\lambda/(\sigma p), 1\}$ *for some* $i \in [p_1]$, $j \in [p_2]$, $k \in [p_3]$, *then*

$$\Big(\frac{\lambda}{\sigma}(\hat{u}_i - u_i), \frac{\lambda}{\sigma}(\hat{v}_j - v_j), \frac{\lambda}{\sigma}(\hat{w}_k - w_k) \Big)^{\top} \xrightarrow{\text{d.}} N(0, I_3) \quad \text{as } p \to \infty. \tag{5.3}$$

If, furthermore, $\sigma/\lambda \ll |u_i|, |v_j|, |w_k| \ll \min\{\lambda/(\sigma p), 1/\sqrt{\log(p)}\}$, *then*

$$\frac{\hat{\mathcal{T}}_{ijk} - \mathcal{T}_{ijk}}{\sigma \sqrt{\hat{u}_i^2 \hat{v}_j^2 + \hat{v}_j^2 \hat{w}_k^2 + \hat{w}_k^2 \hat{u}_i^2}} \xrightarrow{\text{d.}} N(0, 1) \quad \text{as } p \to \infty. \tag{5.4}$$

Theorem 7 establishes the asymptotic distribution for any linear functional $q_1^{\top} \hat{u}$, $q_2^{\top} \hat{v}$, $q_3^{\top} \hat{w}$. Theorem 7 also implies that $[\hat{\mathcal{T}}_{ijk} - z_{\alpha/2} \sigma \sqrt{\hat{u}_i^2 \hat{v}_j^2 + \hat{v}_j^2 \hat{w}_k^2 + \hat{w}_k^2 \hat{u}_i^2}, \hat{\mathcal{T}}_{ijk} + z_{\alpha/2} \times \sigma \sqrt{\hat{u}_i^2 \hat{v}_j^2 + \hat{v}_j^2 \hat{w}_k^2 + \hat{w}_k^2 \hat{u}_i^2}]$ is an asymptotic $(1 - \alpha)$ confidence interval for T_{ijk} under some boundedness conditions of $|u_i|, |v_j|, |w_k|$. Here, the upper bound $|u_i|, |v_j|, |w_k| \ll \min\{\lambda/(\sigma p), 1/\sqrt{\log(p)}\}$ is significantly weaker than the incoherence condition commonly used in the matrix/tensor estimation/inference literature. On the other hand, the lower bound condition, $|u_i|, |v_j|, |w_k| \gg \sigma/\lambda$, is essential to ensure the asymptotic normality of $\hat{\mathcal{T}}$. To see this, consider a special case that $u_i = v_j = w_k = 0$, then (5.3) implies

$$\frac{\lambda^2 \hat{\mathcal{T}}_{ijk}}{\sigma^3} \xrightarrow{\text{d.}} G_1 G_2 G_3 \text{ as } p \to \infty, \quad (G_1, G_2, G_3)^{\top} \sim N(0, I_3).$$

In other words, $\hat{\mathcal{T}}_{ijk}$ satisfies a third moment Gaussian, not a Gaussian distribution.

To cover the broader scenarios that the lower bound conditions are absent, we consider the following lower-thresholding procedure. Let $s(t) = \max\{t, \log(p)\sigma^2 \hat{\lambda}^{-2}\}$[2] for $t \geq 0$ and define the confidence interval for $\mathcal{T}_{ijk}$ as

$$\begin{aligned} \widetilde{CI}_{\alpha}(\hat{\mathcal{T}}_{ijk}) := [&\hat{\mathcal{T}}_{ijk} - z_{\alpha/2} \sigma \sqrt{s(\hat{u}_i^2) s(\hat{v}_j^2) + s(\hat{v}_j^2) s(\hat{w}_k^2) + s(\hat{w}_k^2) s(\hat{u}_i^2)}, \\ &\hat{\mathcal{T}}_{ijk} + z_{\alpha/2} \sigma \sqrt{s(\hat{u}_i^2) s(\hat{v}_j^2) + s(\hat{v}_j^2) s(\hat{w}_k^2) + s(\hat{w}_k^2) s(\hat{u}_i^2)}]. \end{aligned} \tag{5.5}$$

We can prove $\widetilde{CI}_{\alpha}(\hat{\mathcal{T}}_{ijk})$ is a valid $(1 - \alpha)$-level asymptotic confidence interval.

[2]Here, $\log(p)$ can be replaced by any value that grows to infinity as p grows.

THEOREM 8. *Suppose the conditions in Theorem 7 hold. If* $\lambda/\sigma \gg p^{3/4}$ *and* $|u_i|$, $|v_j|$, $|w_k| \ll \min\{\lambda/(\sigma p), 1/\sqrt{\log(p)}\}$ *for* $i \in [p_1]$, $j \in [p_2]$, $k \in [p_3]$, *then*

$$\liminf_{p\to\infty} \mathbb{P}\big(\mathcal{T}_{ijk} \in \widetilde{CI}_\alpha(\hat{\mathcal{T}}_{ijk})\big) \geq 1-\alpha. \tag{5.6}$$

REMARK 4 (Proof sketch of Theorem 7). The proof scheme for Theorem 7 is essentially different from many recent literature on the entrywise inference [12, 21, 72] and we provide a proof sketch here. Without loss generality, we assume $\sigma = 1$ and $\langle u, \hat{u}\rangle$, $\langle v, \hat{v}\rangle$, $\langle w, \hat{w}\rangle \geq 0$. First, we can decompose $\langle \hat{u}, q_1\rangle$ into two terms:

$$\langle q_1, \hat{u}\rangle = \langle \hat{u}, uu^\top q_1\rangle + \langle \hat{u}, (I-uu^\top)q_1\rangle = (q_1^\top u)\hat{u}^\top u + (U_\perp^\top q_1)^\top U_\perp^\top \hat{u}. \tag{5.7}$$

Similar decompositions hold for $\langle \hat{v}, q_2\rangle$ and $\langle \hat{w}, q_3\rangle$. For any $O_i \in \mathbb{O}_{p_i-1}$, we construct three rotation matrices as

$$\tilde{O}_1 = uu^\top + U_\perp O_1 U_\perp^\top \in \mathbb{O}_{p_1},$$
$$\tilde{O}_2 = vv^\top + V_\perp O_2 V_\perp^\top \in \mathbb{O}_{p_2},$$
$$\tilde{O}_3 = ww^\top + W_\perp O_3 W_\perp^\top \in \mathbb{O}_{p_3},$$

where $U_\perp \in \mathbb{O}_{p_1,p_1-1}$, $V_\perp \in \mathbb{O}_{p_2,p_2-1}$, $W_\perp \in \mathbb{O}_{p_3,p_3-1}$ are the orthogonal complement of u, v, w, respectively. A key observation is that $\tilde{\mathcal{A}} = \mathcal{A} \times_1 \tilde{O}_1^\top \times_2 \tilde{O}_2^\top \times_3 \tilde{O}_3^\top$ and $\mathcal{A}$ share the same distribution. Suppose $\tilde{u}$, $\tilde{v}$, $\tilde{w}$ are the outputs of Algorithm 4. Then we have $\tilde{u} = \tilde{O}_1^\top \hat{u}$, $\tilde{v} = \tilde{O}_2^\top \hat{v}$, $\tilde{w} = \tilde{O}_3^\top \hat{w}$ and can further prove that given $\langle u, \hat{u}\rangle$, $\langle v, \hat{v}\rangle$ and $\langle w, \hat{w}\rangle$, $(\frac{\hat{u}^\top U_\perp}{\|U_\perp^\top \hat{u}\|_2}, \frac{\hat{v}^\top V_\perp}{\|V_\perp^\top \hat{v}\|_2}, \frac{\hat{w}^\top W_\perp}{\|W_\perp^\top \hat{w}\|_2})$ and $(\frac{\hat{u}^\top U_\perp}{\|U_\perp^\top \hat{u}\|_2} O_1, \frac{\hat{v}^\top V_\perp}{\|V_\perp^\top \hat{v}\|_2} O_2, \frac{\hat{w}^\top W_\perp}{\|W_\perp^\top \hat{w}\|_2} O_3)$ have the same distribution. By the uniqueness of the Haar measure [65, 66] and Theorem 1, we can further prove for any fixed vectors $f_1 \in \mathbb{S}^{p_1-2}$, $f_2 \in \mathbb{S}^{p_2-2}$, $f_3 \in \mathbb{S}^{p_3-2}$, we have

$$\Big(\lambda\hat{u}^\top U_\perp f_1, \lambda\hat{v}^\top V_\perp f_2, \lambda\hat{w}^\top W_\perp f_3,$$
$$\frac{\hat{u}^\top u - (1-p_1\lambda^{-2}/2)}{\sqrt{p_1/2}\lambda^{-2}}, \frac{\hat{v}^\top v - (1-p_2\lambda^{-2}/2)}{\sqrt{p_2/2}\lambda^{-2}}, \frac{\hat{w}^\top w - (1-p_3\lambda^{-2}/2)}{\sqrt{p_3/2}\lambda^{-2}}\Big)^\top \xrightarrow{\text{d.}} N(0, I_6).$$

This inequality and (5.7) result in (5.2), (5.3) and (5.4).

REMARK 5. The entrywise inference for Tucker low-rank or orthogonal decomposable tensor PCA can be significantly more challenging due to the dependence among different factors. We leave it as future research.

6. Numerical simulations. We now conduct numerical studies to support our theoretical findings in previous sections. Each experiment is repeated for 2000 times, from which we obtain 2000 realizations of the respective statistics. Then we draw histograms or boxplots, and compare with the corresponding baselines. In each histogram, the red line is the density of the standard normal distribution.

We begin with the inference for principal components of Tucker low-rank tensors. Specifically, we randomly draw $\check{U}_j \in \mathbb{R}^{p_j\times r_j}$ with i.i.d. standard normal entries and normalize to $U_j = \mathrm{QR}(\check{U}_j)$. We then draw core tensor $\check{\mathcal{G}} \in \mathbb{R}^{r\times r\times r}$ with i.i.d. standard normal entries and rescale to $\mathcal{G} = \check{\mathcal{G}} \cdot p^\gamma/(\lambda_{\min}(\check{\mathcal{G}}))$. Consequently, U_j is uniform randomly selected from $\mathbb{O}_{p_j,r_j}$ and $\lambda_{\min}(\mathcal{G}) = \lambda = p^\gamma$. For $p_1 = p_2 = p_3 = 200$, $r = 3$ and $\sigma = 1$, each value of $\gamma \in \{0.80, 0.85, 0.90, 0.95\}$, we observe $\mathcal{A}$ under tensor PCA model (1.1) and apply Algorithm 1 to obtain realizations of

$$T_1 = \frac{\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 - p\|\Lambda_1^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p_1}\|\Lambda_1^{-2}\|_{\mathrm{F}}} \quad \text{and} \quad T_2 = \frac{\|\sin\Theta(\hat{U}_1, U_1)\|_{\mathrm{F}}^2 - p\hat{\sigma}^2\|\hat{\Lambda}_1^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p}\hat{\sigma}^2\|\hat{\Lambda}_1^{-2}\|_{\mathrm{F}}}.$$

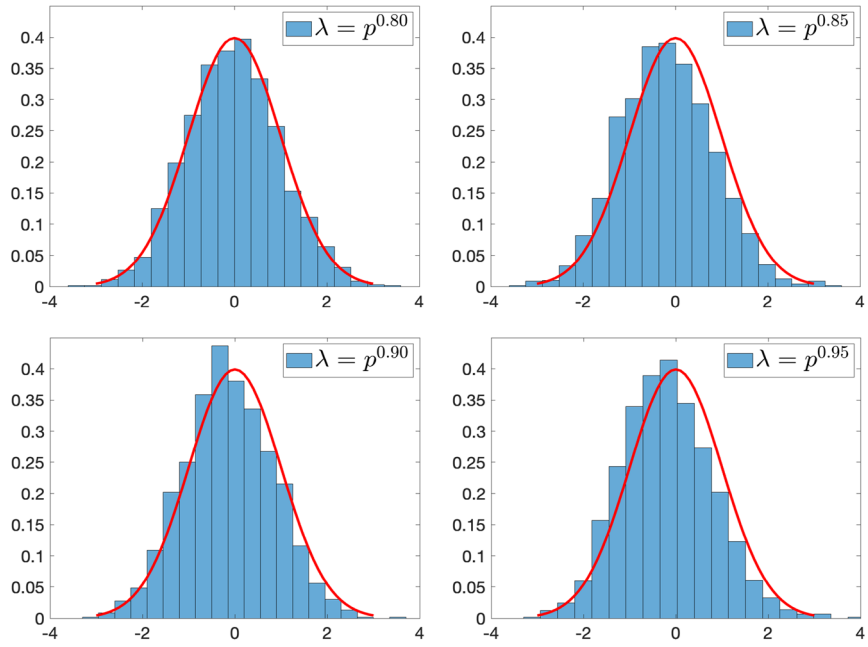

FIG. 1. *Normal approximation of* $\frac{\|\sin\Theta(\hat{U}_1,U_1)\|_{\mathrm{F}}^2 - p\|\Lambda_1^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p}\|\Lambda_1^{-2}\|_{\mathrm{F}}}$ *for order-3 Tucker low-rank tensor PCA model* (1.1). *Here,* $p_1 = p_2 = p_3 = p = 200$, $r = 3$, $\sigma = 1$.

We repeat this procedure for 2000 times, from which we obtain 2000 realizations of the respective statistics and plot the density histograms in Figures 1 and 2, respectively. We can see T_1 and T_2 both achieve good normal approximation in these settings.

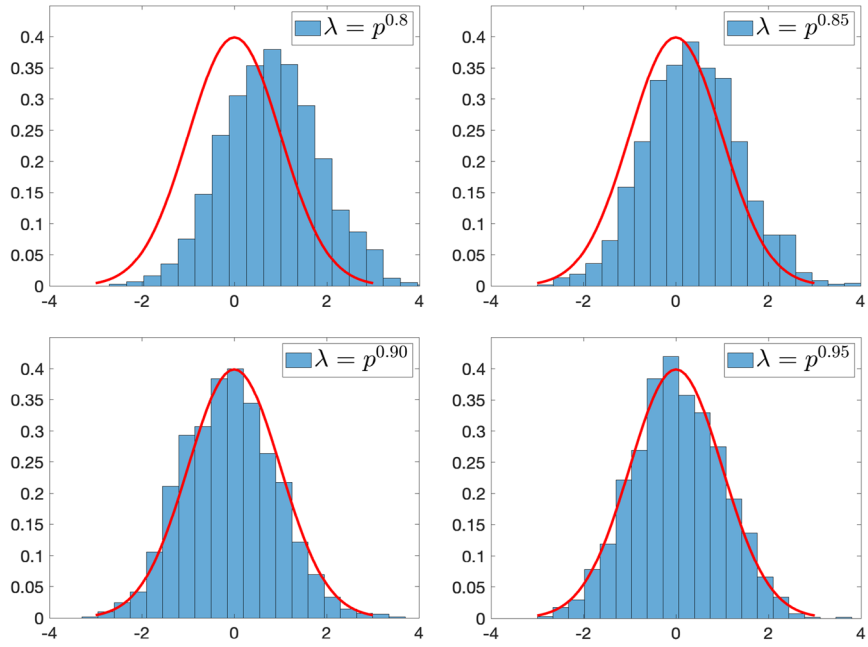

FIG. 2. *Normal approximation of* $\frac{\|\sin\Theta(\hat{U}_1,U_1)\|_{\mathrm{F}}^2 - p\hat{\sigma}^2\|\hat{\Lambda}_1^{-1}\|_{\mathrm{F}}^2}{\sqrt{2p}\hat{\sigma}^2\|\hat{\Lambda}_1^{-2}\|_{\mathrm{F}}}$ *for order-3 Tucker low-rank tensor PCA model* (1.1). *Here,* $p_1 = p_2 = p_3 = p = 200$, $r = 3$, $\sigma = 1$.

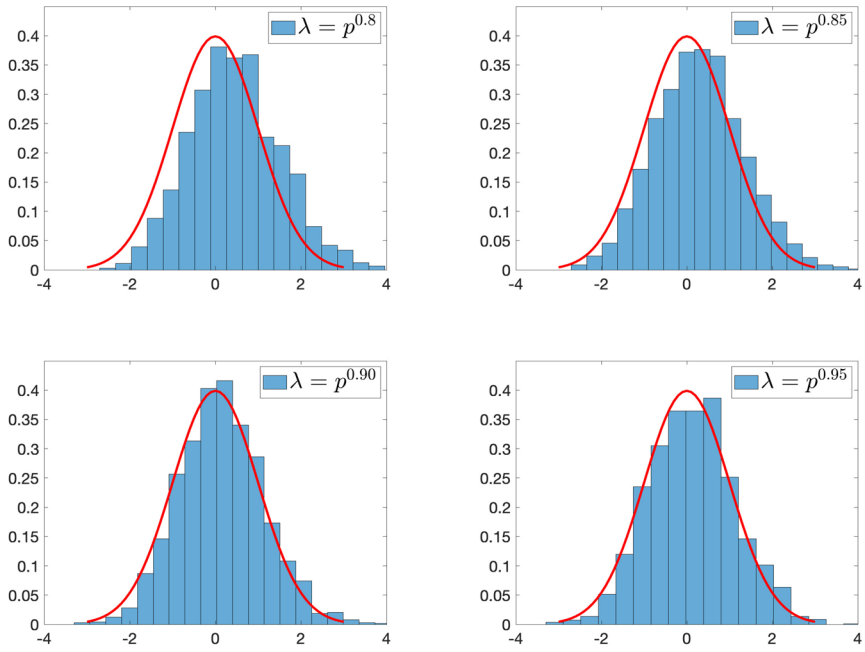

FIG. 3. *Normal approximation of* $\frac{\langle \hat{u}_3, u_3\rangle^2 - (1-p\lambda^{-2})}{\sqrt{2p}\lambda^{-2}}$ *for tensor PCA model* (1.1) *when* $\mathcal{T}$ *is a third-order orthogonally decomposable tensor and* $\sigma = 1$. *Here,* $p_1 = p_2 = p_3 = p = 200$, $r = 3$, $\lambda_{\min} = \lambda$.

We then consider the asymptotic normality in orthogonally decomposable tensors under the tensor PCA model. Similarly, we fix $p = 200$, $r = 3$, and construct the orthogonally decomposable tensor as $\mathcal{T} = \sum_{i=1}^{r}(r+1-i)\lambda \cdot (u_i \otimes v_i \otimes w_i)$, where $[u_1, \ldots, u_r]$, $[v_1, \ldots, v_r]$, $[w_1, \ldots, w_r]$ are drawn uniform randomly from $\mathbb{O}_{p,r}$ similar to the previous setting and $\lambda = p^{\gamma}$ with $\gamma = 0.80, 0.85, 0.90, 0.95$. For each γ, we obtain 2000 replicates of $T = \frac{\langle \hat{u}_3, u_3\rangle^2 - (1-p\lambda^{-2})}{\sqrt{2p}\lambda^{-2}}$, draw the density histogram and plot the results in Figure 3. We can see the normal approximation of T becomes more accurate as the signal strength λ grows.

Though the focus of this paper is on third-order tensors, we will explain later in Section 7 that the results can be generalized to higher-order ones. Next, we conduct simulation study on tensor PCA model for fourth-order orthogonally decomposable tensors when $p = 100$ and $r = 1$. With a few modifications on the proof, we can show $(\langle \hat{u}_1, u_1\rangle^2 - (1 - p\lambda^{-2}))(\sqrt{2p}\lambda^{-2})^{-1}$ is asymptotically normal under the required SNR assumption for efficient computation: SNR $\geq Cp$. The simulation results in Figure 4 show that equipped with a warm initialization, the two-iteration alternating minimization yields an estimator achieving good normal approximation even if SNR $\approx p^{0.9}$, which is strictly weaker than the required SNR assumption for efficient computation. See more discussions in Section 7.

Then we consider the entrywise inference under the rank-1 tensor PCA model. We construct $\mathcal{T} = \lambda \cdot u \otimes v \otimes w \in \mathbb{R}^{p\times p\times p}$, where $u = v = w = (1/\sqrt{p}, \ldots, 1/\sqrt{p})^{\top}$ and $\lambda = p^{\gamma}$ with $\gamma \in \{0.80, 0.85, 0.90, 0.95\}$. For each value of γ, we draw a random observation $\mathcal{A}$ under the tensor PCA model (1.1) and apply Algorithm 4 with $t_{\max} = 10$. We present the histogram in Figure 5 based on 2000 replicate values of $\frac{\hat{\mathcal{T}}_{1,1,1} - \mathcal{T}_{1,1,1}}{\sqrt{\hat{u}_1^2\hat{v}_1^2 + \hat{v}_1^2\hat{w}_1^2 + \hat{w}_1^2\hat{u}_1^2}}$. The simulation results validate the asymptotic normality of $\frac{\hat{\mathcal{T}}_{ijk} - \mathcal{T}_{ijk}}{\sqrt{\hat{u}_i^2\hat{v}_j^2 + \hat{v}_j^2\hat{w}_k^2 + \hat{w}_k^2\hat{u}_i^2}}$ when u, v, w have balanced entry values, which are in line with the theory in Theorem 7.

Finally, we consider the accuracy of the asymptotic entrywise confidence interval proposed in (5.5) under the tensor PCA model. Let $\mathcal{T} = \lambda \cdot u \otimes v \otimes w$ be a rank-1 tensor, where

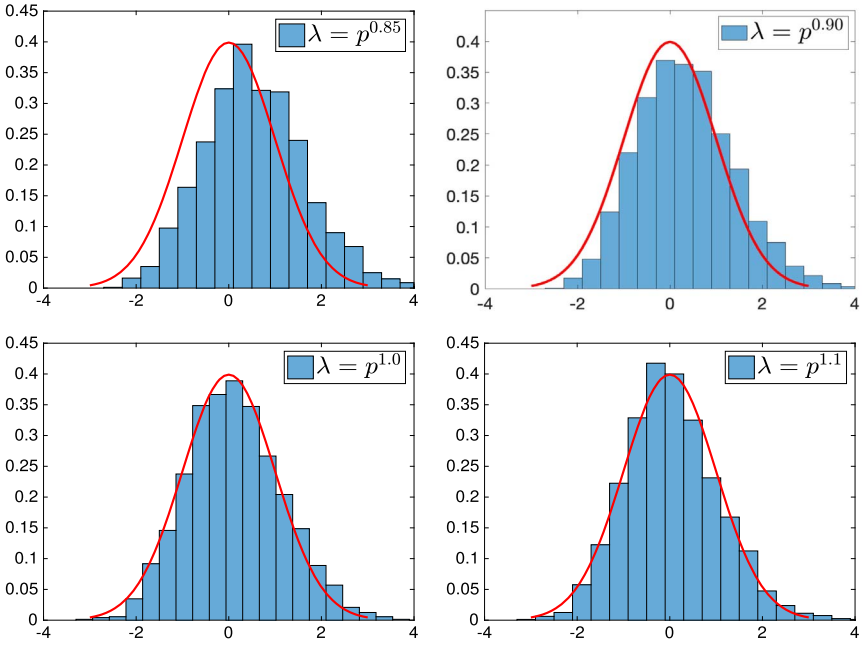

FIG. 4. *Normal approximation of* $\frac{\langle \hat{u}_1, u_1 \rangle^2 - (1 - p\lambda^{-2})}{\sqrt{2p}\lambda^{-2}}$ *for tensor PCA model* (1.1) *when* $\mathcal{T} = \lambda \cdot (u_1 \otimes v_1 \otimes w_1 \otimes q_1)$ *is a fourth-order tensor and* $\sigma = 1$. *Here,* $p_1 = p_2 = p_3 = p_4 = p = 100$, $r = 1$ *and* $\lambda_{\min} = \lambda$.

u, v, w are uniform randomly drawn from $\mathbb{S}^{p-1}$ for $p \in \{100, 200\}$ and $\lambda = p^{\gamma}$ for $\gamma \in \{0.80, 0.85, 0.90, 0.95\}$. For each combination of (p, γ), we report the empirical coverage rates for the 0.95 confidence interval $\widehat{\mathrm{CR}}_{ijk}$ by boxplots in Figure 6. The results show the

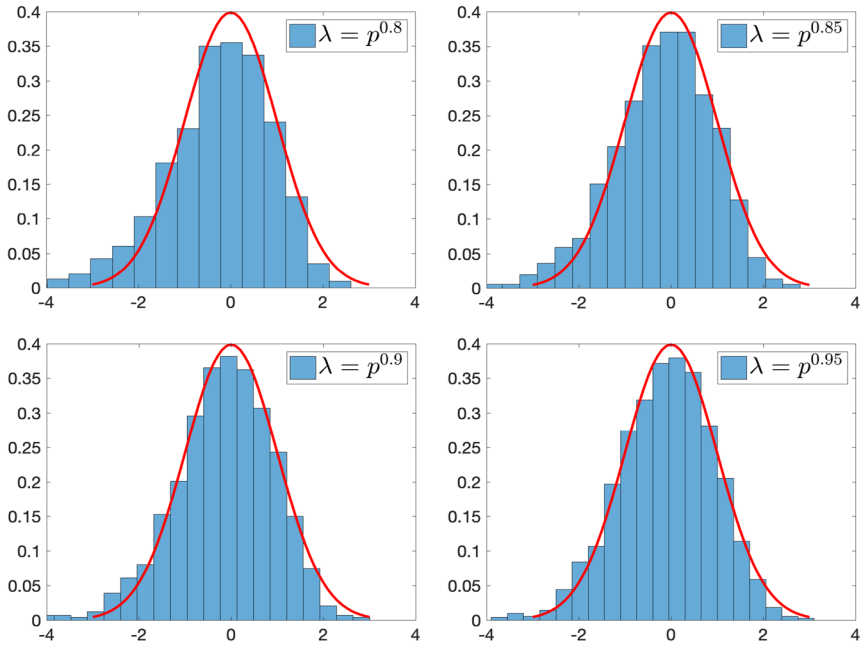

FIG. 5. *Normal approximation of* $\frac{\hat{\mathcal{T}}_{1,1,1} - \mathcal{T}_{1,1,1}}{\sqrt{\hat{u}_1^2\hat{v}_1^2 + \hat{v}_1^2\hat{w}_1^2 + \hat{w}_1^2\hat{u}_1^2}}$ *for tensor PCA model* (1.1) *when* $\mathcal{T}$ *is a rank-*1 *tensor and* $\sigma = 1$. *The parameters are* $p_1 = p_2 = p_3 = p = 200$ *with signal strength* λ.

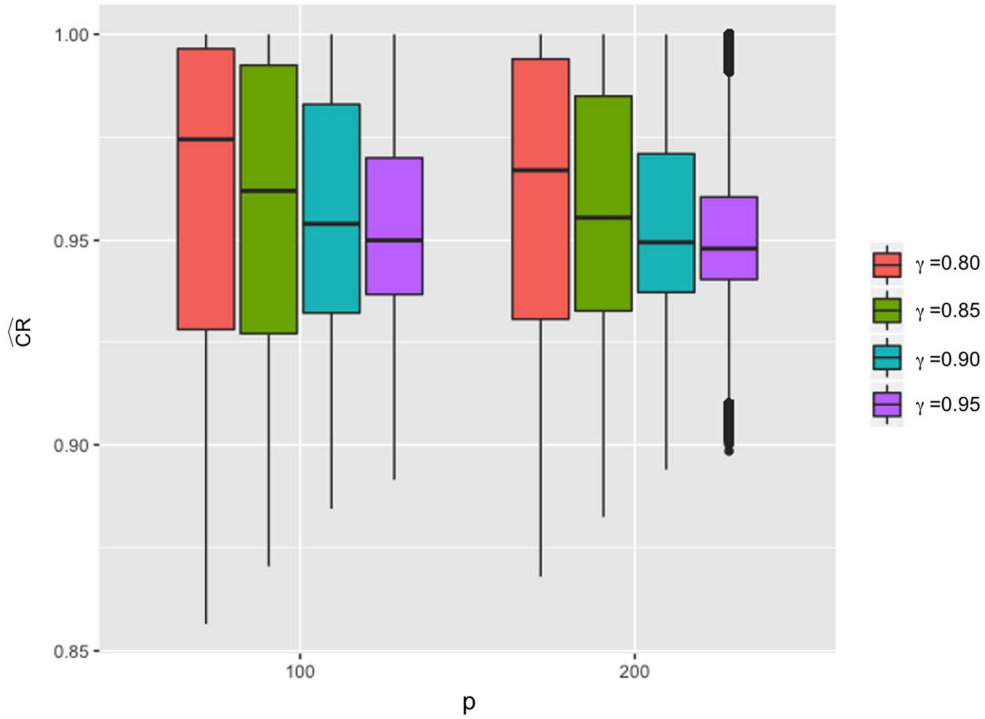

FIG. 6. *Boxplots for empirical coverage of entrywise confidence interval* $\widehat{CR}_{ijk}$.

empirical coverage rates are close to 0.95 in all settings and larger values of (γ, p) lead to more accurate coverage.

7. Discussion. In this paper, we investigate the inference for low-rank tensors under two basic and fundamentally important tensor models: tensor PCA and regression. Based on an initial estimator achieving a reasonable estimation error, we propose to update by a two-iteration alternating minimization algorithm then establish the asymptotic distribution for the singular subspace outcomes. Distributions of general linear forms of the singular vectors are also established for rank-one tensor PCA model, which further enables the entrywise inference on the parameter tensor.

Although our main focus is on third-order tensors, the results in this paper can be extended to higher-order tensors. For example, suppose $m \geq 4$ and $\mathcal{T} = \sum_{j=1}^{r} \lambda_j \cdot u_j^{(1)} \otimes \cdots \otimes u_j^{(m)}$ is orthogonally decomposable. Given $\mathcal{A}$ from the tensor PCA model (1.1) and Assumption 3 holds, we can refine by two power iterations similar to Algorithm 3, then obtain $\{\hat{u}_j^{(1)}, \hat{u}_j^{(2)}, \ldots, \hat{u}_j^{(m)}\}_{j=1}^{r}$. Similar to Theorem 6, we can prove

$$\frac{\langle u_j^{(k)}, \hat{u}_j^{(k)} \rangle^2 - (1 - p_k \lambda_j^{-2} \sigma^2)}{\sqrt{2p_k} \lambda_j^{-2} \sigma^2} \xrightarrow{d.} N(0,1), \quad k = 1, \ldots, m,$$

if $\lambda_{\min}/\sigma \gg p^{3/4}$ and other regularity conditions holds. If $m \geq 4$, the SNR condition $\lambda_{\min} \gg p^{3/4}$ is weaker than the condition that ensure a computationally feasible estimator exists, that is, $\lambda_{\min}/\sigma \gg p^{m/4}$ [77]. In other words, if an sufficiently good initial estimate is already available, a weaker SNR condition $\lambda_{\min}/\sigma \gg p^{3/4}$ is sufficient to guarantee the asymptotic normality of our final estimates. This phenomenon is further justified by the simulation results in Figure 4.

Acknowledgments. The authors thank the Editor, the Associate Editor and three anonymous referees for their comments that helped to improve the presentation of this paper.

The authors are listed alphabetically.

This work was done while Anru R. Zhang and Yuchen Zhou were at the University of Wisconsin-Madison.

Anru R. Zhang is the corresponding author.

Funding. Dong Xia's research was partially supported by Hong Kong RGC Grant ECS 26302019 and GRF 16303320.

Anru R. Zhang and Yuchen Zhou's research was partially supported by NSF Grants CAREER-1944904, NSF DMS-1811868 and grants from Wisconsin Alumni Research Foundation (WARF).

SUPPLEMENTARY MATERIAL

Supplement to "Inference for low-rank tensors—no need to debias" (DOI: 10.1214/21-AOS2146SUPP; .pdf). The supplementary material contains addtional details on algorithms and all technical proofs for the main content of this paper.

REFERENCES

[1] ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832. MR3270750

[2] ANANDKUMAR, A., GE, R. and JANZAMIN, M. (2014). Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates. Preprint. Available at arXiv:1402.5180.

[3] ANANDKUMAR, A., HSU, D. and KAKADE, S. M. (2012). A method of moments for mixture models and hidden Markov models. In *Conference on Learning Theory* 33–1.

[4] AUDDY, A. and YUAN, M. (2020). Perturbation bounds for orthogonally decomposable tensors and their applications in high dimensional data analysis. Preprint. Available at arXiv:2007.09024.

[5] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

[6] BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory* 417–445.

[7] BELKIN, M., RADEMACHER, L. and VOSS, J. (2018). Eigenvectors of orthogonally decomposable functions. *SIAM J. Comput.* **47** 547–615. MR3794351 https://doi.org/10.1137/17M1122013

[8] BEN AROUS, G., MEI, S., MONTANARI, A. and NICA, M. (2019). The landscape of the spiked tensor model. *Comm. Pure Appl. Math.* **72** 2282–2330. MR4011861 https://doi.org/10.1002/cpa.21861

[9] BI, X., QU, A. and SHEN, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Ann. Statist.* **46** 3308–3333. MR3852653 https://doi.org/10.1214/17-AOS1659

[10] BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. Preprint. Available at arXiv:2005.08099.

[11] CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems* 1863–1874.

[12] CAI, C., POOR, H. V. and CHEN, Y. (2020). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. Preprint. Available at arXiv:2006.08580.

[13] CAI, T. T., LIANG, T. and RAKHLIN, A. (2016). Geometric inference for general high-dimensional linear inverse problems. *Ann. Statist.* **44** 1536–1563. MR3519932 https://doi.org/10.1214/15-AOS1426

[14] CAI, T. T. and ZHANG, A. (2014). Sparse representation of a polytope and recovery in sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **60** 122–132. MR3150915 https://doi.org/10.1109/TIT.2013.2288639

[15] CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

[16] CANDÈS, E. J., STROHMER, T. and VORONINSKI, V. (2013). PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.* **66** 1241–1274. MR3069958 https://doi.org/10.1002/cpa.21432

[17] CARPENTIER, A., EISERT, J., GROSS, D. and NICKL, R. (2019). Uncertainty quantification for matrix compressed sensing and quantum tomography problems. In *High Dimensional Probability VIII—the Oaxaca Volume. Progress in Probability* **74** 385–430. Birkhäuser/Springer, Cham. MR4181374 https://doi.org/10.1007/978-3-030-26391-1_18

[18] CHEN, H., RASKUTTI, G. and YUAN, M. (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *J. Mach. Learn. Res.* **20** Paper No. 5, 37. MR3911412

[19] CHEN, J. and SAAD, Y. (2008/09). On the tensor SVD and the optimal low rank orthogonal approximation of tensors. *SIAM J. Matrix Anal. Appl.* **30** 1709–1734. MR2486861 https://doi.org/10.1137/070711621

[20] CHEN, W.-K. (2019). Phase transition in the spiked random tensor with Rademacher prior. *Ann. Statist.* **47** 2734–2756. MR3988771 https://doi.org/10.1214/18-AOS1763

[21] CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

[22] CHI, E. C., GAINES, B. R., SUN, W. W., ZHOU, H. and YANG, J. (2020). Provable convex co-clustering of tensors. *J. Mach. Learn. Res.* **21** Paper No. 214, 58. MR4209500

[23] DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000). On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995

[24] DE SILVA, V. and LIM, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* **30** 1084–1127. MR2447444 https://doi.org/10.1137/06066518X

[25] DE DOMENICO, M., NICOSIA, V., ARENAS, A. and LATORA, V. (2015). Structural reducibility of multilayer networks. *Nat. Commun.* **6** 1–9.

[26] DUDEJA, R. and HSU, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** Paper No. 83, 51. MR4253776

[27] ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1** 211–218.

[28] FAN, J., GONG, W. and ZHU, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *J. Econometrics* **212** 177–202. MR3994013 https://doi.org/10.1016/j.jeconom.2019.04.026

[29] FEIZI, S., JAVADI, H. and TSE, D. (2017). Tensor biclustering. In *Advances in Neural Information Processing Systems* 1311–1320.

[30] FOUCART, S., NEEDELL, D., PLAN, Y. and WOOTTERS, M. (2017). De-biasing low-rank projection for matrix completion. In *Wavelets and Sparsity XVII* **10394** 1039417. International Society for Optics and Photonics. https://doi.org/10.1117/12.2275004.

[31] HAN, R., LUO, Y., WANG, M. and ZHANG, A. R. (2020). Exact clustering in tensor block model: Statistical optimality and computational limit. Preprint. Available at arXiv:2012.09996.

[32] HAN, R., WILLETT, R. and ZHANG, A. (2020). An optimal statistical and computational framework for generalized tensor estimation. Preprint. Available at arXiv:2002.11255.

[33] HAO, B., ZHANG, A. and CHENG, G. (2020). Sparse and low-rank tensor estimation via cubic sketchings. *IEEE Trans. Inf. Theory* **66** 5927–5964. MR4158653 https://doi.org/10.1109/TIT.2020.2982499

[34] HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** Art. 45, 39. MR3144915 https://doi.org/10.1145/2512329

[35] HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory* 956–1006.

[36] HUANG, J., HUANG, D. Z., YANG, Q. and CHENG, G. (2020). Power iteration for tensor PCA. Preprint. Available at arXiv:2012.13669.

[37] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

[38] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079

[39] KARATZOGLOU, A., AMATRIAIN, X., BALTRUNAS, L. and OLIVER, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 79–86.

[40] KE, Z. T., SHI, F. and XIA, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. Preprint. Available at arXiv:1909.06503.

[41] KOLDA, T. G. (2001). Orthogonal tensor decompositions. *SIAM J. Matrix Anal. Appl.* **23** 243–255. MR1856608 https://doi.org/10.1137/S0895479800368354

[42] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

[43] KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869 https://doi.org/10.1214/11-AOS894

[44] KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798
[45] LI, X., LING, S., STROHMER, T. and WEI, K. (2019). Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Appl. Comput. Harmon. Anal.* **47** 893–934. MR3994997 https://doi.org/10.1016/j.acha.2018.01.001
[46] LI, X., XU, D., ZHOU, H. and LI, L. (2018). Tucker tensor regression and neuroimaging analysis. *Stat. Biosci.* **10** 520–545.
[47] LIU, T., YUAN, M. and ZHAO, H. (2017). Characterizing spatiotemporal transcriptome of human brain via low rank tensor decomposition. Preprint. Available at arXiv:1702.07449.
[48] LUO, Y. and ZHANG, A. R. (2020). Tensor clustering with planted structures: Statistical optimality and computational limits. Preprint. Available at arXiv:2005.10743.
[49] LUO, Y. and ZHANG, A. R. (2020). Open problem: Average-case hardness of hypergraphic planted clique detection. In *Conference of Learning Theory* (*COLT*) **125** 3852–3856.
[50] LYNCH, R. E., RICE, J. R. and THOMAS, D. H. (1964). Tensor product analysis of partial difference equations. *Bull. Amer. Math. Soc.* **70** 378–384. MR0169390 https://doi.org/10.1090/S0002-9904-1964-11105-8
[51] MONTANARI, A. and SUN, N. (2018). Spectral algorithms for tensor completion. *Comm. Pure Appl. Math.* **71** 2381–2425. MR3862094 https://doi.org/10.1002/cpa.21748
[52] PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. MR4058987 https://doi.org/10.1214/19-AIHP960
[53] RASKUTTI, G., YUAN, M. and CHEN, H. (2019). Convex regularization for high-dimensional multiresponse tensor regression. *Ann. Statist.* **47** 1554–1584. MR3911122 https://doi.org/10.1214/18-AOS1725
[54] RAUHUT, H., SCHNEIDER, R. and STOJANAC, Ž. (2017). Low rank tensor recovery via iterative hard thresholding. *Linear Algebra Appl.* **523** 220–262. MR3624675 https://doi.org/10.1016/j.laa.2017.02.028
[55] RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.
[56] ROBEVA, E. (2016). Orthogonal decomposition of symmetric tensors. *SIAM J. Matrix Anal. Appl.* **37** 86–102. MR3530392 https://doi.org/10.1137/140989340
[57] SHAH, D. and YU, C. L. (2019). Iterative collaborative filtering for sparse noisy tensor estimation. In 2019 *IEEE International Symposium on Information Theory* (*ISIT*) 41–45. IEEE.
[58] SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** Paper No. 135, 37. MR3763769
[59] SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. MR4047308 https://doi.org/10.1080/01621459.2018.1527701
[60] SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 899–916. MR3641413 https://doi.org/10.1111/rssb.12190
[61] TOMIOKA, R. and SUZUKI, T. (2013). Convex tensor decomposition via structured Schatten norm regularization. In *Advances in Neural Information Processing Systems* 1331–1339.
[62] TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395 https://doi.org/10.1007/BF02289464
[63] VANNIEUWENHOVEN, N., VANDEBRIL, R. and MEERBERGEN, K. (2012). A new truncation strategy for the higher-order singular value decomposition. *SIAM J. Sci. Comput.* **34** A1027–A1052. MR2914314 https://doi.org/10.1137/110836067
[64] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221
[65] VON NEUMANN, J. (1936). The uniqueness of Haar's measure. *Mat. Sb.* **1** 721–734.
[66] WEIL, A. (1940). *L'intégration dans les Groupes Topologiques et Ses Applications*. *Actualités Scientifiques et Industrielles* [*Current Scientific and Industrial Topics*], *No.* 869. Hermann & Cie, Paris. MR0005741
[67] WILMOTH, J. R. and SHKOLNIKOV, V. (2006). Human mortality database.
[68] WU, T., BENSON, A. R. and GLEICH, D. F. (2016). General tensor spectral co-clustering for higher-order data. In *Advances in Neural Information Processing Systems* 2559–2567.
[69] XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900
[70] XIA, D. (2021). Normal approximation and confidence region of singular subspaces. *Electron. J. Stat.* **15** 3798–3851. MR4298986 https://doi.org/10.1214/21-ejs1876
[71] XIA, D. and YUAN, M. (2019). On polynomial time methods for exact low-rank tensor completion. *Found. Comput. Math.* **19** 1265–1313. MR4029842 https://doi.org/10.1007/s10208-018-09408-6

[72] XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 58–77. MR4220984 https://doi.org/10.1111/rssb.12400

[73] XIA, D., ZHANG, A. R. and ZHOU, Y. (2022). Supplement to "Inference for Low-rank Tensors—No Need to Debias." https://doi.org/10.1214/21-AOS2146SUPP

[74] YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

[75] ZHANG, A. (2019). Cross: Efficient low-rank tensor completion. *Ann. Statist.* **47** 936–964. MR3909956 https://doi.org/10.1214/18-AOS1694

[76] ZHANG, A. and HAN, R. (2019). Optimal sparse singular value decomposition for high-dimensional high-order data. *J. Amer. Statist. Assoc.* **114** 1708–1725. MR4047294 https://doi.org/10.1080/01621459.2018.1527227

[77] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

[78] ZHANG, A. R., LUO, Y., RASKUTTI, G. and YUAN, M. (2020). ISLET: Fast and optimal low-rank tensor regression via importance sketching. *SIAM J. Math. Data Sci.* **2** 444–479. MR4106613 https://doi.org/10.1137/19M126476X

[79] ZHANG, C., HAN, R., ZHANG, A. R. and VOYLES, P. M. (2020). Denoising atomic resolution 4D scanning transmission electron microscopy data with tensor singular value decomposition. *Ultramicroscopy* 113123.

[80] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448 https://doi.org/10.1214/07-AOS520

[81] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

[82] ZHANG, T. and GOLUB, G. H. (2001). Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.* **23** 534–550. MR1871328 https://doi.org/10.1137/S0895479899352045

[83] ZHENG, Q. and TOMIOKA, R. (2015). Interpolating convex and non-convex tensor decompositions via the subspace norm. In *Advances in Neural Information Processing Systems* 3106–3113.

[84] ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.* **108** 540–552. MR3174640 https://doi.org/10.1080/01621459.2013.776499