Statistical and Computational Limits for Tensor-on-Tensor Association Detection[1]

Ilias Diakonikolas ILIAS@CS.WISC.EDU
University of Wisconsin-Madison

Daniel M. Kane DAKANE@CS.UCSD.EDU
University of California, San Diego

Yuetian Luo YUETIAN@UCHICAGO.EDU
The University of Chicago

Anru R. Zhang ANRU.ZHANG@DUKE.EDU
Duke University

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

In this paper, we consider the tensor-on-tensor association detection problem, where the goal is to detect whether there is an association between the tensor responses to tensor covariates linked via a low-rank tensor parameter. We first develop tight bounds on the signal-to-noise ratio (SNR) such that the detection problem is statistically possible. We then provide testing procedures that succeed when the SNR is above the threshold. On the other hand, the statistical optimal tests often require computing the largest singular value of a given tensor, which can be NP-hard in general. To complement that, we develop efficient polynomial-time testing procedures with provable guarantees. We also develop matching lower bounds under the Statistical Query model and show that the SNRs required by the proposed polynomial-time algorithms are essential for computational efficiency. We identify a gap that appears between the SNR requirements of the optimal unconstrained-time tests and polynomial-time tests if and only if the sum of the tensor response order and the tensor covariate order is no less than three. To our best knowledge, this is the first complete characterization of the statistical and computational limits for the general tensor-on-tensor association detection problem. Our findings significantly generalize the results in the literature on signal detection in linear regression and low-rank matrix trace regression. Finally, the connection on the computational hardness of the detection problem and the corresponding estimation problem is discussed.

Keywords: Hypothesis testing, minimax separation rate, computational separation rate, statistical and computational gap, tensor

1. Introduction

The analysis of tensor or multiway array data has emerged as an active topic of research in machine learning, statistics, applied mathematics, and signal processing. A general class of problems in tensor learning aims to characterize the association between covariates and responses in the form of scalars, vectors, matrices, or high-order tensors. These tasks can be incorporated in the following *tensor-on-tensor regression* model (Lock, 2018; Raskutti et al., 2019):

$$\boldsymbol{\mathcal{Y}}_i = \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{A}} \rangle_* + \boldsymbol{\mathcal{E}}_i, \quad i = 1, \ldots, n. \tag{1}$$

1. The authors are listed alphabetically.

Here, $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{p_1 \times \cdots \times p_d}, i = 1, \ldots, n$ are the known order-d (or d-way) tensor covariates. $\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{E}}_i \in \mathbb{R}^{p_{d+1} \times \cdots \times p_{d+m}}$ are both order-m tensors and are observations and unknown noise, respectively. $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{p_1 \times \cdots \times p_d \times p_{d+1} \times \cdots \times p_{d+m}}$ is an order-$(d+m)$ tensor parameter of interest. $\langle \cdot, \cdot \rangle_*$ is the contracted tensor inner product defined as $\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{A}} \rangle_* \in \mathbb{R}^{p_{d+1} \times \cdots \times p_{d+m}}$,

$$(\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{A}} \rangle_*)_{[j_1, \ldots, j_m]} = \sum_{\substack{k_l = 1, \\ l = 1, \ldots, d}}^{p_l} \boldsymbol{\mathcal{X}}_{i[k_1, \ldots, k_d]} \boldsymbol{\mathcal{A}}_{[k_1, \ldots, k_d, j_1, \ldots, j_m]}.$$

We also stack all responses and errors to $\boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{E}} \in \mathbb{R}^{p_{d+1} \times \cdots \times p_{d+m} \times n}$, where $\boldsymbol{\mathcal{Y}}_{[:, \ldots, :, i]} = \boldsymbol{\mathcal{Y}}_i$ and $\boldsymbol{\mathcal{E}}_{[:, \ldots, :, i]} = \boldsymbol{\mathcal{E}}_i$. Then the tensor-on-tensor regression model can be written succinctly as $\boldsymbol{\mathcal{Y}} = \mathscr{X}(\boldsymbol{\mathcal{A}}) + \boldsymbol{\mathcal{E}}$, where $\mathscr{X} : \mathbb{R}^{p_1 \times \cdots \times p_{d+m}} \to \mathbb{R}^{p_{d+1} \times \cdots \times p_{d+m} \times n}$ is a linear map such that

$$\mathscr{X}(\boldsymbol{\mathcal{A}})_{[:, \ldots, :, i]} = \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{A}} \rangle_* \quad \text{for} \quad i = 1, \ldots, n. \tag{2}$$

Under different choices for covariate order d and response order m, the generic tensor-on-tensor regression model covers many special regression models in the literature, such as

- Scalar-on-tensor regression (Zhou et al., 2013; Mu et al., 2014): $m = 0, d \geq 3$;
- Tensor-on-vector regression (Li and Zhang, 2017; Sun and Li, 2017): $d = 1, m \geq 3$;
- Scalar-on-matrix regression (or matrix trace regression) (Recht et al., 2010): $m = 0, d = 2$;
- Reduced-rank regression (Izenman, 1975; Reinsel and Velu, 2013): $m = 1, d = 1$.

There has been a surge of interest in estimating the model parameter $\boldsymbol{\mathcal{A}}$ under the generative model (1) (Lock, 2018; Raskutti et al., 2019; Llosa and Maitra, 2022; Luo and Zhang, 2022b; Gahrooei et al., 2021; Liu et al., 2021). A natural question prior to estimating $\boldsymbol{\mathcal{A}}$ is whether the signal $\boldsymbol{\mathcal{A}}$ is significant enough to detect, i.e., detecting whether $\boldsymbol{\mathcal{A}}$ is zero or not. When $m = 0$ and $d = 1$ or 2, the statistical limits of the corresponding detection problems have been studied in Ingster et al. (2010); Verzelen (2012); Arias-Castro et al. (2011) and Carpentier and Nickl (2015), respectively. However, to our best knowledge, the tensor-on-tensor association detection problem with generic choices of m and d is still largely unexplored in the literature.

In this paper, we aim to make a first attempt at this problem. Specifically, we focus on testing whether the observed response tensors and covariate tensors are associated by a rank-one tensor: i.e., given $\{\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i\}_{i=1}^n$, or equivalently $(\boldsymbol{\mathcal{Y}}, \mathscr{X})$, generated from model (1), consider the hypothesis testing problem

$$H_0 : \boldsymbol{\mathcal{A}} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mathcal{A}} \in \mathscr{A}(\lambda), \tag{3}$$

where for $\lambda > 0$,

$$\mathscr{A}(\lambda) = \{\lambda' \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_{d+m} | \lambda' \geq \lambda, \mathbf{u}_j \in \mathcal{S}_{p_j - 1} \text{ for } j = 1, \ldots, d+m\}.$$

Here $\mathcal{S}_{p-1}$ denotes the set of all unit vectors in $\mathbb{R}^p$.

Following the existing literature on signal detection in linear regression and low-rank matrix trace regression (Ingster et al., 2010; Verzelen, 2012; Arias-Castro et al., 2011; Carpentier and Nickl, 2015), we assume the design and the noise are independent Gaussian, i.e., $\boldsymbol{\mathcal{X}}_i$ has i.i.d. $N(0, 1)$

entries and $\boldsymbol{\mathcal{E}}_i$ has i.i.d. $N(0, \sigma^2)$ entries. As usual, our results can be extended to the sub-Gaussian design and noise setting. Moreover, we assume σ is known.

Given any testing procedure $\phi : (\boldsymbol{\mathcal{Y}}, \mathscr{X}) \to \{0, 1\}$, we define its risk as

$$R(\phi) = \mathbb{P}_0(\phi(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1) + \sup_{\boldsymbol{\mathcal{A}} \in \mathscr{A}(\lambda)} \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0),$$

where $\mathbb{P}_0$ is the probability under H_0 and $\mathbb{P}_{\boldsymbol{\mathcal{A}}}$ is the probability under H_1 with the signal tensor $\boldsymbol{\mathcal{A}}$. Given $\lambda > 0$, we say ϕ reliably detects in (3) if for any error tolerance $\alpha > 0$, $R(\phi) \leq \alpha$ for sufficiently large n and p_js.

In this work, we aim to study the statistical and computational limits of λ such that reliable detection for (3) can be achieved by an unconstrained-time algorithm or a polynomial-time algorithm. In particular, in studying the computational limits, we consider the class of Statistical Query (SQ) algorithms (Kearns, 1998). Denote AllAlg^D the class of unconstrained-time algorithms that includes all testing procedures with unlimited computational resources and $\mathrm{PolySQAlg}^D$ the class of SQ testing algorithms that must finish in poly(q) time, where q is the size of the input. As $\min\{n, p_1, \ldots, p_{d+m}\} \to \infty$, we call λ^s the statistical separation rate of the testing problem (3) if

$$\inf_{\phi \in \mathrm{AllAlg}^D} R(\phi) \to 0 \quad \text{when} \quad \lambda/\lambda^s \to \infty \quad \text{and} \quad \inf_{\phi \in \mathrm{AllAlg}^D} R(\phi) \to 1 \quad \text{when} \quad \lambda/\lambda^s \to 0.$$

Similarly, we call λ^c the computational separation rate of the testing problem (3) if

$$\inf_{\phi \in \mathrm{PolySQAlg}^D} R(\phi) \to 0 \quad \text{when} \quad \lambda/\lambda^c \to \infty \quad \text{and} \quad \inf_{\phi \in \mathrm{PolySQAlg}^D} R(\phi) \to 1 \quad \text{when} \quad \lambda/\lambda^c \to 0.$$

1.1. Summary of Contributions

In this paper, we identify the statistical and computational separation rates for the generic tensor-on-tensor association detection problem (3) with every pair of response order m and covariate order d. See Table 1 for a summary of our results. To the best of our knowledge, this is the first complete characterization of the statistical and computational limits for the general tensor-on-tensor association detection problem. On the statistical side, we extend the statistical lower bounds in Ingster et al. (2010); Carpentier and Nickl (2015) for $m = 0$, $d = 1$, and $d = 2$ cases to the general setting. In the reduced-rank regression setting, i.e., $m = 1, d = 1$, to our knowledge, this is also the first result on the corresponding detection problem despite its estimation problem has been widely studied (Izenman, 1975; Reinsel and Velu, 2013). In addition, we find that the statistical lower bounds for $m = 0$, $m = 1$, and $m \geq 2$ cases are all different and they require different proof techniques. We also develop optimal tests which can achieve the corresponding statistical limits in all different scenarios. These optimal statistical tests require computing the largest singular value of a given tensor, which is in general computationally intractable (Hillar and Lim, 2013).

Next, we study the computational separation rates for tensor-on-tensor association detection. We first develop efficient algorithms with optimal guarantees. These tests are based on computing the moments and U-statistics of the parameter of interest. We also develop matching computational lower bounds under the Statistical Query (SQ) framework (Kearns, 1998) to show that the SNRs required by the proposed efficient algorithms are essential. When m and d are even orders, we also provide the matching SQ upper bounds.

Table 1 shows there is a significant gap between the statistical and computational separation rates if and only if $d + m \geq 3$, i.e., in the setting $\boldsymbol{\mathcal{X}}_i$s and $\boldsymbol{\mathcal{Y}}_i$s are associated by a tensor parameter

	λ^s/σ	λ^c/σ	Stat-Comp Gap
$m=0, d=1$ (Ingster et al., 2010)	$\Theta\left(\frac{p^{1/4}}{\sqrt{n}} \wedge n^{-1/4}\right)$	$\Theta\left(\frac{p^{1/4}}{\sqrt{n}} \wedge n^{-1/4}\right)$	No
$m=0, d=2$ (Carpentier and Nickl, 2015)	$\Theta\left(\sqrt{\frac{p}{n}} \wedge n^{-1/4}\right)$	$\Theta\left(\sqrt{\frac{p}{n}} \wedge n^{-1/4}\right)$	No
$m=0, d\geq 3$	$\Theta\left(\sqrt{\frac{p}{n}} \wedge n^{-1/4}\right)$	$\Theta\left(\sqrt{\frac{p^{d/2}}{n}} \wedge n^{-1/4}\right)$	Yes
$m=1, d=0$	$\Theta\left(\frac{p^{1/4}}{\sqrt{n}}\right)$	$\Theta\left(\frac{p^{1/4}}{\sqrt{n}}\right)$	No
$m=1, d=1$	$\Theta\left((\frac{p}{n})^{1/2} \wedge (\frac{p}{n})^{1/4}\right)$	$\Theta\left((\frac{p}{n})^{1/2} \wedge (\frac{p}{n})^{1/4}\right)$	No
$m=1, d\geq 2$	$\Theta\left((\frac{p}{n})^{1/2} \wedge (\frac{p}{n})^{1/4}\right)$	$\Theta\left(\sqrt{\frac{p^{(d+1)/2}}{n}} \wedge (\frac{p}{n})^{1/4}\right)$	Yes
$m=2, d=0$	$\Theta\left(\sqrt{\frac{p}{n}}\right)$	$\Theta\left(\sqrt{\frac{p}{n}}\right)$	No
$m=2, d\geq 1$ or $m\geq 3$	$\Theta\left(\sqrt{\frac{p}{n}}\right)$	$\Theta\left(\sqrt{\frac{p^{(d+m)/2}}{n}} \wedge (\frac{p^m}{n})^{1/4}\right)$	Yes

Table 1: Summary of statistical and computational separation rates for the tensor-on-tensor association detection (3). For simplicity, we consider the setting $p_1 = \cdots = p_{d+m} = p$ and $n, p \to \infty$.

of order three or higher. This echo with a series of results in the literature that a statistical and computational gap shows up in the high-order tensor problems (Richard and Montanari, 2014; Zhang and Xia, 2018; Brennan and Bresler, 2020; Luo and Zhang, 2022a; Dudeja and Hsu, 2021; Han et al., 2022a). But different from these existing works, we consider a supervised problem and both the response and covariate tensors' orders are relevant to the statistical-computational gap. It is also worth noting that although the statistical separation rates for $m = 0$, $m = 1$, and $m \geq 2$ cases are different, the computational separation rates under these scenarios share the same expression $\Theta\left(\sqrt{p^{(d+m)/2}/n} \wedge (p^m/n)^{1/4}\right)$ for every (m, d) pair.

Finally, on the technical side, we find truncation is critical in proving both sharp statistical and computational lower bounds. In particular, we develop a new truncation strategy in applying the second-moment method to show the sharp statistical lower bounds in the $m = 1, d \geq 1$ case. The new strategy involves truncating away a rare "bad" event depending on both covariates and the prior on the parameter tensor. For the computational limits, we developed the first SQ lower bound for truncated distribution. Specifically, we demonstrated that "large" χ^2 correlations are rare events and are dominated on average by the "small" ones in computing the statistical dimension.

1.2. Additional Related Work

The tensor-on-tensor association detection problem is also related to several other detection problems studied in the literature. For example, when $m = 1$ and $d = 0$, our problem is related to the classic signal detection in the Gaussian sequence model (Ermakov, 1991; Ingster et al., 2003; Baraud, 2002). The difference is that here we also have a scalar covariate in the model. Similarly, when $m \geq 3, d = 0$, our problem is closely related to the widely studied tensor PCA detection problem (Montanari et al., 2017; Perry et al., 2020; Jagannath et al., 2020; Kunisky et al., 2022; Dudeja and Hsu, 2021; Brennan and Bresler, 2020; Brennan et al., 2021). Comparing to these results in the literature, we find that in both $m = 1, d = 0$ and $m \geq 3, d = 0$ cases, the statistical and computational separation rates do not change even after introducing an extra random scalar covariate.

Statistical Query is a common framework for providing rigorous evidence for the computational barriers in high-dimensional statistical problems (Feldman et al., 2017b; Diakonikolas et al., 2017, 2019; Feldman et al., 2018; Fan et al., 2018; Kannan and Vempala, 2017). The SQ model was introduced by Kearns (1998) in the context of supervised learning as a natural restriction of the PAC model and has been extensively studied in learning theory. A recent line of work Feldman et al. (2017a); Feldman (2017b); Feldman et al. (2017b) generalized the SQ framework for search problems over distributions. The SQ lower bounds have also been established for several tensor-related problems, such as tensor PCA (Dudeja and Hsu, 2021) and multi-sample hypergraphic planted clique (Brennan et al., 2021).

2. Notation and Preliminaries

Let $[r] = \{1, \ldots, r\}$ for any positive integer r. Lowercase letters (e.g., a), lowercase boldface letters (e.g., $\mathbf{u}$), uppercase boldface letters (e.g., $\mathbf{U}$), and boldface calligraphic letters (e.g., $\boldsymbol{\mathcal{A}}$) denote scalars, vectors, matrices, and order-3-or-higher tensors, respectively. We use bracket subscripts to denote sub-vectors, sub-matrices, and sub-tensors. For an order-d tensor, the Frobenius norm of tensor $\boldsymbol{\mathcal{A}}$ is defined as $\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} = \left(\sum_{i_1,\ldots,i_d} \boldsymbol{\mathcal{A}}^2_{[i_1,\ldots,i_d]}\right)^{1/2}$. The mode-$k$ product of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{p_1\times\cdots\times p_d}$ with a matrix $\mathbf{B} \in \mathbb{R}^{r_k\times p_k}$, denoted by $\boldsymbol{\mathcal{A}}\times_k\mathbf{B}$, is a $p_1\times\cdots\times p_{k-1}\times r_k\times p_{k+1}\times\cdots\times p_d$-dimensional tensor, and its definition is given as $(\boldsymbol{\mathcal{A}}\times_k \mathbf{B})_{[i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_d]} = \sum_{i_k=1}^{p_k} \boldsymbol{\mathcal{A}}_{[i_1,i_2,\ldots,i_d]}\mathbf{B}_{[j,i_k]}$. In addition, we let $\boldsymbol{\mathcal{A}} \times_{k=1}^{d} \mathbf{U}_k := \boldsymbol{\mathcal{A}} \times_1 \mathbf{U}_1 \times \cdots \times_d \mathbf{U}_d$.

For any two sequences $\{a_n\}$, $\{b_n\}$, we say $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$; we say $a_n = \Theta(b_n)$ if $\lim_{n\to\infty}\log(a_n)/\log(n) = \log(b_n)/\log(n)$; we say $a_n \gtrsim b_n$ if $a_n \geq Cb_n$ for all n with some large constant C, this C can depend on m, d or some other constants but it do not depend on n and p. Given any real numbers a, b, denote $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Throughout the paper, let c, C be some absolute constants and C_d, c_d be constants that depend on d only, whose actual values vary from line to line; $c_m, c_{d,m}, c_{d,m,\alpha}$ are noted similarly.

The pairwise correlation of two distributions with probability density functions $D_1, D_2 : \mathbb{R}^q \to \mathbb{R}_+ = \{u \in \mathbb{R} : u \geq 0\}$ with respect to a distribution with density $D : \mathbb{R}^q \to \mathbb{R}_+$, where the support of D contains the supports of D_1 and D_2, is defined as $\chi_D(D_1, D_2) := \int_{\mathbb{R}^m} D_1(x)D_2(x)/D(x)dx - 1$. We remark that when $D_1 = D_2$, the pairwise correlation is identified with the χ^2-divergence between D_1 and D, i.e., $\chi^2(D_1, D) := \int_{\mathbb{R}^m} D_1(x)^2/D(x)dx - 1$.

3. Statistical Lower Bounds

In this section, we provide statistical lower bounds of the tensor-on-tensor association detection problem (3) separately for $m = 0$, $m = 1$, and $m \geq 2$ cases as they are all different and require distinct proof techniques. Since the results for $m = 0$, $d = 1$, or 2 have been established in Ingster et al. (2010); Carpentier and Nickl (2015), we omit the proof here for simplicity. Throughout this section, we let $\underline{p} = \min_j p_j$.

Theorem 1 (Statistical Lower Bound ($m = 0, d \geq 2$)) *If $\lambda/\sigma = o((\underline{p}/n)^{1/2} \wedge n^{-1/4})$, we have $\lim_{n,\underline{p}\to\infty} \inf_\phi R(\phi) = 1$.*

Theorem 2 (Statistical Lower Bound ($m = 1$))

- *($d = 0$) If $\lambda = o(\underline{p}^{1/4}/\sqrt{n})$, we have $\lim_{n,\underline{p}\to\infty} \inf_\phi R(\phi) = 1$.*
- *($d \geq 1$) If $\lambda = o((\underline{p}/n)^{1/2} \wedge (\underline{p}/n)^{1/4})$, we have $\lim_{n,\underline{p}\to\infty} \inf_\phi R(\phi) = 1$.*

Theorem 3 (Statistical Lower Bound ($m \geq 2$)) *If $\lambda/\sigma = o((\underline{p}/n)^{1/2})$, then $\lim_{n,\underline{p}\to\infty} \inf_\phi R(\phi) = 1$.*

Remark 4 (Truncated Second Moment Method) *We prove the statistical lower bounds by reducing the minimax testing risk to a Bayesian testing risk with a uniform prior over the set of parameters. Typically, to show the lower bound, one studies the second moment of the likelihood ratio under the null and proves it tends to 1 when the SNR is below the threshold. However, the second moment of the likelihood ratio can be dominated by an extremely rare "bad" event, causing it to be unbounded. In our problem, such bad events exist when proving the lower bounds in the $m \geq 1$ and large λ settings while do not exist in the $m = 0$ setting considered in Ingster et al. (2010); Carpentier and Nickl (2015). To tackle this challenge, we apply the truncated second-moment technique (Butucea and Ingster, 2013; Arias-Castro and Verzelen, 2014; Perry et al., 2020) to peel away a rare "bad" event in computing the second moment of the likelihood ratio and show the truncated version tends to 1. When $m = 1, d = 0$ or $m \geq 2$, we are able to prove sharp statistical lower bounds by peeling away a rare bad event depending on the covariates tensor only. While in the $m = 1, d \geq 1$ setting, the existing truncation strategy fails; to obtain a sharp lower bound, we perform truncation on both the covariates and the prior on the tensor parameter. To our knowledge, such a truncation strategy on both covariates and the prior on parameters is new. We provide a proof sketch for Theorem 2 in Section 7 to illustrate this new truncation strategy.*

4. Statistical and Computational Upper Bounds

In this section, we introduce the testing procedures that achieve the statistical and computational upper bounds. We first introduce two search statistics that will be critical in developing the statistical optimal tests. Let

$$T_1 = \sup_{\mathbf{v}_j \in \mathcal{S}_{p_j - 1}, j=1,\dots,d+m} \mathscr{X}^*(\boldsymbol{\mathcal{Y}}) \times_1 \mathbf{v}_1^\top \times \cdots \times_{d+m} \mathbf{v}_{d+m}^\top;$$
$$T_2 = \sup_{\mathbf{v}_j \in \mathcal{S}_{p_{d+j}-1}, j=1,\dots,m, \mathbf{v}_{m+1} \in \mathcal{S}_{n-1}} \boldsymbol{\mathcal{Y}} \times_1 \mathbf{v}_1^\top \times \cdots \times_m \mathbf{v}_m^\top \times_{m+1} \mathbf{v}_{m+1}^\top.$$

Here $\mathscr{X}^* : \mathbb{R}^{p_{d+1}\times\cdots\times p_{d+m}\times n} \to \mathbb{R}^{p_1\times\cdots\times p_{d+m}}$ denotes the adjoint operator of $\mathscr{X}$, i.e., $\mathscr{X}^*(\boldsymbol{\mathcal{Y}}) := \sum_{i=1}^n \boldsymbol{\mathcal{Y}}_{[:,\dots,:,i]} \otimes \boldsymbol{\mathcal{X}}_i$.

Based on T_1 and T_2, we introduce the following two testing procedures. Given any pre-specified error tolerance $\alpha > 0$, define

- $\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\left(T_1 \geq Z_{1\alpha}\sqrt{n(\sum_{j=1}^{d+m} p_j)}\right)$;

- $\phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\left(T_2 \geq Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right)$,

where $Z_{1\alpha}, Z_{2\alpha}$ are some sufficiently large constants that depend only α. Next, we provide guarantees for these testing procedures separately.

Theorem 5 (Guarantee of ϕ_1) *Suppose $n \gtrsim (\sum_{j=1}^{d+m} p_j)$. Then for sufficiently large $Z_{1\alpha}$, if $\lambda/\sigma > 4Z_{1\alpha}\sqrt{(\sum_{j=1}^{d+m} p_j)/n}$, we have $R(\phi_1) \leq \alpha$.*

Theorem 6 (Guarantee of ϕ_2) *Suppose $n \gtrsim 1$. Then for sufficiently large $Z_{2\alpha}$, if $\lambda/\sigma > 2\sqrt{2}Z_{2\alpha}\sqrt{1 + \sum_{j=d+1}^{d+m} \frac{p_j}{n}}$, we have $R(\phi_2) \leq \alpha$.*

We note that the test ϕ_1 utilizes the information on $\boldsymbol{\mathcal{Y}}$ only while test ϕ_2 utilizes the information on the interaction between $\boldsymbol{\mathcal{Y}}_i$s and $\boldsymbol{\mathcal{X}}_i$s. We will see later that each of them individually is suboptimal; but the combination of them, $\phi_1 \vee \phi_2$, achieves the statistical upper bound.

On the other hand, when $d+m \geq 3$ (or $m+1 \geq 3$), the statistic T_1 (or T_2) relies on computing the leading singular value of the tensor $\mathscr{X}^*(\boldsymbol{\mathcal{Y}})$ (or $\boldsymbol{\mathcal{Y}}$), for which is NP-hard in general (Hillar and Lim, 2013). Thus, computing ϕ_1 or ϕ_2 can be computationally infeasible. This motivates us to develop the following two computationally efficient tests analogous to ϕ_1 and ϕ_2:

- $\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1(T_3 \geq Z_{3\alpha}/\sqrt{n\prod_{j=d+1}^{d+m} p_j})$, where $T_3 = \sum_{i=1}^{n} \|\boldsymbol{\mathcal{Y}}_i\|_{\mathbf{F}}^2/(n\prod_{j=d+1}^{d+m} p_j) - 1$;
- $\phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\left(T_4 \geq Z_{4\alpha}(\prod_{j=1}^{d+m} p_j)^{1/2}/n\right)$, where $T_4 = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\langle \boldsymbol{\mathcal{Y}}_i \otimes \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_j \otimes \boldsymbol{\mathcal{X}}_j\rangle$.

Here the Kronecker product "$\otimes$" between two tensors satisfies $(\boldsymbol{\mathcal{Y}}_i \otimes \boldsymbol{\mathcal{X}}_i)_{[z_1,\cdots,z_{d+m}]} = \boldsymbol{\mathcal{Y}}_{i[z_{d+1},\cdots,z_{d+m}]}\boldsymbol{\mathcal{X}}_{i[z_1,\cdots,z_d]}$.

We note that T_3 can be viewed as a moment estimator for λ^2 and T_4 is a H_0-centered U-statistic for estimating the tensor parameter $\boldsymbol{\mathcal{A}}$. Next, we provide guarantees for testing procedures ϕ_3 and ϕ_4.

Theorem 7 (Guarantee of ϕ_3) *Suppose $n \gtrsim 1$. Then for sufficiently large $Z_{3\alpha}$, if $\lambda/\sigma > 8Z_{3\alpha}((\prod_{j=d+1}^{d+m} p_j)/n)^{1/4}$, we have $R(\phi_3) \leq \alpha$.*

Theorem 8 (Guarantee of ϕ_4) *Suppose $n \gtrsim (\prod_{j=1}^{d} p_j)^{1/2}$. Then for sufficiently large $Z_{4\alpha}$, if $\lambda/\sigma > \sqrt{2Z_{4\alpha}(\prod_{j=1}^{d+m} p_j)^{1/2}/n}$, we have $R(\phi_4) \leq \alpha$.*

The testing procedure $\phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X})$ has appeared in the literature for solving the detection problems in the linear regression and low-rank matrix trace regression (Ingster et al., 2010; Carpentier and Nickl, 2015), but analyzing $\phi_4(\boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{X}})$ in tensor-on-tensor association detection is much more involved as the response $\boldsymbol{\mathcal{Y}}_i$ here is a tensor with highly correlated entries. In addition, it is crucial to emphasize the significance of leveraging the tensor structure present in $\boldsymbol{\mathcal{Y}}_i$. Tests that treat each entry separately often result in reduced power for hypothesis testing. Specifically, tests ϕ_1, ϕ_2, ϕ_4 effectively utilize the tensor structure of $\boldsymbol{\mathcal{Y}}_i$ and remain invariant to shuffling the entries within $\boldsymbol{\mathcal{Y}}_i$. On the other hand, test ϕ_3 maintains validity when considering each entry of $\boldsymbol{\mathcal{Y}}_i$ independently, but it alone may provide suboptimal results.

Next, we show that combinations of tests ϕ_1-ϕ_4 can achieve the statistical and computational upper bounds listed in Table 1. In the following, we assume $n \gtrsim 1$.

Corollary 9 (Optimal Unconstrained-time and Polynomial-time Tests for $m = 0, d \geq 3$) *(1) If $\lambda/\sigma \gtrsim ((\sum_{j=1}^{d+1} p_j/n)^{1/2} \wedge n^{-1/4})$, the unconstrained-time test $\phi_1 \vee \phi_2 \vee \phi_3$ satisfies $R(\phi_1 \vee \phi_2 \vee \phi_3) \leq \alpha$; (2) if $\lambda/\sigma \gtrsim \left(n^{-1/4} \wedge ((\prod_{j=1}^{d} p_j)^{1/2}/n)^{1/2}\right)$, the polynomial-time test $\phi_3 \vee \phi_4$ satisfies $R(\phi_3 \vee \phi_4) \leq \alpha$.*

Corollary 10 (Optimal Test for $m = 1, d = 0$) *If $\lambda/\sigma \gtrsim \sqrt{p_1^{1/2}/n}$, $R(\phi_4) \leq \alpha$.*

Corollary 11 (Optimal Test for $m = 1, d = 1$) *If $\lambda/\sigma \gtrsim (((p_1 p_2)^{1/2}/n)^{1/2} \wedge (p_2/n)^{1/4})$, $R(\phi_3 \vee \phi_4) \leq \alpha$.*

Corollary 12 (Optimal Unconstrained-time and Polynomial-time Tests for $m = 1, d \geq 2$) *(1) If $\lambda/\sigma \gtrsim ((\sum_{j=1}^{d+1} p_j/n)^{1/2} \wedge (p_{d+1}/n)^{1/4})$, we have the unconstrained-time test $\phi_1 \vee \phi_2 \vee \phi_3$ satisfies $R(\phi_1 \vee \phi_2 \vee \phi_3) \leq \alpha$; (2) if $\lambda/\sigma \gtrsim \left((p_{d+1}/n)^{1/4} \wedge \sqrt{(\prod_{j=1}^{d+1} p_j)^{1/2}/n}\right)$, we have the polynomial-time test $\phi_3 \vee \phi_4$ satisfies $R(\phi_3 \vee \phi_4) \leq \alpha$.*

Corollary 13 (Optimal Test for $m = 2, d = 0$) *If $\lambda/\sigma \gtrsim \sqrt{(p_1 p_2)^{1/2}/n}$, we have the polynomial-time test ϕ_4 satisfies $R(\phi_4) \leq \alpha$.*

Corollary 14 (Optimal Unconstrained-time and Polynomial-time Tests for $m \geq 2, d \geq 1$ or $m \geq 3$) *(1) If $\lambda/\sigma \gtrsim \sqrt{(\sum_{j=1}^{d+m} p_j)/n}$, the unconstrained-time test $\phi_1 \vee \phi_2$ satisfies $R(\phi_1 \vee \phi_2) \leq \alpha$; (2) if $\lambda/\sigma \gtrsim \left(((\prod_{j=d+1}^{d+m} p_j)/n)^{1/4} \wedge \sqrt{(\prod_{j=1}^{d+m} p_j)^{1/2}/n}\right)$, the polynomial-time test $\phi_3 \vee \phi_4$ satisfies $R(\phi_3 \vee \phi_4) \leq \alpha$.*

For readers' convenience, we provide a summary of statistically and computationally optimal testing procedures for different (m, d) pairs in Table 2. From Corollaries 9, 12 and 14, we can see that the efficient test procedures we develop here require a strictly stronger SNR than the statistical optimal tests to solve the tensor-on-tensor association detection problem if and only if $d + m \geq 3$.

5. Statistical Query Lower and Upper Bounds

5.1. Statistical Query Lower Bound

In this section, we demonstrate that the stronger SNRs required by the proposed polynomial-time tests when $d + m \geq 3$ are also required by a fairly broad class of Statistical Query (SQ) algorithms. We start by providing some preliminaries.

Definition 15 (Decision Problem over Distributions) *We denote by $\mathcal{B}(\mathcal{D}, D)$ the decision (or hypothesis testing) problem in which the input distribution D' is promised to satisfy either (a) $D' = D$ or (b) $D' \in \mathcal{D}$, and the goal of the algorithm is to distinguish between these two cases.*

We define SQ algorithms as algorithms that do not have direct access to samples from the distribution but instead have access to an SQ oracle. We consider the following standard oracle.

	Statistically Optimal Test	Computationally Optimal Test
$m = 0, d \geq 3$	$\phi_1 \vee \phi_2 \vee \phi_3$	$\phi_3 \vee \phi_4$
$m = 1, d = 0$	ϕ_4	ϕ_4
$m = 1, d = 1$	$\phi_3 \vee \phi_4$	$\phi_3 \vee \phi_4$
$m = 1, d \geq 2$	$\phi_1 \vee \phi_2 \vee \phi_3$	$\phi_3 \vee \phi_4$
$m = 2, d = 0$	ϕ_4	ϕ_4
$m = 2, d \geq 1$ or $m \geq 3$	$\phi_1 \vee \phi_2$	$\phi_3 \vee \phi_4$

Table 2: Summary of optimal testing procedures for different (m, d) pairs. In the Statistically Optimal Test column, we provide the optimal unconstrained-time testing procedures that achieve the statistical upper bounds and in the Computationally Optimal Test column, we provide the optimal polynomial-time testing procedures (supported by matching SQ lower bounds given in Section 5) that achieve the computational upper bounds.

Definition 16 (VSTAT **Oracle**) *Let D be a distribution over a domain X. For a sample size parameter $n > 0$ and any bounded function $f : \mathbb{R}^q \to [0, 1]$, $\mathrm{VSTAT}(n)$ returns a value $v \in [\mathbb{E}_{x\sim D}[f(x)] - \tau, \mathbb{E}_{x\sim D}[f(x)] + \tau]$, where $\tau = \max\{\frac{1}{n}, \sqrt{\frac{\mathbb{E}_{x\sim D}[f(x)](1-\mathbb{E}_{x\sim D}[f(x)]}{n}}\}$.*

One can prove lower bounds on the complexity of SQ algorithms via an appropriate notion of *Statistical dimension*. Such a complexity measure was introduced in Blum et al. (1994) for PAC learning of Boolean functions and has been generalized to the unsupervised setting in Feldman et al. (2017a); Feldman (2017b). For technical reasons to be mentioned in Remark 22, here we use the definition of statistical dimension from Brennan et al. (2021).

Definition 17 (Statistical Dimension) *Suppose the distributions in $\mathcal{D}$ are indexed by $\mathbf{u} \in \mathcal{U}$. Let μ be a uniform distribution on $\mathcal{U}$. The statistical dimension $\mathrm{SDA}(n)$ for the $\mathcal{B}(\mathcal{D}, D)$ is defined as follows:*

$$\mathrm{SDA}(n) = \max\{q \in \mathbb{N} : \mathbb{E}_{\mathbf{u},\mathbf{v}\sim\mu}\left[|\chi_{D_0}(D_\mathbf{u}, D_\mathbf{v}) - 1| \,\middle|\, \mathcal{E}\right] \leq \frac{1}{n} \text{ for all events } \mathcal{E} \text{ s.t. } \mathbb{P}_{\mathbf{u},\mathbf{v}\sim\mu}[\mathcal{E}] \geq \frac{1}{q^2}\}.$$

If one can bound below the SDA of the given problem, then it implies an unconditional lower bound on the complexity of any SQ algorithm for the problem using the following standard result.

Lemma 18 (Theroem 1.3 in Brennan et al. (2021) and Theorem 2.7 in Feldman et al. (2017a)) *Let $\mathcal{B}(\mathcal{D}, D)$ be a decision problem, where D is the reference distribution and $\mathcal{D}$ is a class of distributions. Any SQ algorithm for solving $\mathcal{B}$ requires at least $\mathrm{SDA}(n)$ queries to the $\mathrm{VSTAT}(1/3n)$ oracle.*

Next, we provide a lower bound on the SQ dimension for the tensor-on-tensor association detection problem.

Theorem 19 (Statistical Query Lower Bound) *Given any $0 < \epsilon < 1/4$ and sufficiently large p. Suppose $\lambda/\sigma \leq 1/2$ for $m = 0$, $\lambda/\sigma \leq c_1 \underline{p}^{1/4-\epsilon}$ for $m = 1$ and $\lambda/\sigma \leq c_1 \underline{p}^{1/2-2\epsilon}$ for $m \geq 2$ with* *sufficiently small constant $c_1 > 0$. Then if $n \le c_2((\sigma/\lambda)^2 \underline{p}^{(1/2-2\epsilon)(d+m)} + (\sigma/\lambda)^4 \underline{p}^{(1-4\epsilon)m})$ for some sufficiently small constant $c_2 > 0$, we have the statistical query dimension of tensor-on-tensor association detection problem is at least $\exp(cp^{2\epsilon})$ for some $c > 0$.*

By Lemma 18 and the interpretation for the VSTAT oracle, Theorem 19 shows that for any $0 < \epsilon < \frac{1}{4}$, if $n \le c_2((\sigma/\lambda)^2 \underline{p}^{(1/2-2\epsilon)(d+m)} + (\sigma/\lambda)^4 \underline{p}^{(1-4\epsilon)m})$, then we need at least super polynomial many queries to solve the tensor-on-tensor association detection problem via the SQ algorithm. Since the number of queries is a proxy for runtime, this is equivalent to say efficient SQ algorithms require SNR at least $\lambda/\sigma \ge C\left(\sqrt{\underline{p}^{(d+m)/2}/n} \wedge (\underline{p}^m/n)^{1/4}\right)$. Finally, our SQ lower bound also implies a computational lower bound for the detection problem restricted to the class of low-degree polynomial algorithms based on the recent work Brennan et al. (2021). For completeness, we provide details in Appendix A.

Remark 20 (Optimality of the SQ Lower Bound) *Our SQ lower bound is proved under the SNR condition $\lambda/\sigma \le 1/2$ for $m = 0$, $\lambda/\sigma \le c_1\underline{p}^{1/4-\epsilon}$ for $m = 1$ and $\lambda/\sigma \le c_1\underline{p}^{1/2-2\epsilon}$ for $m \ge 2$. This is the best we can hope for as when $\lambda \gg 1$, $\lambda \gg p^{1/4}$ and $\lambda \gg p^{1/2}$ for $m = 0$, $m = 1$ and $m = 2$, respectively, we can query the statistical optimal test given in Table 2 via VSTAT(4), then SQ can solve the detection problem with only $O(1)$ samples as suggested in statistical separation rates given in Table 1, while these test can not be computed efficiently in general as we have mentioned in Section 5.2. Such a restriction (also termed "one-shot problem" versus "multi-sample problem" issue in Brennan et al. (2021)) has also appeared when we try to prove SQ lower bounds for planted clique and tensor PCA problems (Brennan et al., 2021) for the same reason that when the SNR is large, SQ algorithms can query a high-degree function and solve the problem with only one sample, but that high-degree function can not be computed efficiently.*

Remark 21 (Proof Techniques) *To prove the sharp SQ lower bound, we again apply truncation technique. This choice is motivated by the fact that the pairwise correlation between different distributions tends to escalate rapidly for large values of λ. To our best knowledge, this work is the first SQ lower bound specifically designed for truncated distributions. To bound the statistical dimension, we embark on a two-step process. Initially, we identify the set that attains the largest SDA and subsequently illustrate that occurrences of "large" χ^2 correlations are infrequent and are predominantly overshadowed by the prevalence of "smaller" correlations on average. We believe that these techniques can be adapted and applied effectively in other contexts, particularly when instances of pairwise correlation intermittently exhibit substantial amplification.*

Remark 22 (Comparisons on Different Statistical Dimensions) *We also investigated a more straightforward way to show SQ lower bounds via other statistical dimension notions based on the averaged pairwise correlation or pairwise correlation, as suggested in the seminal paper Feldman et al. (2017a). However, we succeeded in proving the sharp SQ lower bound only when $\lambda/\sigma = O(\mathrm{Poly}(\log p))$, which is strictly suboptimal compared to the ones in Theorem 19 when $m \ge 1$. It is interesting to explore whether there is some gap between the notion of statistical dimension in Definition 17 and the ones in Feldman et al. (2017a). For readers' reference, in Appendix B, we provide a simpler proof for the SQ lower bound based on pairwise correlation when $\lambda/\sigma \le 1/2$.*

5.2. Statistical Query Upper Bound

In this section, we assume d and m are even numbers and the class of tensor parameters of interest has the form $\mathscr{A} = \{\lambda \mathbf{u}^{\otimes(d+m)}, \mathbf{u} \in \mathcal{S}_{p-1}\}$. Then we show that we can provide a matching SQ upper bound for the corresponding tensor-on-tensor association detection problem.

In showing the SQ upper bound, we will use the following two statistics:

- $T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) = \sum_{j_1,\ldots,j_m=1}^{p} \boldsymbol{\mathcal{Y}}_{1[j_1,\ldots,j_m]}^2$;
- $T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) = \sum_{j_1,\ldots,j_{\frac{d+m}{2}}=1}^{p} (\boldsymbol{\mathcal{Y}}_1 \otimes \boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,j_2,j_2,\ldots,j_{\frac{d+m}{2}},j_{\frac{d+m}{2}}]}$.

Notice that $T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)$ and $T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)$ are one-sample versions of statistics T_3 and T_4. The target quantity we would like to estimate via the SQ algorithm is $\phi_5 \vee \phi_6$, where

$$\phi_5 = 1(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) \geq p^m + \lambda^2/2) \quad \text{and} \quad \phi_6 = 1(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) \geq \lambda/2).$$

Then we have the following guarantee on the SQ upper bound.

Theorem 23 (Statistical Query Upper Bound) *Suppose*

$$n = C\left(\left(\frac{\sigma^2 p^{(d+m)/2}}{\lambda^2} \wedge \frac{\sigma^4 p^m}{\lambda^4}\right) \vee 1\right) \log^2 n \tag{4}$$

for large $C > 0$.Then there exists a statistical query algorithm that distinguishes H_0 from H_1 by estimating $\phi_5 \vee \phi_6$ with $O(\log(nB/(\lambda^2 \wedge \lambda^4)))$ number of queries to VSTAT(n)*, where $B := \max(2\lambda^2 + p^{d/2}\lambda^2 + p^{(d+m)/2}, 2p^{2m} + 2p^m\lambda^2 + 4\lambda^2 + 3\lambda^4)$.*

Theorem 23 shows that when d and m are even and $\boldsymbol{\mathcal{A}} = \lambda \mathbf{u}^{\otimes(d+m)}$ for some $\mathbf{u} \in \mathcal{S}_{p-1}$, there exists an efficient SQ algorithm for solving the tensor-on-tensor association detection problem (3) if $\lambda/\sigma = \Theta\left(\sqrt{p^{(d+m)/2}/n} \wedge (p^m/n)^{1/4}\right)$.

Remark 24 *We note the current SQ lower and upper bounds are only matched in the special setting considered in this section. When d or m are not even, we conjecture the actual SQ lower and upper bounds might be slightly higher than the one in (4), even though we believe the computational separate rates given in Table 1 are still correct, i.e., SQ algorithms are suboptimal when d or m are not even. Similar issues have also appeared when we try to provide tight SQ lower and upper bounds for tensor PCA, where Dudeja and Hsu (2021) showed that SQ algorithms require an unnecessarily larger SNR when the tensor order is odd and provided a matching SQ lower bound based on the Fourier analytic approach (Feldman et al., 2018; Li et al., 2019).*

5.3. Connection of Detection and Estimation

In this section, we establish a connection between the testing problem (3) and its corresponding estimation problem, which revolves around the estimation of $\boldsymbol{\mathcal{A}}$. We aim to investigate how and under what conditions we can effectively estimate $\boldsymbol{\mathcal{A}}$. Typically, the computational hardness of an estimation problem is studied by examining the computational hardness of its corresponding testing problem. However, in the case of hypothesis testing problem (3), we find that it is actually easier than the corresponding estimation problem. As the noise level σ approaches zero, the hypothesis testing problem (3) becomes trivial, while the estimation problem remains highly nontrivial. This disparity arises due to the varying entrywise variances of $\boldsymbol{\mathcal{Y}}$ under the null and alternative hypotheses, which allows simple tests to succeed.

Given this intriguing observation, it becomes compelling to explore the computational limits when we match the first two moments of the null and alternative distributions. How does this modification impact the computational complexity? In the subsequent discussion, we present an SQ lower bound for a variant of the testing problem (3) where the first two moments of $\boldsymbol{\mathcal{Y}}$ are matched, specifically considering the case when $m = 0$.

Without loss of generality, we assume $\sigma^2 \in [0, 1)$ (see Appendix E.3 for an explanation) and consider the following hypothesis testing problem for $m = 0$:

$$H_0 : (\boldsymbol{\mathcal{X}}_i, y_i)_{i=1}^n \overset{i.i.d.}{\sim} (N(0,1)^{\otimes p^{\otimes d}}, N(0,1))$$
$$H_1 : (\boldsymbol{\mathcal{X}}_i, y_i)_{i=1}^n \overset{i.i.d.}{\sim} y_i = \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{X}}_i \rangle + \varepsilon_i, \text{ with } \boldsymbol{\mathcal{A}} = \mathbf{a}^{\otimes d}, \|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2 = 1 - \sigma^2, \mathbf{a} \in \mathbb{R}^p, \text{ and } \boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} N(0, \sigma^2). \tag{5}$$

Note that for the testing problem in (5), the first and second moments of y_is under null and alternative are matched. Then we have the following SQ lower bound.

Theorem 25 (SQ Hardness for (5)) *For any $0 < \epsilon < \frac{1}{2}$, any SQ algorithm with $\mathrm{VSTAT}(n)$ oracle distinguishing H_0 and H_1 in (5) needs either $2^{\Omega(p^\epsilon)}$ number of queries or requires a query with $n \geq C\left(\frac{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2 + \sigma^2}{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2} p^{d(1/2-\epsilon)}\right)$ for some $C > 0$.*

We also have the following matching SQ upper bound when d is even.

Theorem 26 (SQ Upper Bound for (5)) *Suppose $n\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2 / \left((\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2 + \sigma^2) p^{d/2} \log^2(n)\right) \to \infty$. Then there exists a statistical query algorithm that distinguishes H_0 and H_1 in (5) with vanishing type I+II error by estimating $T_6(y_1, \boldsymbol{\mathcal{X}}_1) := \sum_{j_1,\ldots,j_{d/2} \in [p]} (y_1 \boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,j_2,j_2,\ldots,j_{d/2},j_{d/2}]}$ with $O(\log(nB \log^2 n))$ number of queries to* VSTAT(n)*, where $B = p^{d/2} + 2(1 - \sigma^2)$.*

Notably, the computational lower bound of the sample complexity predicted by the SQ argument for solving (5) matches the existing upper bound of efficient algorithms for the corresponding estimation problem, as demonstrated in (Zhang et al., 2020; Han et al., 2022b; Luo and Zhang, 2021). This alignment between the predicted lower bound and the existing upper bounds provides further support for the validity and accuracy of the SQ argument.

When comparing the sample complexities of the SQ lower bounds for two testing problems, namely (3) with $m = 0$ and (5), we observe a significant increase in the required sample complexity. Specifically, the sample complexity for (3) is on the order of $\Theta(\frac{p^{d/2}\sigma^2}{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2} \wedge \frac{\sigma^4}{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^4})$, while for (5), it becomes $\Theta(\frac{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2 + \sigma^2}{\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}^2} p^{d/2})$.

When $m \geq 1$, the situation becomes more intricate since the variances of each entry in the tensor response differ, making it challenging to determine which null hypothesis should be considered. As a result, we leave this aspect as a topic for future exploration and investigation.

6. Conclusion and Discussions

In this work, we study the tensor-on-tensor association detection problem and provide the first complete characterization of the statistical and computational separation rates for it with all different

(m, d) pairs. We show a gap between the statistical and computational separation rates appears if and only if $d + m \geq 3$. These results significantly extend the results in the literature on signal detection in linear regression and low-rank matrix trace regression.

7. A Proof Sketch for Theorem 2

Preliminaries. To prove the statistical lower bound for a hypothesis testing problem, such as the one in (3), a common strategy is to first reduce it to a Bayesian hypothesis testing problem, then apply the classic Neyman-Pearson Lemma.

Lemma 27 (Neyman-Pearson Lemma, e.g. see Theorem 6.1 in Shao (2006)) *Let μ be a probability measure on $\mathscr{A}$, and let $\{\mathbb{P}_{\boldsymbol{\mathcal{A}}} : \boldsymbol{\mathcal{A}} \in \mathscr{A}\}$ be a family of probability measures indexed by $\boldsymbol{\mathcal{A}} \in \mathscr{A}$ on the domain X. Then, for any probability measure $\mathbb{P}_0$ on X,*

$$\inf_{\phi} \left(\mathbb{P}_0(\phi = 1) + \sup_{\boldsymbol{\mathcal{A}} \in \mathscr{A}} \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi = 0) \right) \geq 1 - \mathrm{TV}(\mathbb{P}_\mu, \mathbb{P}_0),$$

where $\mathbb{P}_\mu$ the mixture probability measure $\mathbb{P}_\mu = \int_{\mathscr{A}} \mathbb{P}_{\boldsymbol{\mathcal{A}}} \mu(d\boldsymbol{\mathcal{A}})$ and $\inf_\phi$ is the infimum over all $\{0, 1\}$-valued statistics.

Oftentimes, working directly with TV distance can be hard and a common strategy is to further bound it by the χ^2-divergence:

$$2\mathrm{TV}(\mathbb{P}_\mu, \mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0}(|D_\mu/D_0 - 1|) \leq \left(\mathbb{E}_{\mathbb{P}_0}\left((D_\mu/D_0 - 1)^2\right)\right)^{1/2} = \sqrt{\chi^2(D_\mu, D_0)}, \tag{6}$$

where D_μ and D_0 denote the density function of $\mathbb{P}_\mu$ and $\mathbb{P}_0$, respectively. Thus, we can see it suffices to choose a suitable measure μ on $\mathscr{A}$ to bound $\chi^2(D_\mu, D_0)$ by a small enough value in order to obtain a desirable lower bound on the minimax risk. Since $\chi^2(D_\mu, D_0) + 1 = \mathbb{E}_{\mathbb{P}_0}\left((D_\mu/D_0)^2\right)$, the second moment of the likelihood ratio under the null, this strategy is also called "second-moment method" in the literature.

For proving our lower bound, we consider the special parameter space $\mathscr{A} = \{\boldsymbol{\mathcal{A}} : \boldsymbol{\mathcal{A}} = \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}, \mathbf{u}_1, \ldots, \mathbf{u}_{d+m} \in \mathcal{S}_{p-1}\}$. Moreover, without loss of generality, we assume the noise tensor has i.i.d. $N(0, 1)$ entries. This is because σ is assumed to be known, we can rescale the data by dividing σ and replace λ by λ/σ in the end.

Naive second-order moment method fails even in $m = 1, d = 0$. In $m = 1, d = 0$, we have $\mathbf{y}_i = \mathbf{a} x_i + \mathbf{e}_i$ for $i \in [n]$, or compactly $\mathbf{Y} = \mathbf{a}\mathbf{x}^\top + \mathbf{E} \in \mathbb{R}^{p \times n}$, where $\mathbf{a} = \lambda \mathbf{u}$ and $\mathbf{x} = (x_1, \ldots, x_n)^\top$. Let $\mathcal{C} := \{\mathbf{u} \in \mathbb{R}^p : \mathbf{u}_i \in \{1/\sqrt{p}, -1/\sqrt{p}\}, i \in [p]\}$, a natural prior, μ, on the parameter space would be $\mathbf{a} = \lambda \mathbf{u}$ where $\mathbf{u}$ is generated uniformly at random from $\mathcal{C}$. Then by Lemma 38 Eq. (47) in Appendix F, we have

$$\begin{aligned} \chi^2(D_\mu, D_0) &= \mathbb{E}_{\mathbf{u},\mathbf{v} \sim \mathrm{Unif}[\mathcal{C}]} \left(\mathbb{E}_{\mathbf{x}}(\exp(\lambda^2 \langle \mathbf{u}, \mathbf{v} \rangle \|\mathbf{x}\|_2^2))\right) \\ &= \mathbb{E}_{\mathbf{u},\mathbf{v} \sim \mathrm{Unif}[\mathcal{C}]} \left(\int \frac{1}{(2\pi)^{n/2}} \exp((\lambda^2 \langle \mathbf{u}, \mathbf{v} \rangle - \frac{1}{2})\|\mathbf{x}\|_2^2) dx_1 \cdots dx_n \right). \end{aligned}$$

Notice that the inner expectation $\mathbb{E}_{\mathbf{x}}(\exp(\lambda^2 \langle \mathbf{u}, \mathbf{v} \rangle \|\mathbf{x}\|_2^2))$ is infinite if $\lambda^2 \langle \mathbf{u}, \mathbf{v} \rangle > 1/2$. To avoid this, we peel away a rare "bad" event when $\|\mathbf{x}\|_2^2$ is large. By the property of TV distance, truncating away a rare event only affects $\mathrm{TV}(\mathbb{P}_\mu, \mathbb{P}_0)$ by $o(1)$. In this example, we let the good event be $A := \{|\|\mathbf{x}\|_2^2 - n| < n/2\}$. By the concentration of χ_n^2, we know the "bad" event A^c happens with probability at most $\exp(-cn)$ for some $c > 0$. Finally, condition on the good event A, we are able to show $\chi^2(D_\mu, D_0) = o(1)$ when $\lambda = o(p^{1/4}/\sqrt{n})$ and the conclusion follows by (6).

Truncation on both covariate and prior is needed when $m = 1, d \geq 1$**.** In this setting, we first show that to prove the lower bound in $m = 1, d \geq 1$, it is enough to show it for the special $m = 1, d = 1$ case, i.e., $\mathbf{y}_i = \lambda\langle \mathbf{x}_i, \mathbf{u}_1\rangle \mathbf{u}_2 + \mathbf{e}_i$ for $i = 1, \ldots, n$ or compactly $\mathbf{Y} = \lambda \mathbf{u}_2 \mathbf{u}_1^\top \mathbf{X}^\top + \mathbf{E} \in \mathbb{R}^{p\times n}$. A natural prior on the parameter space would be $\mathbf{A} = \lambda \mathbf{u}_1 \mathbf{u}_2^\top$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are independently drawn from Unif$[\mathcal{C}]$. Similar to the $m = 1, d = 0$ case, we find a bad event exists here as well. Moreover, since the $\mathbf{x}_i$s and $\mathbf{u}_1$ are entangled together, we need to truncate away a bad event depending jointly on $\mathbf{X}$ and $\mathbf{u}_1$. Specifically, we find a proper good event is the following:

$$A = A_1 \bigcap A_2 = \{\mathbf{X}, \mathbf{u}_1 : \|\mathbf{X}^\top\mathbf{X}\|_{\mathbf{F}}^2 \leq C_1(n\vee p)^2(n\wedge p), \|\mathbf{X}\| \leq C_2\sqrt{n\vee p}, \|\mathbf{X}\mathbf{u}_1\|_2 \leq C_3\sqrt{n}\},$$

where $A_1 = \{\mathbf{X} : \|\mathbf{X}^\top\mathbf{X}\|_{\mathbf{F}}^2 \leq C_1(n\vee p)^2(n\wedge p), \|\mathbf{X}\| \leq C_2\sqrt{n\vee p}\}$, $A_2 := \{\|\mathbf{X}\mathbf{u}_1\|_2 \leq C_3\sqrt{n}\}$ and C_1, C_2, C_3 are some sufficiently large constants. Here the motivation of this good event comes from the latter part analysis when we apply the Hansen-Wright inequality. By the concentration theory for random matrices, we have A happens with probability at least $1 - \exp(-c(n\wedge p))$. By Jensen's inequality and (6), we are able to show

$$\begin{aligned}
&\mathrm{TV}(\mathbb{P}_\mu, \mathbb{P}_0)\\
\leq&\sqrt{\frac{1}{4}\left(\mathbb{E}_{\mathbf{X}}\left[1_{A_1}\left(1 + \int_1^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1}\left(|\mathbf{u}_1^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_1| \geq \Delta\right)dz\right)\right] - \mathbb{P}_{\mathbf{X},\mathbf{u}_1}(A)\right)} + \exp(-c(n\wedge p)),
\end{aligned}$$

where $\Delta := \sqrt{p\log z/(c\lambda^4)}$. Next, we apply the Hansen-Wright inequality

$$\mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(|\mathbf{u}_1^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_1| \geq \Delta\right) \leq \exp\left(-c\left(\frac{\Delta^2 p^2}{\|\mathbf{X}^\top\mathbf{X}\|_{\mathbf{F}}^2} \wedge \frac{\Delta p}{\|\mathbf{X}^\top\mathbf{X}\|}\right)\right)$$

and with this we can show condition on event A_1, when $\lambda = o((p/n)^{1/2} \wedge (p/n)^{1/4})$, we have

$$\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(|\mathbf{u}_1^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_1| \geq \sqrt{\frac{p\log z}{c\lambda^4}}\right)dz = o(1).$$

This further implies $\mathrm{TV}(\mathbb{P}_\mu, \mathbb{P}_0) = o(1)$ as $\mathbb{P}_{\mathbf{X}}(A_1) \to 1, \mathbb{P}_{\mathbf{X},\mathbf{u}_1\sim\mathrm{Unif}[\mathcal{C}]}(A) \to 1$ as $n, p \to \infty$.

Acknowledgments

I. Diakonikolas is supported by NSF Medium Award CCF-2107079, NSF Award CCF-1652862 (CAREER), a Sloan Research Fellowship, and a DARPA Learning with Less Labels (LwLL) grant; D. M. Kane is supported by NSF Award CCF-1553288 (CAREER) and a Sloan Research Fellowship; A.R. Zhang and Y. Luo are in part supported by NSF Award DMS-220374 (CAREER). This work was started while the authors were visiting the Simons Institute and the authors would like to thank the institute for its hospitality and support. Y. Luo would like to thank Chao Gao for helpful discussions during the project.

References

Ery Arias-Castro and Nicolas Verzelen. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940–969, 2014.

Ery Arias-Castro, Emmanuel J Candès, and Yaniv Plan. Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *The Annals of Statistics*, 39(5):2533–2556, 2011.

Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, pages 577–606, 2002.

Avrim Blum, Merrick Furst, Jeffrey Jackson, Michael Kearns, Yishay Mansour, and Steven Rudich. Weakly learning dnf and characterizing statistical query learning using fourier analysis. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 253–262, 1994.

Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*, pages 648–847. PMLR, 2020.

Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.

Cristina Butucea and Yuri I Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

Alexandra Carpentier and Richard Nickl. On signal detection and confidence sets for low rank inference problems. *Electronic Journal of Statistics*, 9(2):2675–2688, 2015.

Alexandra Carpentier, Olivier Collier, Laëtitia Comminges, Alexandre B Tsybakov, and Yu Wang. Minimax rate of testing in sparse linear regression. *Automation and Remote Control*, 80(10): 1817–1834, 2019.

Eungchun Cho. Inner product of random vectors. *International Journal of Pure and Applied Mathematics*, 56(2):217–221, 2009.

Ilias Diakonikolas and Daniel M Kane. The sample complexity of robust covariance testing. In *Conference on Learning Theory*, pages 1511–1521. PMLR, 2021.

Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.

Ilias Diakonikolas, Weihao Kong, and Alistair Stewart. Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2745–2754. SIAM, 2019.

Rishabh Dudeja and Daniel Hsu. Statistical query lower bounds for tensor pca. *Journal of Machine Learning Research*, 22:83–1, 2021.

Michael Sergeevich Ermakov. Minimax detection of a signal in a gaussian white noise. *Theory of Probability & Its Applications*, 35(4):667–679, 1991.

Jianqing Fan, Han Liu, Zhaoran Wang, and Zhuoran Yang. Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. *arXiv preprint arXiv:1808.06996*, 2018.

Vitaly Feldman. Dealing with range anxiety in mean estimation via statistical queries. In *International Conference on Algorithmic Learning Theory*, pages 629–640. PMLR, 2017a.

Vitaly Feldman. A general characterization of the statistical query complexity. In *Conference on Learning Theory*, pages 785–830. PMLR, 2017b.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2): 1–37, 2017a.

Vitaly Feldman, Cristobal Guzman, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1265–1277. SIAM, 2017b.

Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. *SIAM Journal on Computing*, 47(4):1294–1338, 2018.

Mostafa Reisi Gahrooei, Hao Yan, Kamran Paynabar, and Jianjun Shi. Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics*, 63(2):147–159, 2021.

Rungang Han, Yuetian Luo, Miaoyan Wang, Anru R Zhang, et al. Exact clustering in tensor block model: Statistical optimality and computational limit. *Journal of the Royal Statistical Society Series B*, 84(5):1666–1698, 2022a.

Rungang Han, Rebecca Willett, and Anru R Zhang. An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29, 2022b.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

Yuri Ingster, Jurij I Ingster, and IA Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2003.

Yuri I Ingster, Alexandre B Tsybakov, and Nicolas Verzelen. Detection boundary in sparse regression. *Electronic Journal of Statistics*, 4:1476–1526, 2010.

Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, 5(2):248–264, 1975.

Aukosh Jagannath, Patrick Lopatto, and Leo Miolane. Statistical thresholds for tensor pca. *The Annals of Applied Probability*, 30(4):1910–1933, 2020.

Ravindran Kannan and Santosh Vempala. The hidden hubs problem. In *Conference on Learning Theory*, pages 1190–1213, 2017.

Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.

Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2022.

Jerry Li, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. On mean estimation for general norms with statistical queries. In *Conference on Learning Theory*, pages 2158–2172. PMLR, 2019.

Lexin Li and Xin Zhang. Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146, 2017.

Jiani Liu, Ce Zhu, Zhen Long, Yipeng Liu, et al. Tensor regression. *Foundations and Trends® in Machine Learning*, 14(4):379–565, 2021.

Carlos Llosa and Ranjan Maitra. Reduced-rank tensor-on-tensor regression and tensor-variate analysis of variance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Eric F Lock. Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27 (3):638–647, 2018.

Yuetian Luo and Anru R Zhang. Low-rank tensor estimation via riemannian gauss-newton: Statistical optimality and second-order convergence. *arXiv preprint arXiv:2104.12031*, 2021.

Yuetian Luo and Anru R Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *The Annals of Statistics*, 50(1):584–613, 2022a.

Yuetian Luo and Anru R Zhang. Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap, and their interplay. *arXiv preprint arXiv:2206.08756*, 2022b.

Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.

Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank one perturbations of gaussian tensors. *IEEE Transactions on Information Theory*, 63(3):1572–1579, 2017.

Cun Mu, Bo Huang, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *International conference on machine learning*, pages 73–81, 2014.

Amelia Perry, Alexander S Wein, and Afonso S Bandeira. Statistical limits of spiked tensor models. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 230–264. Institut Henri Poincaré, 2020.

Garvesh Raskutti, Ming Yuan, and Han Chen. Convex regularization for high-dimensional multiresponse tensor regression. *The Annals of Statistics*, 47(3):1554–1584, 2019.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

Gregory C Reinsel and Raja P Velu. *Multivariate reduced-rank regression: theory and applications*, volume 136. Springer Science & Business Media, 2013.

Emile Richard and Andrea Montanari. A statistical model for tensor pca. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.

Jun Shao. *Mathematical statistics*. Springer Science & Business Media, 2006.

Will Wei Sun and Lexin Li. Store: sparse tensor response regression and neuroimaging analysis. *The Journal of Machine Learning Research*, 18(1):4908–4944, 2017.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Nicolas Verzelen. Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electronic Journal of Statistics*, 6:38–90, 2012.

Anru Zhang and Dong Xia. Tensor svd: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, 2018.

Anru R Zhang and Yuchen Zhou. On the non-asymptotic and sharp lower tail bounds of random variables. *Stat*, 9(1):e314, 2020.

Anru R Zhang, Yuetian Luo, Garvesh Raskutti, and Ming Yuan. Islet: Fast and optimal low-rank tensor regression via importance sketching. *SIAM journal on mathematics of data science*, 2(2): 444–479, 2020.

Hua Zhou, Lexin Li, and Hongtu Zhu. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552, 2013.

Appendix A. Hardness Against Low-Degree Polynomials

In this section, we leverage the recent result on the connection of SQ model and low-degree polynomials (Brennan et al., 2021) and give the low-degree hardness of the tensor-on-tensor association detection problem in the conjectured hard regime. To this end, we consider a simpler tensor-on-tensor association detection problem with a uniform prior under the alternative.

Problem 28 *We consider the following hypothesis testing problem:*

H_0*:* $(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i) \overset{i.i.d.}{\sim} D_0 := (N(0,1)^{\otimes p^{\otimes d}}, N(0,1)^{\otimes p^{\otimes m}})$ *and* $\boldsymbol{\mathcal{Y}}_i$ *is independent of* $\boldsymbol{\mathcal{X}}_i$*.*

H_1*: First, a vector* $\mathbf{u}$ *is chosen uniformly at random from* $\mathcal{S}_{p-1}$*.* $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{p^{d+m}}$ *has i.i.d.* $N(0,1)$ *entries.* $\boldsymbol{\mathcal{Y}}_i = \langle \boldsymbol{\mathcal{X}}_i, \lambda \mathbf{u}^{\otimes(d+m)} \rangle_* + \boldsymbol{\mathcal{E}}_i$*;* $\boldsymbol{\mathcal{E}}_i$ *has i.i.d.* $N(0, \sigma^2)$ *entries. We denote this distribution by* $D_{\mathbf{u}}$*.*

Next, we introduce a few preliminaries about low-degree polynomials and we refer readers to Brennan et al. (2021) for more details.

Notation. For a distribution D, we denote by $D^{\otimes n}$ the joint distribution of n independent samples from D. For $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ and a distribution D on $\mathbb{R}$, we define the inner product $\langle f, g \rangle_D = \mathbb{E}_{X \sim D}[f(X)g(X)]$ and the norm $||f||_D = \sqrt{\langle f, f \rangle_D}$.

Definition 29 (n-sample ϵ-distinguisher) *We say that the polynomial $p : (\mathbb{R}^{p^d + p^m})^{\otimes n} \to \mathbb{R}$ is an n-sample ϵ-distinguisher for the hypothesis testing Problem 28 if $|\mathbb{E}_{X \sim D_0^{\otimes n}}[p(X)] - \mathbb{E}_{\mathbf{u} \sim \mu} \mathbb{E}_{X \sim D_{\mathbf{u}}^{\otimes n}}[p(X)]| \geq \epsilon \sqrt{\mathbb{V}ar_{X \sim D_0^{\otimes n}}[p(X)]}$. We call ϵ the advantage of the distinguisher. If $\epsilon > 1$, we call p a good distinguisher.*

Let $\mathcal{V}$ be the linear space of polynomials with degree at most k. The best possible advantage achieved by polynomials in this class is given by the *low-degree likelihood ratio* (see (Brennan et al., 2021, Fact 2.1))

$$\max_{p \in \mathcal{V}, \mathbb{E}_{X \in D_0^{\otimes n}}[p^2(X)] \leq 1} \left| \mathbb{E}_{X \sim D_0^{\otimes n}}[p(X)] - \mathbb{E}_{\mathbf{u} \sim \mu} \mathbb{E}_{X \sim D_{\mathbf{u}}^{\otimes n}}[p(X)] \right| = \left\| \mathbb{E}_{\mathbf{u} \sim \mu} \left[(\bar{D}_{\mathbf{u}}^{\otimes n})^{\leq k} \right] - 1 \right\|_{D_0^{\otimes n}},$$

where $\bar{D}_{\mathbf{u}} := D_{\mathbf{u}}/D_0$ denotes the likelihood ratio and the notation $f^{\leq k}$ denotes the orthogonal projection of f onto $\mathcal{V}$.

Another notation we will use regarding a finer notion of degrees is the following: we say that the polynomial $f(\{\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i\}_{i=1}^n) : (\mathbb{R}^{p^d + p^m})^{\otimes n} \to \mathbb{R}$ for Problem 28 has *samplewise* degree-(r, k) if it is a polynomial, where each monomial uses at most k different samples from $\{\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i\}_{i=1}^n$ and uses degree at most r for each of them. Note that a function of samplewise degree-(r, k) has degree at most rk. In analogy to what was stated for the best degree-k distinguisher, the best unit norm distinguisher of samplewise degree-(r, k) achieves advantage $\left\| \mathbb{E}_{\mathbf{u} \sim \mu} \left[(\bar{D}_{\mathbf{u}}^{\otimes n})^{\leq r, k} \right] - 1 \right\|_{D_0^{\otimes n}}$, where the notation $f^{\leq r, k}$ now means the orthogonal projection of f onto the space of all samplewise degree-(r, k) polynomials.

A.1. Hardness of Tensor-on-tensor Association Detection Against Low-degree Polynomials

In this section, we establish the following result.

Theorem 30 *Consider the hypothesis testing Problem 28 with sufficiently large p. Given any $0 < \epsilon < 1/4$ and sufficiently large p. Suppose $\lambda/\sigma \leq 1/2$ for $m = 0$, $\lambda/\sigma \leq c_1 \underline{p}^{1/4 - \epsilon}$ for $m = 1$ and $\lambda/\sigma \leq c_1 \underline{p}^{1/2 - 2\epsilon}$ for $m \geq 2$ with sufficiently small $c_1 > 0$. Then if $n \leq c_2((\sigma/\lambda)^2 \underline{p}^{(1/2 - 2\epsilon)(d+m)} + (\sigma/\lambda)^4 \underline{p}^{(1 - 4\epsilon)m})$ for some sufficiently small $c_2 > 0$, then there is no n-sample good distinguisher with samplewise degree-(r, k) for any $r \in \mathbb{N}$ and any even integer $k < p^\epsilon$.*

Since evaluating a degree rk polynomial with n samples is a proxy for running time $O((p^{d+m} n)^{rk})$, Theorem 30 provides strong evidence for the hardness of solving the hypothesis testing Problem 28.

To prove Theorem 30, we need the following two results from Brennan et al. (2021).

Fact 31 *(Brennan et al., 2021, Fact 2.2) If $\left\| \mathbb{E}_{\mathbf{u} \sim \mu} \left[(\bar{D}_{\mathbf{u}}^{\otimes n})^{\leq r, k} \right] - 1 \right\|_{D_0^{\otimes n}} \leq \epsilon$, then the hypothesis testing Problem 28 has no n-sample ϵ-distinguisher in polynomials with samplewise degree-(r, k).*

Theorem 32 *(Brennan et al., 2021, Theorem 4.1) Let $\mathcal{S}$ be a hypothesis testing problem on $\mathbb{R}^N$ with respect to null hypothesis D_0. Let $n, k \in \mathbb{N}$ with k even. Suppose that for all $1 \le n' \le n$, $\mathrm{SDA}(\mathcal{S}, \mu, n') \ge 100^k \cdot (n/n')^k$. Then for all $r \in \mathbb{N}$, $\left\|\mathbb{E}_{\mathbf{u}\sim\mu}\left[(\bar{D}_{\mathbf{u}}^{\otimes n})^{\le r, \Omega(k)}\right] - 1\right\|_{D_0^{\otimes n}} \le 1$.*

Proof [Proof of Theorem 30] First notice $\mathrm{SDA}(\mathcal{S}, \mu, n') \ge \mathrm{SDA}(\mathcal{S}, \mu, n)$ for $n' \le n$. So by taking $q = 100^k(n/n')^k$ with $k < p^\epsilon$, $n \le c_2((\sigma/\lambda)^2 \underline{p}^{(1/2-2\epsilon)(d+m)} + (\sigma/\lambda)^4 \underline{p}^{(1-4\epsilon)m})$, we have $\mathrm{SDA}(\mathcal{S}, \mu, n') \ge 100^k(n/n')^k$ by Theorem 19. This further implies

$$\left\|\mathbb{E}_{\mathbf{u}\sim\mu}\left[(\bar{D}_{\mathbf{u}}^{\otimes n})^{\le r, \Omega(k)}\right] - 1\right\|_{D_0^{\otimes n}} \le 1$$

for all $r \in \mathbb{N}$ by Theorem 32. Thus the result follows from Fact 31. ■

Appendix B. A Simpler Proof of Statistical Query Lower Bound via Pairwise Correlation with Suboptimal Dependence on λ

In this section, we provide a simpler proof of the statistical query lower bound via the statistical dimension defined by pairwise correlation comparing the proof of Theorem 19, but we will see the dependence on λ is suboptimal.

Definition 33 *We say that a set of s distributions $\mathcal{D} = \{D_1, \ldots, D_s\}$ over $\mathbb{R}^q$ is (γ, β)-correlated relative to a distribution D if $|\chi_D(D_i, D_j)| \le \gamma$ for all $i \ne j$, and $|\chi_D(D_i, D_j)| \le \beta$ for $i = j$.*

We are now ready to define the SQ dimension based on pairwise correlation. To distinguish it with the one in Definition 17, we use notation SD for it.

Definition 34 (Statistical Query Dimension Based on Pairwise Correlation) *For $\beta, \gamma > 0$, a decision problem $\mathcal{B}(\mathcal{D}, D)$, where D is a fixed distribution and $\mathcal{D}$ is a family of distributions over $\mathbb{R}^q$, let s be the maximum integer such that there exists a finite set of distributions $\mathcal{D}_D \subseteq \mathcal{D}$ such that $\mathcal{D}_D$ is (γ, β)-correlated relative to D and $|\mathcal{D}_D| \ge s$. We define the* Statistical Query dimension *with pairwise correlations (γ, β) of $\mathcal{B}$ to be s and denote it by $\mathrm{SD}(\mathcal{B}, \gamma, \beta)$.*

If one can bound below the SD of the given problem, then it also implies an unconditional lower bound on the complexity of any SQ algorithm for the problem using the following result.

Lemma 35 (Corollary 3.12 in Feldman et al. (2017a)) *Let $\mathcal{B}(\mathcal{D}, D)$ be a decision problem, where D is the reference distribution and $\mathcal{D}$ is a class of distributions. For $\gamma, \beta > 0$, let $s = \mathrm{SD}(\mathcal{B}, \gamma, \beta)$. For any $\gamma' > 0$, any SQ algorithm for solving $\mathcal{B}$ requires at least $s \cdot \gamma'/(\beta - \gamma)$ queries to the $\mathrm{VSTAT}(1/(3(\gamma + \gamma')))$ oracle.*

Next, we provide a lower bound on the SQ dimension for the tensor-on-tensor association detection problem when $\lambda/\sigma \le 1/2$.

Theorem 36 (Statistical Query Lower Bound) *Suppose $\lambda/\sigma \le 1/2$. Given any $0 < c < \frac{1}{2}$, the statistical query dimension of the tensor-on-tensor association detection with pairwise correlation $(c_0((\lambda/\sigma)^2 \underline{p}^{(c-1/2)(d+m)} + (\lambda/\sigma)^4 \underline{p}^{2(c-1/2)m}), (1 - 2(\lambda/\sigma)^2)^{-1/2} - 1)$ for some constants $c_0 > 0$ is at least $O(2^{\Omega(\underline{p}^c)})$, where $\underline{p} = \min_j p_j$.*

Proof We consider a subclass of parameters $\mathscr{A} = \{\boldsymbol{\mathcal{A}} = \lambda \mathbf{u}^{d+m}, \mathbf{u} \in \mathcal{S}_{p-1}\}$ and without loss of generality assume $\sigma = 1$. Recall we use notation $D_{\mathbf{u}}$ to denote the joint distribution of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ with tensor parameter $\mathscr{A} = \lambda \mathbf{u}^{\otimes(d+m)}$; and notation D_0 to denote the distribution of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ under H_0. Our goal is to bound the χ^2-divergence between $D_{\mathbf{u}}$ and D_0 and to find a large collection of distributions $\{D_{\mathbf{u}}\}_{\mathbf{u}\in\mathcal{U}}$ with weak pairwise correlation.

First by Lemma 38, we have

$$\chi^2(D_{\mathbf{u}}, D_0) + 1 = (1 - 2\lambda^2)^{-1/2},$$

as $\lambda \leq 1/2$ by assumption.

Moreover, for $\mathbf{u} \neq \mathbf{v}$, we have

$$\begin{aligned}
&\chi_{D_0}(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i), D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)) + 1 \\
&= \int \frac{D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i) D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)}{D_0(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)} d\boldsymbol{\mathcal{Y}}_i d\boldsymbol{\mathcal{X}}_i \\
&\overset{\text{Lemma 38}}{=} \mathbb{E}_{\boldsymbol{\mathcal{X}}_i} \exp\left(\lambda^2 \langle \mathbf{u}, \mathbf{v}\rangle^m \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle \langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle\right) \\
&\overset{(a)}{\leq} \left(1 - \langle \mathbf{u}, \mathbf{v}\rangle^{d+m}\lambda^2\right)^{-1/2} \cdot \left(1 - \lambda^2 \langle \mathbf{u}, \mathbf{v}\rangle^{d+m} - \frac{\lambda^4 \langle \mathbf{u}, \mathbf{v}\rangle^{2m}}{1 - \lambda^2 \langle \mathbf{u}, \mathbf{v}\rangle^{d+m}}\right)^{-1/2} \\
&\overset{(b)}{\leq} \left(1 + c_0(\lambda^2 \langle \mathbf{u}, \mathbf{v}\rangle^{d+m} + \lambda^4 \langle \mathbf{u}, \mathbf{v}\rangle^{2m})\right),
\end{aligned} \tag{7}$$

where (a) is because of Lemma 38 Eq. (49) and the fact when $\lambda \leq 1/2$, we have $\lambda^2 \leq 1/4$, so $1 - t^d - \lambda^2 t^m (1 - t^{2d}) = (1 - t^d)(1 - \lambda^2 t^m (1 + t^d)) > 0$ and $t^d + \lambda^2 t^m (1 - t^{2d}) + 1 = (1 + t^d)(1 + \lambda^2 t^m (1 - t^d))) > 0$, i.e., $|t^d + \lambda^2 t^m (1 - t^{2d})| < 1$, where $t = \langle \mathbf{u}, \mathbf{v}\rangle$; (b) is because $\lambda \leq 1/2$ and Lemma 41.

By Lemma 45, for any $0 < c < 1/2$ we can construct at least $N = 2^{\Omega(p^c)}$ number of unit vectors $\mathcal{U} = \{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(N)}\}$ such that $\max_{\mathbf{u}\neq\mathbf{v}\in\mathcal{U}} |\mathbf{u}^\top \mathbf{v}| \leq O(p^{-1/2+c})$. This further implies that

$$\max_{\mathbf{u}\neq\mathbf{v}\in\mathcal{U}} \chi_{D_0}(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}, \mathscr{X}), D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}, \mathscr{X})) \leq c_0(\lambda^2 p^{(-1/2+c)(d+m)} + \lambda^4 p^{(-1/2+c)(2m)}).$$

This shows the statistical query dimension of the tensor-on-tensor association detection with pairwise correlation $(c_0((\lambda/\sigma)^2 p^{(c-1/2)(d+m)} + (\lambda/\sigma)^4 p^{2(c-1/2)m}), (1 - 2(\lambda/\sigma)^2)^{-1/2} - 1)$ for some constants $c_0 > 0$ is at least $O(2^{\Omega(p^c)})$. ■

We note that no truncation is performed in the proof of Theorem 36. If we perform truncation, then we can improve λ/σ to be of scale up to Poly($\log p$), but it is unclear for us how to show the sharp SQ lower bound when λ/σ scales up polynomially in p.

Appendix C. Proofs in Section 3

C.1. Proof of Theorem 1

Consider the special class of parameters $\mathscr{A} = \{\lambda \mathbf{u}^{\otimes(d+m)}, \mathbf{u} \in \mathcal{C}\}$ and we consider the following prior on $\mathscr{A}$: $\boldsymbol{\mathcal{A}} = \lambda \mathbf{u}^{d+m}$ with $\mathbf{u}$ uniformly at random chosen from $\mathcal{C}$. Then based on Lemma 27, it is enough to show $\chi^2(D_\mu(\boldsymbol{\mathcal{Y}}, \mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}, \mathscr{X})) = o(1)$, or equivalently

$\chi^2(D_\mu(\boldsymbol{\mathcal{Y}},\mathscr{X}),D_0(\boldsymbol{\mathcal{Y}},\mathscr{X}))+1 = 1+o(1)$, to prove our results, where D_μ and D_0 denote the density function under the mixture model and null, respectively.

Then

$$\begin{aligned}
&\chi^2(D_\mu(\boldsymbol{\mathcal{Y}},\mathscr{X}),D_0(\boldsymbol{\mathcal{Y}},\mathscr{X}))+1\\
&=\int \frac{[\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}[\mathcal{C}]}D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}},\mathscr{X})]^2}{D_0(\boldsymbol{\mathcal{Y}},\mathscr{X})}d\boldsymbol{\mathcal{Y}}d\mathscr{X}\\
&=\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}[\mathcal{C}],\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\int \frac{D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}},\mathscr{X})D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}},\mathscr{X})}{D_0(\boldsymbol{\mathcal{Y}},\mathscr{X})}d\boldsymbol{\mathcal{Y}}d\mathscr{X}\\
&\overset{\text{Lemma 38}}{=}\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}[\mathcal{C}],\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\mathbb{E}_{\mathscr{X}}\exp\left(\lambda^2\sum_{i=1}^n\langle\boldsymbol{\mathcal{X}}_i,\mathbf{u}^{\otimes d}\rangle\langle\boldsymbol{\mathcal{X}}_i,\mathbf{v}^{\otimes d}\rangle\right)\\
&\overset{(a)}{\le}\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}[\mathcal{C}],\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\exp(n\lambda^2\langle\mathbf{u}^{\otimes d},\mathbf{v}^{\otimes d}\rangle)(1+C_0c^2)\\
&=\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}[\mathcal{C}],\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\exp(n\lambda^2\langle\mathbf{u},\mathbf{v}\rangle^d)(1+C_0c^2)\\
&\overset{(b)}{\le} g(\frac{n\lambda^2}{p})(1+C_0c^2)\to 1+o(1),
\end{aligned}$$

where in (a), we use the fact $\lambda\le cn^{-1/4}$ for some small enough $c>0$ and Lemma 39, moreover $C_0>0$ here is some universal constant; (b) is by Lemma 40 as $d\ge 2$ and $\lambda=o(\sqrt{p/n})$; the final conclusion follows from the assumption that $\lambda=o(\sqrt{p/n})$, the function g satisfies $g(0+)=1$ and the fact we can let c goes to zero as $\lambda=o(n^{-1/4})$ by assumption. This finishes the proof of this theorem.

C.2. Proof of Theorem 2

We divide the proof into two steps, in Step 1, we prove the lower bound for $m=1,d=0$ case and in Step 2, we prove the lower bound for $m=1,d\ge 1$.

Step 1. We follow the idea described in Section 7. Recall, we have model $\mathbf{Y}=\lambda\mathbf{u}\mathbf{x}^\top+\mathbf{E}\in\mathbb{R}^{p\times n}$. Denote $D_\mu(\mathbf{Y},\mathbf{x})$ and $D_0(\mathbf{Y},\mathbf{x})$ as the density function of the data $(\mathbf{Y},\mathbf{x})$ under distribution $\mathbb{P}_\mu$ and $\mathbb{P}_0$, respectively. Similarly, let $D_\mu(\mathbf{Y}|\mathbf{x})$ and $D_0(\mathbf{Y}|\mathbf{x})$ to denote the conditional density of $\mathbf{Y}$ given $\mathbf{x}$. For notation simplicity, we use notation $\mathrm{TV}(D_\mu(\mathbf{Y},\mathbf{x}),D_0(\mathbf{Y},\mathbf{x}))$ to mean $\mathrm{TV}(\mathbb{P}_\mu,\mathbb{P}_0)$, similarly for $\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}),D_0(\mathbf{Y}|\mathbf{x}))$.

First, recall the good event is $A:=\{|\|\mathbf{x}\|_2^2-n|<n/2\}$, then

$$\begin{aligned}
\mathrm{TV}(D_\mu(\mathbf{Y},\mathbf{x}),D_0(\mathbf{Y},\mathbf{x}))&=\frac{1}{2}\int|D_\mu(\mathbf{Y},\mathbf{x})-D_0(\mathbf{Y},\mathbf{x})|d\mathbf{Y}d\mathbf{x}\\
&=\mathbb{E}_{\mathbf{x}}\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}),D_0(\mathbf{Y}|\mathbf{x}))\\
&=\mathbb{E}_{\mathbf{x}}\left((1_A+1_{A^c})\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}),D_0(\mathbf{Y}|\mathbf{x}))\right)\\
&\overset{(a)}{\le}\exp(-cn)+\mathbb{E}_{\mathbf{x}}1_A\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}),D_0(\mathbf{Y}|\mathbf{x})),
\end{aligned}\tag{8}$$

here (a) is because TV distance is at most 1 and the probability of event A^c is at most $\exp(-cn)$.

Next, we show condition on A, $\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}), D_0(\mathbf{Y}|\mathbf{x}))$ is small:

$$\begin{aligned}\mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{x}), D_0(\mathbf{Y}|\mathbf{x})) &\overset{(6)}{\leq} \frac{1}{2}\sqrt{\chi^2(D_\mu(\mathbf{Y}|\mathbf{x}), D_0(\mathbf{Y}|\mathbf{x}))} \\ &\leq \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \mathrm{Unif}[\mathcal{C}]}\left[\int \frac{D_\mathbf{u}(\mathbf{Y}|x)D_\mathbf{u}(\mathbf{Y}|x)}{D_0(\mathbf{Y}|\mathbf{x})}d\mathbf{Y}\right]} \\ &\overset{(a)}{\leq} \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \mathrm{Unif}[\mathcal{C}]}\exp(\lambda^2(\sum_{i=1}^n x_i^2)\langle \mathbf{u},\mathbf{v}\rangle) - 1} \\ &\overset{(b)}{\leq} \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \mathrm{Unif}[\mathcal{C}]}\exp(3n\lambda^2\langle \mathbf{u},\mathbf{v}\rangle/2) - 1} \\ &\overset{(c)}{\leq} \frac{1}{2}\sqrt{\exp(Cn^2\lambda^4/p) - 1} \overset{(d)}{\to} o(1),\end{aligned} \tag{9}$$

here (a) is by Lemma 38 Eq. (46); (b) is because we condition on event A; (c) is because $\langle \mathbf{u},\mathbf{v}\rangle$ follows a sub-Gaussian distribution with sub-Gaussian norm $1/\sqrt{p}$ and the bound for the moment generating function for the sub-Gaussian random variable, see (Vershynin, 2010, Eq. (5.12)); (d) is because $\lambda^4 n^2/p = o(1)$ by assumption.

By plugging (9) into (8), we have $\mathrm{TV}(D_\mu(\mathbf{Y},\mathbf{x}), D_0(\mathbf{Y},\mathbf{x})) = o(1)$ when $\lambda^4 n^2/p = o(1)$ and the result follows from Lemma 27.

Step 2. First, we claim that to prove the lower bound for $m = 1, d \geq 1$, it is enough to consider the $m = 1, d = 1$ case. The reason is the following: suppose under H_1, the model is

$$\mathbf{y}_i = \langle \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+1}, \boldsymbol{\mathcal{X}}_i\rangle_* + \boldsymbol{\mathcal{E}}_i \overset{(51)}{=} \langle \lambda \mathbf{u}_d \otimes \mathbf{u}_{d+1}, \boldsymbol{\mathcal{X}} \times_{i=1}^{d-1} \mathbf{u}_i^\top\rangle_* + \boldsymbol{\mathcal{E}}_i;$$

if there is an oracle tells us $\mathbf{u}_1,\ldots,\mathbf{u}_{d-1}$, then $\boldsymbol{\mathcal{X}} \times_{i=1}^{d-1} \mathbf{u}_i^\top$ still has i.i.d. $N(0,1)$ entries. So we can view $\boldsymbol{\mathcal{X}} \times_{i=1}^{d-1} \mathbf{u}_i^\top$ as the new covariate, and the model is the same as the model in H_1 when $m = 1, d = 1$ case.

To prove the result in the $m = 1, d = 1$ case, we assume the following prior $\mathbf{A} = \lambda \mathbf{u}_1\mathbf{u}_2^\top$, where $\mathbf{u}_1, \mathbf{u}_2$ are generated independently and uniformly at random from $\mathcal{C}$. We denote this prior by μ. By stacking $\{\mathbf{y}_i, \mathbf{x}_i\}$s together, we have $\mathbf{Y} = \lambda \mathbf{u}_2\mathbf{u}_1^\top \mathbf{X}^\top + \mathbf{E} \in \mathbb{R}^{p\times n}$, where each column of $\mathbf{Y}$ is equal to $\mathbf{y}_i$ for $i = 1,\ldots,n$.

We also need the truncation argument as in Step 1, but now we not only need to truncate $\mathbf{X}$, but also $\mathbf{u}_1$. Denote the good event as

$$A = A_1 \bigcap A_2 = \{\mathbf{X}, \mathbf{u}_1 : \|\mathbf{X}^\top\mathbf{X}\|_\mathbf{F}^2 \leq C_1(n\vee p)^2(n\wedge p), \|\mathbf{X}\| \leq C_2\sqrt{n\vee p}, \|\mathbf{X}\mathbf{u}_1\|_2 \leq C_3\sqrt{n}\},$$

where $A_1 = \{\mathbf{X} : \|\mathbf{X}^\top\mathbf{X}\|_\mathbf{F}^2 \leq C_1(n\vee p)^2(n\wedge p), \|\mathbf{X}\| \leq C_2\sqrt{n\vee p}\}$ and $A_2 := \{\|\mathbf{X}\mathbf{u}_1\|_2 \leq C_3\sqrt{n}\}$. Then

$$\begin{aligned}&\mathbb{P}_{\mathbf{X},\mathbf{u}_1}(A^c) \\ &\leq \mathbb{P}_{\mathbf{X},\mathbf{u}_1}(\|\mathbf{X}^\top\mathbf{X}\|_\mathbf{F}^2 > C_1(n\vee p)^2(n\wedge p)) + \mathbb{P}_{\mathbf{X},\mathbf{u}_1}(\|\mathbf{X}\| > C_2\sqrt{n\vee p}) + \mathbb{P}_{\mathbf{X},\mathbf{u}_1}(\|\mathbf{X}\mathbf{u}_1\|_2 > C_3\sqrt{n}) \\ &\leq \exp(-c(n\wedge p)),\end{aligned} \tag{10}$$

here we use a few facts to prove this inequality, the first two quantities are bounded by the fact that $\|\mathbf{X}\| > C\sqrt{n \vee p}$ with probability at most $\exp(-c(n \wedge p))$ by Corollary 5.35 of Vershynin (2010) and $\mathbf{X}^\top \mathbf{X}$ has rank at most $n \wedge p$; the final quantity is bounded by the fact $\mathbf{X}\mathbf{u}_1$ has i.i.d. $N(0,1)$ entries, $\|\mathbf{X}\mathbf{u}_1\|_2^2 \sim \chi_n^2$ and the concentration bound for the χ_n^2 distribution (Vershynin, 2010, Proposition 5.16).

Next, we bound $\mathrm{TV}(D_\mu(\mathbf{Y},\mathbf{X}), D_0(\mathbf{Y},\mathbf{X}))$.

$$\begin{aligned}
&\mathrm{TV}(D_\mu(\mathbf{Y},\mathbf{X}), D_0(\mathbf{Y},\mathbf{X})) \\
&= \mathbb{E}_{\mathbf{X}} \mathrm{TV}(D_\mu(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X})) \\
&= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} \left((1_A + 1_{A^c}) \mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right) \\
&\overset{(10)}{\leq} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} \left(1_A \mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right) + \exp(-c(n \wedge p)) \\
&= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} 1_A \left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right) + \exp(-c(n \wedge p)) \\
&= \mathbb{E}_{\mathbf{X}} 1_{A_1} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} 1_{A_2} \left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right) + \exp(-c(n \wedge p)),
\end{aligned} \tag{11}$$

here the notation $D_{\mathbf{u}_1,\mu_2}$ denotes the distribution of the data $(\mathbf{Y},\mathbf{X})$ when $\mathbf{u}_1$ is fixed while $\mathbf{u}_2 \sim \mathrm{Unif}[\mathcal{C}]$.

Then we have

$$\begin{aligned}
\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X})) &\overset{(6)}{\leq} \frac{1}{2}\sqrt{\chi^2(_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))} \\
&\overset{(a)}{\leq} \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u}_2,\mathbf{u}_2' \sim \mathrm{Unif}[\mathcal{C}]} \exp(\lambda^2 \langle \mathbf{u}_2, \mathbf{u}_2' \rangle \mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1) - 1} \\
&\overset{(b)}{\leq} \frac{1}{2}\sqrt{\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1)^2\right) - 1},
\end{aligned} \tag{12}$$

where (a) is by a similar argument as in (50) and (b) is because $\langle \mathbf{u}_2, \mathbf{u}_2' \rangle$ follows a sub-Gaussian distribution with sub-Gaussian norm $1/\sqrt{p}$

Based on (12) and the Jensen's inequality, we have

$$\begin{aligned}
&\left(\mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} \left\{1_{A_1} 1_{A_2} \left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right)\right\}\right)^2 \\
\leq &\mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} \left\{1_{A_1} 1_{A_2} \left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\mathbf{Y}|\mathbf{X}), D_0(\mathbf{Y}|\mathbf{X}))\right)^2\right\} \\
\overset{(12)}{\leq} &\frac{1}{4} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} 1_{A_1} 1_{A_2} \left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1)^2\right) - 1\right) \\
= &\frac{1}{4} \mathbb{E}_{\mathbf{X}} 1_{A_1} \mathbb{E}_{\mathbf{u}_1 \sim \mathrm{Unif}[\mathcal{C}]} 1_{A_2} \left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1)^2\right) - 1\right).
\end{aligned} \tag{13}$$

Since $\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)$ is a nonnegative random variable, we have given on $\mathbf{X}$,

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]} 1_{A_2}\left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\right)\\
&= \int_0^\infty \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(1_{A_2}\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\geq z\right)dz\\
&\overset{(a)}{=} \int_0^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(1_{A_2}\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\geq z\right)dz\\
&\leq \int_0^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\geq z\right)dz\\
&= 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\geq z\right)dz\\
&= 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \mathbb{P}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]}\left(|\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1|\geq \sqrt{\frac{p\log z}{c\lambda^4}}\right)dz\\
&\overset{(b)}{\leq} 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \exp\left(-c(\frac{\Delta^2p^2}{\|\mathbf{X}^\top\mathbf{X}\|_{\mathbf{F}}^2}\wedge\frac{\Delta p}{\|\mathbf{X}^\top\mathbf{X}\|})\right)dz,
\end{aligned}
\tag{14}
$$

here (a) is because on event A_2, we have $|\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1|^2 = \|\mathbf{X}\mathbf{u}_1\|_2^4 \leq Cn^2$; (b) is by the Hanson-Wright inequality (Rudelson and Vershynin, 2013) with $\Delta := \sqrt{\frac{p\log z}{c\lambda^4}}$ and entries of $\mathbf{u}_1$ are i.i.d. sub-Gaussian random variable with sub-Gaussian norm $1/\sqrt{p}$.

Thus condition on event A_1, from (14) we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{u}_1\sim \mathrm{Unif}[\mathcal{C}]} 1_{A_2}\left(\exp\left(\frac{c\lambda^4}{p}(\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1)^2\right)\right)\\
&\leq 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \exp\left(-c(\frac{\Delta^2p^2}{\|\mathbf{X}^\top\mathbf{X}\|_{\mathbf{F}}^2}\wedge\frac{\Delta p}{\|\mathbf{X}^\top\mathbf{X}\|})\right)dz\\
&\leq 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \exp\left(-c'\left(\frac{\Delta^2p^2}{(n\vee p)^2(n\wedge p)}\wedge\frac{\Delta p}{n\vee p}\right)\right)dz\\
&\overset{(a)}{\leq} 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \exp\left(-c'\left(\frac{\Delta^2p^2}{(n\vee p)^2(n\wedge p)}\right)\right)dz\\
&= 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} \exp\left(-c'\left(\frac{p^3\log z}{(n\vee p)^2(n\wedge p)\lambda^4}\right)\right)dz\\
&= 1+\int_1^{\exp\frac{Cn^2\lambda^4}{p}} z^{\left(-c'\left(\frac{p^3}{(n\vee p)^2(n\wedge p)\lambda^4}\right)\right)}dz\\
&\overset{(b)}{=} 1+o(1),
\end{aligned}
\tag{15}
$$

here (a) is because when $z \leq \exp(\frac{Cn^2\lambda^4}{p})$, we have $\frac{\Delta^2 p^2}{(n\vee p)^2(n\wedge p)} \leq c\frac{\Delta p}{n\vee p}$; (b) is because when $\lambda = o((p/n)^{1/2} \wedge (p/n)^{1/4})$, we have $-c'\left(\frac{p^3}{(n\vee p)^2(n\wedge p)\lambda^4}\right) \to -\infty$ and thus $\int_1^{\exp\frac{Cn^2\lambda^4}{p}} z^{\left(-c'\left(\frac{p^3}{(n\vee p)^2(n\wedge p)\lambda^4}\right)\right)} dz = o(1)$.

By plugging (15) into (13), we have

$$\begin{aligned}
&\left(\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{u}_1\sim\mathrm{Unif}[\mathcal{C}]}1_{A_1}1_{A_2}\left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\boldsymbol{\mathcal{Y}}|\mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}|\mathscr{X}))\right)\right)^2\\
\overset{(15)}{\leq}&\frac{1}{4}\left(\mathbb{E}_{\mathbf{X}}1_{A_1}(1+o(1)) - \mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{u}_1\sim\mathrm{Unif}[\mathcal{C}]}1_A\right) \overset{(a)}{\to} o(1),
\end{aligned}$$

here (a) is because both event A and A_1 holds with probability goes to 1 when $n, p \to \infty$ by (10). This concludes that $\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{u}_1\sim\mathrm{Unif}[\mathcal{C}]}1_{A_1}1_{A_2}\left(\mathrm{TV}(D_{\mathbf{u}_1,\mu_2}(\boldsymbol{\mathcal{Y}}|\mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}|\mathscr{X}))\right) = o(1)$ when $\lambda = o((p/n)^{1/2} \wedge (p/n)^{1/4})$ and the result follows by considering (11) and Lemma 27.

C.3. Proof of Theorem 3

We first consider an easier hypothesis testing problem. Suppose other than the data $\{\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i\}_{i=1}^n$, the oracle provides the following additional information: under H_0, it provides $\{x_i\}_{i=1}^n$, where $x_i := \langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle$ for some arbitrary $\mathbf{v}$; under H_1, suppose $\boldsymbol{\mathcal{A}} = \lambda\mathbf{u}^{\otimes(d+m)}$, then the oracle provides $\{x_i\}_{i=1}^n$ where $x_i := \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle$. Since under H_1, $\boldsymbol{\mathcal{Y}}_i = \lambda\mathbf{u}^{\otimes m}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle + \boldsymbol{\mathcal{E}}_i$, given x_i, $\boldsymbol{\mathcal{Y}}_i$ is conditionally independent of $\boldsymbol{\mathcal{X}}_i$, an easier hypothesis testing problem comparing the original one is the following: given i.i.d. data $\{\boldsymbol{\mathcal{Y}}_i, x_i\}_{i=1}^n$, $\boldsymbol{\mathcal{Y}}_i = x_i\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{E}}_i$ where x_i and $\boldsymbol{\mathcal{E}}_i$ has i.i.d. $N(0,1)$ entries, we would like to test

$$H_0 : \boldsymbol{\mathcal{A}}' = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mathcal{A}}' \in \mathscr{A}(\lambda), \tag{16}$$

where for $\lambda > 0$, $\mathscr{A}(\lambda) = \{\lambda\mathbf{u}^{\otimes m} : \mathbf{u} \in \mathcal{S}_{p-1}\}$. This is the original testing problem with $m \geq 2, d = 0$. In summary, to prove the original statement, we only need to show there is no reliable testing procedure when $\lambda/\sigma = o((p/n)^{1/2})$ in the $d = 0, m \geq 2$ case.

Let the good event be $A := \{|\sum_{i=1}^n x_i^2 - n| < n/2\}$, where $x_i \overset{i.i.d.}{\sim} N(0,1)$, and by the same argument as in (8), if we can show $\mathrm{TV}(D_\mu(\boldsymbol{\mathcal{Y}}|\mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}|\mathscr{X}))$ is small on event A when $\lambda/\sigma = o((p/n)^{1/2})$, where μ is the prior on $\mathscr{A}$ chosen in the same way as in the proof of Theorem 1, then we are done.

Given A,

$$\begin{aligned}
\mathrm{TV}(D_\mu(\boldsymbol{\mathcal{Y}}|\mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}|\mathscr{X})) &\leq \frac{1}{2}\sqrt{\chi^2(D_\mu(\boldsymbol{\mathcal{Y}}|\mathscr{X}), D_0(\boldsymbol{\mathcal{Y}}|\mathscr{X}))}\\
&\overset{(a)}{\leq} \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\exp(\lambda^2(\sum_{i=1}^n x_i^2)\langle\mathbf{u},\mathbf{v}\rangle^m) - 1}\\
&\overset{(b)}{\leq} \frac{1}{2}\sqrt{\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\mathrm{Unif}[\mathcal{C}]}\exp(3n\lambda^2\langle\mathbf{u},\mathbf{v}\rangle^m/2) - 1}\\
&\overset{(c)}{\leq} \frac{1}{2}\sqrt{g(Cn\lambda^2/p) - 1} \overset{(c)}{\to} o(1),
\end{aligned} \tag{17}$$

here (a) is by Lemma 38 Eq. (46); (b) is because we condition on event A; (c) is by Lemma 40 and $\lambda/\sigma = o((p/n)^{1/2})$. This finishes the proof of this theorem.

Appendix D. Proofs in Section 4

Throughout the proof in this section, we will again assume $\sigma = 1$ without loss of generality and we consider $\boldsymbol{\mathcal{A}} = \lambda' \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}$ with $\lambda' = \lambda$ as detection only gets easier as λ' increases. Moreover, the following representation for the model (1) is used a lot in the proof. By Lemma 44 Eq. (53), we have $\boldsymbol{\mathcal{Y}}_i = \lambda \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d \rangle \mathbf{u}_{d+1} \otimes \cdots \otimes \mathbf{u}_{d+m} + \boldsymbol{\mathcal{E}}_i$. So under H_1, we can write $\boldsymbol{\mathcal{Y}} = \lambda \|\mathbf{w}\| \mathbf{u}_{d+1} \otimes \cdots \otimes \mathbf{u}_{d+m} \otimes \mathbf{w}/\|\mathbf{w}\| + \boldsymbol{\mathcal{E}}$ where

$$\mathbf{w} = (\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d \rangle, \cdots, \langle \boldsymbol{\mathcal{X}}_n, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d \rangle)^\top. \tag{18}$$

D.1. Proof of Theorem 5

First, we control the type I error,

$$\mathbb{P}_0 \left(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1 \right) = \mathbb{P}_0 \left(T_1 \geq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) \overset{\text{Lemma 43}(b)}{\leq} \exp(-c_\alpha (\sum_{j=1}^{d+m} p_i)) \leq \alpha/4,$$

here the last inequality holds by picking $Z_{1\alpha}$ to be large enough.

Now, let us compute the type II error. Suppose $\boldsymbol{\mathcal{A}} = \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}$. Recall $\boldsymbol{\mathcal{Y}} = \mathscr{A}(\boldsymbol{\mathcal{X}}) + \boldsymbol{\mathcal{E}}$, then

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0 \right) &= \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(T_1 \leq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) \\
&= \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\sup_{\mathbf{v}_j \in \mathcal{S}_{p_j - 1}, j=1,\ldots,d+m} \mathscr{X}^*(\boldsymbol{\mathcal{Y}}) \times_{i=1} \mathbf{v}_1^\top \times \cdots \times_{d+m} \mathbf{v}_{d+m}^\top \leq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) \\
&= \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\sup_{\mathbf{v}_j \in \mathcal{S}_{p_j - 1}, j=1,\ldots,d+m} \langle \mathscr{X}^*(\boldsymbol{\mathcal{Y}}), \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_{d+m} \rangle \leq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) \\
&\leq \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\langle \boldsymbol{\mathcal{Y}}, \mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}) \rangle \leq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) \\
&\leq \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\lambda \|\mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m})\|_{\mathbf{F}}^2 \leq Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} + |\langle \boldsymbol{\mathcal{E}}, \mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}) \rangle| \right) \\
&\overset{(a)}{\leq} \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\lambda \|\mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m})\|_{\mathbf{F}}^2 \leq 2 Z_{1\alpha} \sqrt{n(\sum_{j=1}^{d+m} p_j)} \right) + \alpha/4 \\
&\overset{(b)}{\leq} \mathbb{P}_{\boldsymbol{\mathcal{A}}} \left(\|\mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m})\|_{\mathbf{F}}^2 / n - 1 \leq -1/2 \right) + \alpha/4 \\
&\overset{(c)}{\leq} 2\exp(-cn) + \alpha/4 \leq 3\alpha/4,
\end{aligned}$$

where (a) is because we have shown in the H_0 case that $Z_{1\alpha}$ is chosen such that

$$|\langle \boldsymbol{\mathcal{E}}, \mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m})\rangle| \le Z_{1\alpha}\sqrt{n(\sum_{j=1}^{d+m} p_j)}$$

holds with probability at least $1-\alpha/4$; (b) is because $\lambda > 4Z_{1\alpha}\sqrt{\frac{\sum_{j=1}^{d+m} p_j}{n}}$ by our assumption; (c) is because by Lemma 44 Eq. (53), we have

$$\|\mathscr{X}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m})\|_{\mathbf{F}}^2 = \sum_{i=1}^{n} \|\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}\rangle_*\|_{\mathbf{F}}^2 = \sum_{i=1}^{n} \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d\rangle^2,$$

and by the fact of Lemma 42. The last inequality holds by choosing n to be large enough. This finishes the proof of this theorem.

D.2. Proof of Theorem 6

First, we have

$$\mathbb{P}_0\left(\phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) = \mathbb{P}_0\left(T_2 \ge Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right) \overset{\text{Lemma 43}(a)}{\le} \exp(-c_\alpha(n + \sum_{j=d+1}^{d+m} p_j)) \le \alpha/4.$$

Here the last inequality can be hold if we take $Z_{2\alpha}$ to be large enough.

Under the H_1, we use the representation $\boldsymbol{\mathcal{Y}} = \lambda\|\mathbf{w}\|\mathbf{u}_{d+1} \otimes \cdots \otimes \mathbf{u}_{d+m} \otimes \mathbf{w}/\|\mathbf{w}\| + \boldsymbol{\mathcal{E}}$ where $\mathbf{w}$ is presented in (18). Then

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0\right)\\
=&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\sup_{\mathbf{v}_j \in \mathcal{S}_{p_{d+j}-1}, j=1,\ldots,m, \mathbf{v}_{m+1}\in\mathcal{S}_{n-1}} \boldsymbol{\mathcal{Y}} \times \mathbf{v}_1^\top \times \cdots \times_m \mathbf{v}_m^\top \times_{m+1} \mathbf{v}_{m+1}^\top \le Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right)\\
\le&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\boldsymbol{\mathcal{Y}} \times_1 \mathbf{u}_{d+1}^\top \times \cdots \times_m \mathbf{u}_{d+m}^\top \times_{m+1} \mathbf{w}^\top/\|\mathbf{w}\| \le Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right)\\
=&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\lambda\|\mathbf{w}\| + \boldsymbol{\mathcal{E}} \times_1 \mathbf{u}_{d+1}^\top \times \cdots \times_m \mathbf{u}_{d+m}^\top \times_{m+1} \mathbf{w}^\top/\|\mathbf{w}\| \le Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right)\\
\overset{(a)}{\le}&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\lambda\|\mathbf{w}\| \le 2Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right) + \alpha/4\\
\overset{(b)}{\le}&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\lambda\sqrt{n/2} \le 2Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}\right) + 2\exp(-cn) + \alpha/4\\
\overset{(c)}{\le}&0 + 2\exp(-cn) + \alpha/4 \le 3\alpha/4.
\end{aligned}$$

Here (a) is because under H_0, $Z_{2\alpha}$ is chosen such that $|\boldsymbol{\mathcal{E}} \times_1 \mathbf{u}_{d+1}^\top \times \cdots \times_m \mathbf{u}_{d+m}^\top \times_{m+1} \mathbf{w}^\top / \|\mathbf{w}\|| \le Z_{2\alpha}\sqrt{n + \sum_{j=d+1}^{d+m} p_j}$ holds with probability at least $1-\alpha/4$; (b) is because $\|\mathbf{w}\| \ge \sqrt{n/2}$ with probability at least $1 - 2\exp(-cn)$ by Lemma 42; (c) is because $\lambda > 2\sqrt{2} Z_{2\alpha}\sqrt{1 + \sum_{j=d+1}^{d+m} \frac{p_j}{n}}$ by our assumption. The last inequality holds by taking n to be large enough. This finishes the proof of this theorem.

D.3. Proof of Theorem 7

First, with a little abuse of notation, we let the entries of $\boldsymbol{\mathcal{E}}_i$ to be $\boldsymbol{\mathcal{E}}_{i,l}$ for $l = 1, \ldots, (\prod_{j=d+1}^{d+m} p_j)$. Under H_0, we have

$$
\begin{aligned}
\mathbb{P}_0\left(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) &= \mathbb{P}_0\left(\sum_{i=1}^n \|\boldsymbol{\mathcal{Y}}_i\|_{\mathrm{F}}^2 / (n \prod_{j=d+1}^{d+m} p_j) - 1 \ge Z_{3\alpha} / \sqrt{n \prod_{j=d+1}^{d+m} p_j}\right) \\
&= \mathbb{P}_0\left(\sum_{i=1}^n \sum_{l=1}^{\prod_{j=d+1}^{d+m} p_j} (\boldsymbol{\mathcal{E}}_{i,l}^2 - 1) \ge Z_{3\alpha}\sqrt{n \prod_{j=d+1}^{d+m} p_j}\right) \le \alpha/6,
\end{aligned}
$$

where in the last inequality, we use the Bernstein inequality for the sum of sub-exponential random variables and let $Z_{3\alpha}$ to be large enough.

Now let us compute the type II error. Denote $\Delta_1 = \sum_{i=1}^n \langle \boldsymbol{\mathcal{X}}, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d \rangle^2$ and $\Delta_2 = \sum_{i=1}^n \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d \rangle \langle \boldsymbol{\mathcal{E}}_i, \mathbf{u}_{d+1} \otimes \cdots \otimes \mathbf{u}_{d+m} \rangle$.

Let A be the event $A := \{|\Delta_1/n - 1| \le 1/2\}$. By Lemma 42, we know event A holds with probability at least $1 - 2\exp(-cn)$. And we assume $2\exp(-cn)$ is less than $\alpha/6$ by taking large enough n. Then we have

$$
\begin{aligned}
&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\phi_3(\boldsymbol{\mathcal{Y}},\mathscr{X})=0\right)\\
=&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\sum_{i=1}^{n}\|\boldsymbol{\mathcal{Y}}_i\|_{\mathbf{F}}^2/(n\prod_{j=d+1}^{d+m}p_j)-1<Z_{3\alpha}/\sqrt{n\prod_{j=d+1}^{d+m}p_j}\right)\\
=&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{\lambda^2}{n\prod_{j=d+1}^{d+m}p_j}\Delta_1-\frac{Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}<\frac{-2\lambda}{n\prod_{j=d+1}^{d+m}p_j}\Delta_2-\frac{\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)}{n\prod_{j=d+1}^{d+m}p_j}\right)\\
\leq&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{\lambda^2}{2\prod_{j=d+1}^{d+m}p_j}-\frac{Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}<\frac{-2\lambda}{n\prod_{j=d+1}^{d+m}p_j}\Delta_2-\frac{\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)}{n\prod_{j=d+1}^{d+m}p_j},A\right)+\alpha/6\\
\overset{(a)}{\leq}&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{\lambda^2}{4\prod_{j=d+1}^{d+m}p_j}+\frac{Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}<\frac{-2\lambda}{n\prod_{j=d+1}^{d+m}p_j}\Delta_2-\frac{\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)}{n\prod_{j=d+1}^{d+m}p_j},A\right)+\alpha/6\\
\overset{(b)}{\leq}&\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{\lambda^2}{4\prod_{j=d+1}^{d+m}p_j}<\left|\frac{2\lambda}{n\prod_{j=d+1}^{d+m}p_j}\Delta_2\right|,A\right)+\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}<\left|\frac{\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)}{n\prod_{j=d+1}^{d+m}p_j}\right|\right)\\
&+\alpha/6.
\end{aligned}
\tag{19}
$$

Here (a) is because $\frac{\lambda^2}{4\prod_{j=d+1}^{d+m}p_j}\geq\frac{2Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}$ by the assumption on λ; (b) is by the union bound.

Next, we bound the last two probabilities at the end of (19) one by one. First,

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(\frac{Z_{3\alpha}}{\sqrt{n\prod_{j=d+1}^{d+m}p_j}}<\left|\frac{\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)}{n\prod_{j=d+1}^{d+m}p_j}\right|\right)&=\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(Z_{3\alpha}\sqrt{n\prod_{j=d+1}^{d+m}p_j}<\left|\sum_{i=1}^{n}\sum_{l=1}^{\prod_{j=d+1}^{d+m}p_j}(\boldsymbol{\mathcal{E}}_{i,l}^2-1)\right|\right)\\
&\leq\alpha/3
\end{aligned}
\tag{20}
$$

by the choice of $Z_{3\alpha}$ and n discussed in H_0 case.

Second,

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{A}}\left(\frac{\lambda^2}{4\prod_{j=d+1}^{d+m}p_j}<\left|\frac{2\lambda}{n\prod_{j=d+1}^{d+m}p_j}\Delta_2\right|, A\right)\\
&=\mathbb{P}_{\mathcal{A}}\left(\frac{\lambda n}{8}<|\Delta_2|, A\right)\\
&=\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}_{\mathcal{A}}\left(\frac{\lambda n}{8}<|\Delta_2|, A\right)\Big|\{\mathcal{X}_i\}_{i=1}^n\right)\\
&\overset{(a)}{\leq}\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}_{\mathcal{A}}\left(Z_{3\alpha}n^{3/4}(\prod_{j=d+1}^{d+m}p_j)^{1/4}<|\Delta_2|, A\right)\Big|\{\mathcal{X}_i\}_{i=1}^n\right)\\
&\overset{(b)}{\leq}\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}_{\mathcal{A}}\left(Z_{3\alpha}n^{3/4}(\prod_{j=d+1}^{d+m}p_j)^{1/4}\sqrt{2\Delta_1/(3n)}<|\Delta_2|\right)\Big|\{\mathcal{X}_i\}_{i=1}^n\right)\\
&\overset{(c)}{\leq}2\exp\left(-\frac{Z_{3\alpha}^2n^{3/2}(\prod_{j=d+1}^{d+m}p_j)^{1/2}2\Delta_1/(3n)}{2\Delta_1}\right)\leq\alpha/3,
\end{aligned}
\tag{21}
$$

where (a) is by the assumption on λ; (b) is by the property of event A ;(c) is because condition on $\{\mathcal{X}_i\}_{i=1}^n$, we can apply the Hoeffding inequality on the weighted sum of Gaussian random variables $\{\langle\mathcal{E}_i, \mathbf{u}_{d+1}\otimes\cdots\otimes\mathbf{u}_{d+m}\rangle\}_{i=1}^n$ in Δ_2 with weights $\{\langle\mathcal{X}_i, \mathbf{u}_1\otimes\cdots\otimes\mathbf{u}_d\rangle\}_{i=1}^n$, and the last inequality holds if we take n to be large enough.

By plugging (20) and (21) into (19), we prove that the type II error is bounded by $5\alpha/6$ and this finishes the proof of this theorem.

D.4. Proof of Theorem 8

The proof is long and involved. The main idea of the proof is to compute the expectation and variance of the statistic T_4 under both H_0 and H_1 and then use the Chebyshev inequality to bound the type I and type II errors. We decompose the proof into several steps. In Step 1, we compute $\mathbb{E}_0T_4$, $\mathbb{E}_{\mathcal{A}}T_4$, $\mathbb{V}ar_0T_4$ and bound the type I error; in Step 2, we prepare for bounding $\mathbb{V}ar_{\mathcal{A}}T_4$; in step 3, we compute a few key quantities in bounding $\mathbb{V}ar_{\mathcal{A}}T_4$; in the last step, we get the bound for $\mathbb{V}ar_{\mathcal{A}}T_4$ and the type II error. In this proof, we assume $\mathcal{A}=\lambda\mathbf{u}_1\otimes\cdots\otimes\mathbf{u}_{d+m}$.

Recall

$$
\begin{aligned}
&(\mathcal{Y}_i\otimes\mathcal{X}_i)_{[z_1,\ldots,z_{d+m}]}=\mathcal{Y}_{i[z_{d+1},\ldots,z_{d+m}]}\mathcal{X}_{i[z_1,\ldots,z_d]}\\
&\langle\mathcal{Y}_i\otimes\mathcal{X}_i,\mathcal{Y}_j\otimes\mathcal{X}_j\rangle=\sum_{z_1,\ldots,z_{d+m}}\mathcal{Y}_{i[z_{d+1},\ldots,z_{d+m}]}\mathcal{Y}_{j[z_{d+1},\ldots,z_{d+m}]}\mathcal{X}_{i[z_1,\ldots,z_d]}\mathcal{X}_{j[z_1,\ldots,z_d]}.
\end{aligned}
\tag{22}
$$

Step 1. It is easy to check $\mathbb{E}_0 T_4 = 0$ and $\mathbb{E}_{\boldsymbol{\mathcal{A}}} \boldsymbol{\mathcal{Y}}_i \otimes \boldsymbol{\mathcal{X}}_i = \boldsymbol{\mathcal{A}}$. So we have $\mathbb{E}_{\boldsymbol{\mathcal{A}}} T_4 = \lambda^2$. Next

$$\begin{aligned}\mathbb{V}ar_0 T_4 = \mathbb{E}_0 T_4^2 &= \frac{4}{n^2(n-1)^2}\mathbb{E}_0\left(\sum_{i<j}\langle \boldsymbol{\mathcal{Y}}_i \otimes \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_j \otimes \boldsymbol{\mathcal{X}}_j\rangle\right)^2 \\ &= \frac{2}{n(n-1)}\mathbb{E}_0\left(\langle \boldsymbol{\mathcal{Y}}_i \otimes \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_j \otimes \boldsymbol{\mathcal{X}}_j\rangle\right)^2 \\ &= \frac{2\prod_{j=1}^{d+m} p_j}{n(n-1)}.\end{aligned}$$

Then by Chebyshev inequality, we have

$$\mathbb{P}_0\left(\phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) = \mathbb{P}_0\left(T_4 \geq Z_{4\alpha}(\prod_{j=1}^{d+m} p_j)^{1/2}/n\right) \leq \frac{\mathbb{E}_0 T_4^2}{(Z_{4\alpha}(\prod_{j=1}^{d+m} p_j)^{1/2}/n)^2} \leq \alpha/2,$$

where in the last inequality, we take $Z_{4\alpha}$ to be large enough.

Step 2. In this step, we do preparation for computing $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}} T_4$. For simplicity, we introduce a few new notation. Given any $k_z \in [p_{d+z}]$ for $z = 1, \ldots, m$, let $\boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)} = \boldsymbol{\mathcal{A}}_{[:,\ldots,:,k_1,\ldots,k_m]} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$, $\boldsymbol{\mathcal{Y}}_i^{(k_1,\ldots,k_m)} = \boldsymbol{\mathcal{Y}}_{i[k_1,\ldots,k_m]}$ and $\boldsymbol{\mathcal{E}}_i^{(k_1,\ldots,k_m)} = \boldsymbol{\mathcal{E}}_{i[k_1,\ldots,k_m]}$. With this notation, we have

$$\boldsymbol{\mathcal{Y}}_i^{(k_1,\ldots,k_m)} = \langle \boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}, \boldsymbol{\mathcal{X}}_i\rangle + \boldsymbol{\mathcal{E}}_i^{(k_1,\ldots,k_m)}, \quad \text{for } i = 1, \ldots, n.$$

Also we use notation $\boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)}$ and $\boldsymbol{\mathcal{X}}_{i,l}$ to denote the lth entry in the vectorization of $\boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}$ and $\boldsymbol{\mathcal{X}}_i$, respectively.

Denote

$$\begin{aligned}K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j)) &= \sum_{k_1,\ldots,k_m}\langle \boldsymbol{\mathcal{Y}}_i^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}, \boldsymbol{\mathcal{Y}}_j^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{X}}_j - \boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}\rangle \\ \delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i) &= \sum_{k_1,\ldots,k_m}\langle \boldsymbol{\mathcal{Y}}_i^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}, \boldsymbol{\mathcal{A}}^{(k_1,\ldots,k_m)}\rangle.\end{aligned}$$

Notice that $K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j))$ and $\delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i)$ are centered, moreover simple calculation yields

$$\begin{aligned}T_4 &= \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\left(K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j)) + \delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i) + \delta(\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j) + \|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^2\right) \\ &= \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j)) + \frac{2}{n}\sum_{i=1}^n \delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i) + \lambda^2.\end{aligned}$$

Notice that $K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j))$ for different (i, j) pairs are uncorrelated under H_1; similarly $K((\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i), (\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j))$ is uncorrelated with $\delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i)$ and $\delta(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i)$ is uncorrelated with $\delta(\boldsymbol{\mathcal{X}}_j, \boldsymbol{\mathcal{Y}}_j)$ for $i \neq j$. Thus the variance of T_4 has the following formula:

$$\begin{aligned}\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_4) &= \frac{4}{n^2(n-1)^2}\cdot\frac{n(n-1)}{2}\mathbb{E}(K((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1), (\boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{Y}}_2))^2) + \frac{4}{n^2}n\mathbb{E}_{\boldsymbol{\mathcal{A}}}(\delta^2(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1)) \\ &= \frac{2}{n(n-1)}\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1), (\boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{Y}}_2))) + \frac{4}{n}\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1)).\end{aligned} \tag{23}$$

Next, we bound $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1), (\boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{Y}}_2)))$ separately. We will use the notation $\mathbb{E}_{\boldsymbol{\mathcal{X}}}$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{X}}}$ to denote the expectation and variance computed under the marginal distribution of the covariate tensor $\boldsymbol{\mathcal{X}}$ and notation $\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}$ to denote the expectation and variance taken under the conditional distribution of $\boldsymbol{\mathcal{Y}}$ given $\boldsymbol{\mathcal{X}}$ with parameter $\boldsymbol{\mathcal{A}}$. Finally, we will also use the notation p^d to denote $\prod_{j=1}^d p_j$ for simplicity.

The key formula in bounding the variance of $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1), (\boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{Y}}_2)))$ is that for any random variable $f(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}})$,

$$\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(f(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}})) = \mathbb{E}_{\boldsymbol{\mathcal{X}}}(\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(f(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}))) + \mathbb{V}ar_{\boldsymbol{\mathcal{X}}}(\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(f(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}))). \tag{24}$$

Next, we bound $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1), (\boldsymbol{\mathcal{X}}_2, \boldsymbol{\mathcal{Y}}_2)))$ separately in **Step 3.1** and **Step 3.2**.

Step 3.1 (Bound for $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))$). Let δ_{ij} be the Kronecker delta. Then

$$\delta((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1)) = \sum_{k_1,\ldots,k_m} \sum_{j=1}^{p^d} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \left(\boldsymbol{\mathcal{E}}_1^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{X}}_{1,j} + \sum_{l=1}^{p^d} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} (\boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l} - \delta_{jl}) \right). \tag{25}$$

Then we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\delta((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))) &= \sum_{k_1,\ldots,k_m} \sum_{j=1}^{p^d} \sum_{l=1}^{p^d} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} (\boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l} - \delta_{jl}) \\ &= \sum_{j=1}^{p^d} \sum_{l=1}^{p^d} \left(\sum_{k_1,\ldots,k_m} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} \right) (\boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l} - \delta_{jl}) \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\delta((\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{Y}}_1))) =& \mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\sum_{k_1,\ldots,k_m} \sum_{j=1}^{p^d} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{E}}_1^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{X}}_{1,j}) \\ =& \sum_{k_1,\ldots,k_m} \sum_{k_1',\ldots,k_m'} \sum_{j=1}^{p^d} \sum_{l=1}^{p^d} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} \mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l} \boldsymbol{\mathcal{E}}_1^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{E}}_1^{(k_1',\ldots,k_m')}) \\ \overset{(a)}{=}& \sum_{k_1,\ldots,k_m} \sum_{j=1}^{p^d} \sum_{l=1}^{p^d} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l} \\ =& \sum_{j=1}^{p^d} \sum_{l=1}^{p^d} \left(\sum_{k_1,\ldots,k_m} \boldsymbol{\mathcal{A}}_j^{(k_1,\ldots,k_m)} \boldsymbol{\mathcal{A}}_l^{(k_1,\ldots,k_m)} \right) \boldsymbol{\mathcal{X}}_{1,j} \boldsymbol{\mathcal{X}}_{1,l}, \end{aligned}$$

where in (a) the nonzero part appears when $(k_1, \ldots, k_m) = (k_1', \ldots, k_m')$.

Thus

$$
\begin{aligned}
&\mathbb{V}ar_{\boldsymbol{\mathcal{X}}}\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\delta((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1)))\\
=&\sum_{j=1}^{p^d}\sum_{l=1}^{p^d}\sum_{r=1}^{p^d}\sum_{s=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_l^{(k_1,\dots,k_m)}\right)\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}\right)\\
&\times\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}-\delta_{jl})(\boldsymbol{\mathcal{X}}_{1,r}\boldsymbol{\mathcal{X}}_{1,s}-\delta_{rs})\\
\overset{(a)}{\le}&2\sum_{j=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)^2+2\sum_{1\le j,r\le p^d,j\neq r}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}\right)^2\\
\overset{(b)}{\le}&2\left(\sum_{j=1}^{p^d}\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)^2+2\sum_{1\le j,r\le p^d,j\neq r}\Big(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\Big)\Big(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)2}\Big)\\
\le&2\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4+2\left(\sum_{j=1}^{p^d}\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)^2=4\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4,
\end{aligned}
\tag{26}
$$

where in (a), we use the fact

$$
\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}-\delta_{jl})(\boldsymbol{\mathcal{X}}_{1,r}\boldsymbol{\mathcal{X}}_{1,s}-\delta_{rs})=\begin{cases}2, & \text{if } j=l=r=s\\ 1, & \text{if } j=r\neq l=s \text{ or } j=s\neq l=r\\ 0, & \text{otherwise.}\end{cases}
\tag{27}
$$

In (b) we use the Cauchy-Schwarz inequality.

Finally,

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\mathcal{X}}}\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\delta((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1)))&=\mathbb{E}_{\boldsymbol{\mathcal{X}}}\sum_{j=1}^{p^d}\sum_{l=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_l^{(k_1,\dots,k_m)}\right)\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}\\
&\overset{(a)}{=}\sum_{j=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)=\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^2.
\end{aligned}
\tag{28}
$$

Here (a) is because the nonzero part appears when $j=l$.

Based on (26), (28) and (24), we have

$$
\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\delta((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1)))\le 4\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4+\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^2=4\lambda^4+\lambda^2.
\tag{29}
$$

Step 3.2 (Bound for $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2)))$). First, we have the following decomposition for $K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2))$ under H_1,

$$
\begin{aligned}
&K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2)) \\
&= \sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\left(\boldsymbol{\mathcal{E}}_1^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{1,j} + \sum_{r=1}^{p^d}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})\right) \\
&\quad\times\left(\boldsymbol{\mathcal{E}}_2^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{2,j} + \sum_{s=1}^{p^d}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})\right) \\
&= \sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\boldsymbol{\mathcal{E}}_1^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{E}}_2^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{2,j} + \sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\boldsymbol{\mathcal{E}}_1^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj}) \\
&\quad+\sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})\boldsymbol{\mathcal{E}}_2^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{X}}_{2,j} \\
&\quad+\sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\left(\sum_{r=1}^{p^d}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})\right)\left(\sum_{s=1}^{p^d}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})\right).
\end{aligned}
\tag{30}
$$

Thus we have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2))) \\
&= \sum_{k_1,\dots,k_m}\sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\sum_{s=1}^{p^d}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj}) \\
&= \sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\sum_{s=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}\right)(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})
\end{aligned}
$$

By noticing that each summation term in the last equality of (30) are uncorrelated to each other given $\{\boldsymbol{\mathcal{X}}_i\}_{i=1}^n$, we have

$$
\begin{aligned}
&\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2)))\\
=&\sum_{k_1,\ldots,k_m}\sum_{k_1',\ldots,k_m'}\sum_{j=1}^{p^d}\sum_{l=1}^{p^d}\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{E}}_1^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{E}}_2^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{E}}_1^{(k_1',\ldots,k_m')}\boldsymbol{\mathcal{E}}_2^{(k_1',\ldots,k_m')}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{X}}_{2,l})\\
&+\sum_{k_1,\ldots,k_m}\sum_{k_1',\ldots,k_m'}\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\sum_{l=1}^{p^d}\sum_{r=1}^{p^d}\\
&\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{E}}_1^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{E}}_1^{(k_1',\ldots,k_m')}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1',\ldots,k_m')}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})(\boldsymbol{\mathcal{X}}_{2,l}\boldsymbol{\mathcal{X}}_{2,r}-\delta_{rl}))\\
&+\sum_{k_1,\ldots,k_m}\sum_{k_1',\ldots,k_m'}\sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\sum_{l=1}^{p^d}\sum_{s=1}^{p^d}\\
&\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{E}}_2^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{E}}_2^{(k_1',\ldots,k_m')}\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,l}\boldsymbol{\mathcal{A}}_s^{(k_1',\ldots,k_m')}\boldsymbol{\mathcal{A}}_r^{(k_1,\ldots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})(\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{X}}_{1,s}-\delta_{sl}))\\
\overset{(a)}{=}&\sum_{k_1,\ldots,k_m}\sum_{j=1}^{p^d}\sum_{l=1}^{p^d}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{X}}_{2,l}\\
&+\sum_{k_1,\ldots,k_m}\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\sum_{l=1}^{p^d}\sum_{r=1}^{p^d}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1,\ldots,k_m)}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})(\boldsymbol{\mathcal{X}}_{2,l}\boldsymbol{\mathcal{X}}_{2,r}-\delta_{rl}))\\
&+\sum_{k_1,\ldots,k_m}\sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\sum_{l=1}^{p^d}\sum_{s=1}^{p^d}\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,l}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1,\ldots,k_m)}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})(\boldsymbol{\mathcal{X}}_{1,l}\boldsymbol{\mathcal{X}}_{1,s}-\delta_{sl})),
\end{aligned}
$$

where in (a) the nonzero part appears when $(k_1,\ldots,k_m)=(k_1',\ldots,k_m')$.

Thus, we have

$$
\begin{aligned}
&\mathbb{V}ar_{\boldsymbol{\mathcal{X}}}(\mathbb{E}^{\boldsymbol{\mathcal{X}}}_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2))))\\
=&\sum_{j=1}^{p^d}\sum_{r=1}^{p^d}\sum_{s=1}^{p^d}\sum_{l=1}^{p^d}\sum_{u=1}^{p^d}\sum_{v=1}^{p^d}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}\right)\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_u^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_v^{(k_1,\dots,k_m)}\right)\\
&\times\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,r}-\delta_{rj})(\boldsymbol{\mathcal{X}}_{1,u}\boldsymbol{\mathcal{X}}_{1,l}-\delta_{ul})\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{2,v}\boldsymbol{\mathcal{X}}_{2,l}-\delta_{vl})(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})\\
\overset{(27)}{\le}&\sum_{j=1}^{p^d}4\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)^2+2\sum_{j=u=l=r\neq s=v}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}\right)^2\\
&+\sum_{s=r=l\neq u=j=v}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)2}\right)\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_u^{(k_1,\dots,k_m)2}\right)\\
&+\sum_{j=l\neq r=u,s=v\neq l}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)}\right)^2\\
\overset{(a)}{\le}&4\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4+3\sum_{1\le j,s\le p^d,j\neq s}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2}\right)\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)2}\right)\\
&+2\sum_{j=1}^{p^d}\sum_{s\neq j,r\neq j}\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)2}\right)\left(\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_s^{(k_1,\dots,k_m)2}\right)\\
\le&4\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4+3(\sum_{j=1}^{p^d}\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_j^{(k_1,\dots,k_m)2})^2+p^d(\sum_{r=1}^{p^d}\sum_{k_1,\dots,k_m}\boldsymbol{\mathcal{A}}_r^{(k_1,\dots,k_m)2})^2\\
\le&3(\prod_{j=1}^d p_j)\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^4=3(\prod_{j=1}^d p_j)\lambda^4.
\end{aligned}
\tag{31}
$$

Here in (a) we apply the Cauchy-Schwarz inequality to the second and fourth summation terms and combine the second summation term with the third summation term.

Moreover

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2))))\\
=&(\prod_{j=d+1}^{d+m} p_j)\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\sum_{j=1}^{p^d}\sum_{l=1}^{p^d}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{2,l})\\
&+2\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\sum_{l=1}^{p^d}\sum_{r=1}^{p^d}\left(\sum_{k_1,\ldots,k_m}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1,\ldots,k_m)}\right)\mathbb{E}_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{X}}_{1,j}\boldsymbol{\mathcal{X}}_{1,l}(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})(\boldsymbol{\mathcal{X}}_{2,l}\boldsymbol{\mathcal{X}}_{2,r}-\delta_{rl}))\\
\overset{(a)}{=}&\prod_{j=1}^{d+m}p_j+2\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\sum_{r=1}^{p^d}\left(\sum_{k_1,\ldots,k_m}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)}\boldsymbol{\mathcal{A}}_r^{(k_1,\ldots,k_m)}\right)\mathbb{E}_{\boldsymbol{\mathcal{X}}}((\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,r}-\delta_{rj}))\\
\overset{(b)}{\le}&\prod_{j=1}^{d+m}p_j+4\sum_{j=1}^{p^d}\sum_{s=1}^{p^d}\left(\sum_{k_1,\ldots,k_m}\boldsymbol{\mathcal{A}}_s^{(k_1,\ldots,k_m)2}\right)\\
=&\prod_{j=1}^{d+m}p_j+4(\prod_{k=1}^{d}p_k)\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}^2=\prod_{j=1}^{d+m}p_j+4(\prod_{k=1}^{d}p_k)\lambda^2.
\end{aligned}
\tag{32}
$$

Here (a) is because the nonzero part appears when $j = l$ and (b) is because the nonzero part appears when $s = r$ and $\mathbb{E}_{\boldsymbol{\mathcal{X}}}((\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})(\boldsymbol{\mathcal{X}}_{2,j}\boldsymbol{\mathcal{X}}_{2,s}-\delta_{sj})) \le 2$.

Combining (31), (32) and (24), we have

$$
\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(K((\boldsymbol{\mathcal{X}}_1,\boldsymbol{\mathcal{Y}}_1),(\boldsymbol{\mathcal{X}}_2,\boldsymbol{\mathcal{Y}}_2)))\le 3(\prod_{j=1}^{d}p_j)\lambda^4+\prod_{j=1}^{d+m}p_j+4(\prod_{k=1}^{d}p_k)\lambda^2. \tag{33}
$$

Step 4. By plugging (29) and (33) into (23), we have

$$
\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_4)\le\frac{2}{n(n-1)}(3(\prod_{j=1}^{d}p_j)\lambda^4+\prod_{j=1}^{d+m}p_j+4(\prod_{k=1}^{d}p_k)\lambda^2)+\frac{4}{n}(4\lambda^4+\lambda^2). \tag{34}
$$

By Chebyshev inequality, we have

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_4(\boldsymbol{\mathcal{Y}},\mathscr{A})=0)=&\,\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(T_4<Z_{4\alpha}(\prod_{j=1}^{d+m}p_j)^{1/2}/n\right)\\
\le&\,\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(T_4-\mathbb{E}_{\boldsymbol{\mathcal{A}}}T_4<Z_{4\alpha}(\prod_{j=1}^{d+m}p_j)^{1/2}/n-\mathbb{E}_{\boldsymbol{\mathcal{A}}}T_4\right)\\
\overset{(a)}{\le}&\,\mathbb{P}_{\boldsymbol{\mathcal{A}}}\left(T_4-\lambda^2<-\lambda^2/2\right)\\
\le&\,\frac{\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_4)}{\lambda^4/4}\overset{(b)}{\le}\alpha/2.
\end{aligned}
$$

Here (a) is by the assumption on λ and (b) is by (34) and the assumptions on λ and $n \ge C_\alpha(\prod_{k=1}^{d} p_k)^{1/2}$. This finishes the proof of this theorem.

D.5. Proofs for Corollary 9 - Corollary 14

The proofs for these corollaries are similar. Here we provide the proof for Corollary 14 and omit others for simplicity.

Let us first consider the guarantee for $\phi_1 \vee \phi_2$. First,

$$\mathbb{P}_0\left(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) \leq \mathbb{P}_0\left(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) + \mathbb{P}_0\left(\phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) \leq \alpha/2$$

for proper choices of C'_α as we have shown Theorems 5 and 6.

Under H_1, when $n \geq C_{\alpha,d,m}(\sum_{j=1}^{d+m} p_j)$, by Theorem 5, we have if λ satisfies the condition in the statement, then

$$\mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \alpha/2.$$

When $n \leq C_{\alpha,d,m}(\sum_{j=1}^{d+m} p_j)$, then we have $\sum_{j=1}^{d+m} p_j/n \geq c_{\alpha,d,m}$. So by picking C'_α large enough. The assumption on λ implies that the requirement of λ in Theorem 6 holds. So we have

$$\mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_1(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_2(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \alpha/2.$$

This finishes the proof for the statistical upper bound part.

Now let us consider the guarantee for $\phi_3 \vee \phi_4$. Under H_0, we still have

$$\mathbb{P}_0\left(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) \leq \mathbb{P}_0\left(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) + \mathbb{P}_0\left(\phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 1\right) \leq \alpha/2$$

for proper choices of C'_α as we have shown Theorems 7 and 8.

Under H_1, when $n \geq C_\alpha \prod_{j=d+1}^{d+m} p_j$, we have $\sqrt{(\prod_{j=1}^{d+m} p_j)^{1/2}/n} \leq c_\alpha((\prod_{j=d+1}^{d+m} p_j)/n)^{1/4}$, by Theorem 8, we have if $\lambda \geq C'_\alpha\sqrt{(\prod_{j=1}^{d+m} p_j)^{1/2}/n}$, we have

$$\mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \alpha/2.$$

When $n \leq C_\alpha \prod_{j=d+1}^{d+m} p_j$, then we have $\sqrt{(\prod_{j=1}^{d+m} p_j)^{1/2}/n} \geq c_\alpha((\prod_{j=d+1}^{d+m} p_j)/n)^{1/4}$. So when $\lambda \geq C'_\alpha((\prod_{j=d+1}^{d+m} p_j)/n)^{1/4}$, by Theorem 7, we have

$$\mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) \vee \phi_4(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \mathbb{P}_{\boldsymbol{\mathcal{A}}}(\phi_3(\boldsymbol{\mathcal{Y}}, \mathscr{X}) = 0) \leq \alpha/2.$$

This finishes the proof for the computational upper bound part.

Appendix E. Proofs in Section 5

We will follow the convention introduced at the beginning of Appendix C and consider the special setting $p_1 = p_2 = \cdots = p_{d+m}$ and $\sigma = 1$ without loss of generality.

E.1. Proof of Theorem 19

Here we provide a unified proof for $m = 0$, $m = 1$ and $m \geq 2$ together, although we note the truncation strategy we introduced below is actually not needed in the $m = 0$ case.

We consider a subclass of parameters $\mathscr{A} = \{\boldsymbol{\mathcal{A}} = \lambda \mathbf{u}^{d+m}, \mathbf{u} \in \mathcal{S}_{p-1}\}$. The prior we put on the alternative is $\mathbf{u} \sim \text{Unif}(\mathcal{S}_{p-1})$. Recall we use notation $D_{\mathbf{u}}$ to denote the joint distribution of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ with tensor parameter $\mathscr{A} = \lambda \mathbf{u}^{\otimes(d+m)}$; and notation D_0 to denote the distribution of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ under H_0. Based on the definition of the statistical dimension, it is critical to understand the behavior of $|\chi_{D_0}(D_{\mathbf{v}}, D_{\mathbf{u}})|$. By Lemma 38 Eq. (49), we can see this term can easily blow up when $\lambda \gg 1$, which will happen in $m = 1$ or $m \geq 2$. Thus, here we need to perform truncation under H_1. Given $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i) \sim D_{\mathbf{u}}$, we keep the data if $|\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle| \leq M := C\sqrt{\log p}$ for sufficiently large C and resample otherwise. Let C_* be the normalization factor $1/C_* := \mathbb{P}\left(|\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle| \leq M\right)$. Notice that $C_* \approx 1$ for large enough C. Let us denote this truncated distribution as $\widetilde{D}_{\mathbf{u}}$. It is easy to check $\text{TV}(D_{\mathbf{u}}, \widetilde{D}_{\mathbf{u}}) = p^{-C}$, which goes to zero as $p \to \infty$. If we denote the risk for this modified hypothesis testing problem

$$H_0 : (\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i) \sim D_0 \quad \text{v.s. } H_1 : (\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i) \sim \widetilde{D}_{\mathbf{u}} \text{ for } \mathbf{u} \sim \text{Unif}(\mathcal{S}_{p-1}) \tag{35}$$

as R', then we have for any test ϕ, $R(\phi) \geq R'(\phi) + o(1)$ as $p \to \infty$. Thus to show the SQ lower bound for the original hypothesis testing problem, it is enough to show the same lower bound hold for the hypothesis testing problem (35).

We divide the rest of the proof into two steps. In step 1, we provide a good bound for $|\chi_{D_0}(\widetilde{D}_{\mathbf{v}}, \widetilde{D}_{\mathbf{u}})|$ based on different values of $|\mathbf{u}^\top \mathbf{v}|$ and in the second step we bound the statistical dimension.

Step 1. Given any $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{p-1}, \mathbf{u} \neq \mathbf{v}$, let $t := \mathbf{u}^\top \mathbf{v}$. If $|\lambda^2 t^m| \leq 1/4$, then we can easily check $|(1 - t^{2d})\lambda^2 t^m + t^d| \leq (1 - |t^d|^2)/4 + |t^d| < 1$ as $|t^d| < 1$, and we have $|\lambda^2 t^{d+m}| \leq 1/4$. Then by Lemma 38 Eq. (49), we have

$$\begin{aligned}
|\chi_{D_0}(\widetilde{D}_{\mathbf{v}}, \widetilde{D}_{\mathbf{u}})| &\leq C_*^2 |\chi_{D_0}(D_{\mathbf{v}}, D_{\mathbf{u}})| \\
&\leq C_*^2 \left| \left((1 - t^{d+m}\lambda^2)^2 - \lambda^4 t^{2m} \right)^{-1/2} - 1 \right| \\
&\overset{\text{Lemma 41}}{\leq} C_*^2 |1 - c(t^{2(d+m)}\lambda^4 - 2t^{d+m}\lambda^2 - \lambda^4 t^{2m}) - 1| \\
&\leq C \left| t^{d+m}\lambda^2 + \lambda^4 t^{2m} \right|.
\end{aligned}$$

At the same time, for any $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{p-1}, \mathbf{u} \neq \mathbf{v}$, we also have the following bound by truncation:

$$\begin{aligned}
|\chi_{D_0}(\widetilde{D}_{\mathbf{v}}, \widetilde{D}_{\mathbf{u}})| \overset{\text{Lemma 38}}{=} & C_*^2 \mathbb{E}_{\boldsymbol{\mathcal{X}}_i : |\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle| \leq M, |\langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle| \leq M} \left(\exp(\lambda^2 t^m \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle \langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle) \right) \\
\leq & C_*^2 \exp\left(\lambda^2 M^2 |t|^m\right).
\end{aligned}$$

In summary, we have

$$|\chi_{D_0}(\widetilde{D}_{\mathbf{v}}, \widetilde{D}_{\mathbf{u}})| \leq \begin{cases} C|t^{d+m}\lambda^2 + \lambda^4 t^{2m}|, & \text{if } |\lambda^2 t^m| \leq 1/4; \\ C_*^2 \exp\left(\lambda^2 M^2 |t|^m\right), & \text{otherwise} . \end{cases} \tag{36}$$

Step 2. By the definition of the statistical dimension, we find given any positive integer q, and event A such that $P_{\mathbf{u},\mathbf{v} \sim \text{Unif}(\mathcal{S}_{p-1})}(A) \geq 1/q^2$, to maximize $\mathbb{E}_{\mathbf{u},\mathbf{v} \sim \text{Unif}(\mathcal{S}_{p-1})} \left(|\chi_{D_0}(\widetilde{D}_{\mathbf{v}}, \widetilde{D}_{\mathbf{u}})| | A \right)$:

(1) incorperating positive t is better than negative t where $t = \mathbf{u}^\top \mathbf{v}$; (2) for positive t, the upper bound of $|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|$ is an increasing function with respect to t, i.e., choosing large positive t is better than small positive t.

Given any $0 < \epsilon < 1/4$, let the event $A^* := \{\mathbf{u}, \mathbf{v} \in \mathcal{S}_{p-1} : \mathbf{u}^\top \mathbf{v} > p^{-1/2+\epsilon}\}$. First by Lemma 46, we known $(\mathbf{u}^\top \mathbf{v} + 1)/2 \sim \text{Beta}\left(\frac{p-1}{2}, \frac{p-1}{2}\right)$. Then by Zhang and Zhou (2020) Theorem 8, we have there exist $C' > c' > 0$ such that

$$\exp(-C' p^{2\epsilon}) \leq P_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}(A^*) \leq \exp(-c' p^{2\epsilon}). \tag{37}$$

We set $1/q^2 = P_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}(A^*)$, i.e., $q \geq \exp(-c' p^{2\epsilon}/2)$. If we can show

$$\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|\Big|A^*\right] \leq \frac{1}{n}, \tag{38}$$

then we are done as for any other event A with $P_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}(A) \geq 1/q^2$, we will have

$$\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|\Big|A^*\right] \geq \mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|\Big|A\right]$$

for the reason we have mentioned before and this implies the statistical dimension is at least q which is greater or equal to $\exp(-c' p^{2\epsilon}/2)$.

Next, we show (38). Let us define two more events

$$A^{**} := \{\mathbf{u}, \mathbf{v} | \mathbf{u}^\top \mathbf{v} > p^{-1/2+2\epsilon}\} \quad \text{and} \quad A^{***} := \{\mathbf{u}, \mathbf{v} | p^{-1/2+\epsilon} < \mathbf{u}^\top \mathbf{v} < p^{-1/2+2\epsilon}\}.$$

Then

$$\begin{aligned}
&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|\Big|A^*\right] \\
&= \mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})| 1_{A^*}\right] / P_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}(A^*) \\
&= \mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})|(1_{A^{**}} + 1_{A^{***}})\right] / P_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}(A^*).
\end{aligned} \tag{39}$$

Next, we bound $\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})| 1_{A^{**}}\right]$ and $\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_\mathbf{v}, \widetilde{D}_\mathbf{u})| 1_{A^{***}}\right]$ separately below.

First,

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_{\mathbf{v}},\widetilde{D}_{\mathbf{u}})|1_{A^{**}}\right]\\
\overset{(36)}{\leq}&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}C_*^2\left[\exp(\lambda^2|t|^m M^2)1_{A^{**}}\right]\\
\overset{(a)}{=}&C\int_{p^{-1+4\epsilon}}^1 \exp(\lambda^2 b^{m/2}M^2)\frac{b^{-1/2}(1-b)^{(p-3)/2}\Gamma(p/2)}{\Gamma(1/2)\Gamma((p-1)/2)}db\\
\overset{(b)}{\leq}&Cp^{1-2\epsilon}\int_{p^{-1+4\epsilon}}^1 \exp(\lambda^2 b^{m/2}M^2)(1-b)^{(p-3)/2}db\\
=&Cp^{1-2\epsilon}\int_{p^{-1+4\epsilon}}^1 \exp(\lambda^2 b^{m/2}M^2-\frac{p-3}{2}\log\frac{1}{1-b})db\\
\overset{(c)}{\leq}&Cp^{1-2\epsilon}\int_{p^{-1+4\epsilon}}^1 \exp(\lambda^2 b^{m/2}M^2-\frac{p-3}{2}b)db\\
\overset{(d)}{\leq}&\begin{cases} Cp^{1-2\epsilon}\exp(\lambda^2M^2)\int_{p^{-1+4\epsilon}}^1\exp(-\frac{p-3}{2}b)db, & \text{if } m=0\\ Cp^{1-2\epsilon}\int_{p^{-1+4\epsilon}}^1\exp(\lambda^2 b^{1/2}M^2-cpb)db, & \text{if } m=1\\ Cp^{1-2\epsilon}\int_{p^{-1+4\epsilon}}^1\exp(\lambda^2 bM^2-cpb)db, & \text{if } m\geq 2\end{cases}\\
\overset{(e)}{\leq}&Cp^{1-2\epsilon}p^C\int_{p^{-1+4\epsilon}}^1\exp(-c'pb)db\\
\leq&Cp^{C+1-2\epsilon}\exp(\lambda^2M^2)\left(\frac{\exp(-c'p)}{-c'p}-\frac{\exp(-c'p\cdot p^{-1+4\epsilon})}{-c'p}\right)\overset{(f)}{\leq}\exp(-cp^{4\epsilon}).
\end{aligned}
\tag{40}
$$

here (a) is because $B:=(\mathbf{u}^\top\mathbf{v})^2\sim \text{Beta}(\frac{1}{2},\frac{p-1}{2})$ by Lemma 46 and we change the integration variable to B; (b) is because $\Gamma(p/2)/(\Gamma(1/2)\Gamma((p-1)/2))\leq C\sqrt{p}$ by the property of Gamma function and we bound $b^{-1/2}$ by $p^{1/2-2\epsilon}$; (c) is because $\log\frac{1}{1-b}\geq b$; (d) is because $b\in(0,1)$; (e) is because

- (for $m=1$) $\frac{\lambda^2 b^{1/2}M^2}{cpb}=\frac{\lambda^2M^2}{cpb^{1/2}}\leq\frac{\lambda^2M^2}{cp\cdot p^{-1/2+2\epsilon}}=\frac{\lambda^2M^2}{p^{1/2+2\epsilon}}=o(1)$ as $\lambda\leq c_1p^{1/4-\epsilon}$,
- (for $m\geq 2$) $\frac{\lambda^2 bM^2}{cpb}=\frac{\lambda^2M^2}{cp}=o(1)$ as $\lambda\leq c_1p^{1/2-2\epsilon}$;

and (f) holds when p is sufficiently large.

Next, on event A^{***}, let us first check $|\lambda^2t^m|\leq 1/4$ holds in all $m=0$, $m=1$ and $m\geq 2$ cases. First, it is true in $m=0$ case as $\lambda\leq 1/2$. When $m=1$, $|\lambda^2 t|\leq c_1^2p^{1/2-2\epsilon}\cdot p^{-1/2+2\epsilon}\leq c_1\leq 1/4$ as long as $c_1\leq 1/2$. Finally for $m\geq 2$, $|\lambda^2t^m|\leq|\lambda^2t^2|\leq c_1^2p^{1-4\epsilon}\cdot p^{-1+4\epsilon}\leq 1/4$ as long as $c_1\leq 1/2$.

Thus on event A^{***}, we can bound $|\chi_{D_0}(\widetilde{D}_{\mathbf{v}},\widetilde{D}_{\mathbf{u}})|$ via $C|t^{d+m}\lambda^2+\lambda^4t^{2m}|$ by (36). Thus

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_{\mathbf{v}},\widetilde{D}_{\mathbf{u}})|1_{A^{***}}\right]\\
\leq&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}\left[C|t^{d+m}\lambda^2+\lambda^4t^{2m}|1_{A^{***}}\right]\\
\leq&C(p^{(-1/2+2\epsilon)(d+m)}\lambda^2+\lambda^4p^{2m(-1/2+2\epsilon)})\mathbb{P}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}(A^{***})\\
\leq&C(p^{(-1/2+2\epsilon)(d+m)}\lambda^2+\lambda^4p^{2m(-1/2+2\epsilon)})\mathbb{P}_{\mathbf{u},\mathbf{v}\sim \text{Unif}(\mathcal{S}_{p-1})}(A^{*}).
\end{aligned}
\tag{41}
$$

By plugging (40) and (41) into (39), we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{u},\mathbf{v}\sim\text{Unif}(\mathcal{S}_{p-1})}\left[|\chi_{D_0}(\widetilde{D}_{\mathbf{v}},\widetilde{D}_{\mathbf{u}})|\Big|A^*\right]\\ &\overset{(37)}{\leq} C\cdot\exp(-cp^{4\epsilon})\cdot\exp(C'p^{2\epsilon})+C(p^{(-1/2+2\epsilon)(d+m)}\lambda^2+\lambda^4p^{2m(-1/2+2\epsilon)})\leq\frac{1}{n},\end{aligned}$$

where the last inequality holds as p is sufficiently large and $n \leq c_2(\lambda^2 p^{(1/2-2\epsilon)(d+m)} + \lambda^4 p^{(1-4\epsilon)m})$ for some sufficiently small c_2. This finishes the proof of this theorem.

E.2. Proof of Theorem 23

One key property regarding the SQ algorithm we will use is the following one.

Lemma 37 (Theorem 1.4 of Feldman (2017a)) *Let D be a distribution over the domain X. There exists a statistical query algorithm that given $n, \xi, B > 0$, $\varphi : X \to \mathbb{R}$ satisfying $\mathbb{E}_{x\sim D}[\varphi^2(x)] \leq B$ and access to $\text{VSTAT}(n)$, outputs an estimate of $\mathbb{E}_{x\sim D}[\varphi(x)]$, denoted by $\hat{\mathbb{E}}_{x\sim D}[\varphi(x)]$, such that*

$$\left|\hat{\mathbb{E}}_{x\sim D}[\varphi(x)]-\mathbb{E}_{x\sim D}[\varphi(x)]\right|\leq 8\log(n)\sqrt{\frac{\mathbb{V}ar_{x\sim D}(\varphi(x))}{n}}+\xi$$

after making at most $3\log(4nB/\xi^2)$ queries.

Recall we use the following two critical statistics to design the test:

- $T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) = \sum_{j_1,\ldots,j_m=1}^{p} \boldsymbol{\mathcal{Y}}_{1[j_1,\ldots,j_m]}^2$;
-

$$\begin{aligned}&T_6(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1)\\ &=\sum_{j_1,\ldots,j_{\frac{d+m}{2}}=1}^{p}(\boldsymbol{\mathcal{Y}}_1\otimes\boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,j_2,j_2,\ldots,j_{\frac{d+m}{2}},j_{\frac{d+m}{2}}]}\\ &=\left(\sum_{j_{(d+1)/2},\ldots,j_{(d+m)/2}=1}^{p}\boldsymbol{\mathcal{Y}}_{1[j_{(d+1)/2},j_{(d+1)/2},\ldots,j_{(d+m)/2},j_{(d+m)/2}]}\right)\cdot\left(\sum_{j_1,\ldots,j_{d/2}=1}^{p}\boldsymbol{\mathcal{X}}_{1[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}\right)\end{aligned}$$

Under H_0, we have

$$\begin{aligned}&\mathbb{E}_0(T_5(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))=p^m,\quad \mathbb{V}ar_0(T_5(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))=2p^m,\quad \mathbb{E}_0(\varphi_1^2(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))\leq 3p^{2m};\\ &\mathbb{E}_0(T_6(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))=0,\quad \mathbb{V}ar_0(T_6(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))=p^{(d+m)/2},\quad \mathbb{E}_0(\varphi_2^2(\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{X}}_1))=p^{d+m}.\end{aligned}$$

Under H_1, we have $\boldsymbol{\mathcal{Y}}_1 = \lambda\langle\boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle\mathbf{u}^{\otimes m} + \boldsymbol{\mathcal{E}}_1$. We use the notation $\mathbb{E}_{\boldsymbol{\mathcal{X}}_1}$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{X}}_1}$ to denote the expectation and variance computed under the marginal distribution of the covariate tensor $\boldsymbol{\mathcal{X}}_1$ and notation $\mathbb{E}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}_1}$ and $\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{\mathcal{X}}_1}$ to denote the expectation and variance taken under the conditional distribution of $\boldsymbol{\mathcal{Y}}_1$ given $\boldsymbol{\mathcal{X}}_1$ with parameter $\boldsymbol{\mathcal{A}}$.

Since $T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)|\boldsymbol{\mathcal{X}}_1 \sim \chi^2_{p^m}(\lambda^2\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle^2)$, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) &= \mathbb{E}_{\boldsymbol{\mathcal{X}}_1}(p^m + \lambda^2\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle^2) = \lambda^2 + p^m;\\
\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) &= \mathbb{E}_{\boldsymbol{\mathcal{X}}_1}(\mathbb{V}ar^{\boldsymbol{\mathcal{X}}_1}_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1))) + \mathbb{V}ar_{\boldsymbol{\mathcal{X}}_1}(\mathbb{E}^{\boldsymbol{\mathcal{X}}_1}_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))\\
&= \mathbb{E}_{\boldsymbol{\mathcal{X}}_1}(2(p^m + 2\lambda^2\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle^2)) + \mathbb{V}ar_{\boldsymbol{\mathcal{X}}_1}(p^m + \lambda^2\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle^2)\\
&= 2p^m + 4\lambda^2 + 2\lambda^4;\\
\mathbb{E}_{\boldsymbol{\mathcal{A}}}(\varphi_1^2(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) &\le p^{2m} + 2p^m\lambda^2 + 2p^m + 4\lambda^2 + 3\lambda^4 \le 3p^{2m} + 2p^m\lambda^2 + 4\lambda^2 + 3\lambda^4.
\end{aligned}$$

Moreover, under H_1, denote $\xi_1 = \sum_{j_1,\dots,j_{m/2}=1}^{p} \boldsymbol{\mathcal{Y}}_{1[j_1,j_1,\dots,j_{m/2},j_{m/2}]}$ and $\xi_2 = \sum_{j_1,\dots,j_{d/2}=1}^{p} \boldsymbol{\mathcal{X}}_{1[j_1,j_1,\dots,j_{d/2},j_{d/2}]}$, then $T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) = \xi_1\xi_2$. $\xi_1|\boldsymbol{\mathcal{X}}_1 \sim N(\lambda\langle \boldsymbol{\mathcal{X}}_1, \mathbf{u}^{\otimes d}\rangle, p^{m/2})$, so $\xi_1 \sim N(0, \lambda^2 + p^{m/2})$. In addition $\xi_2 \sim N(0, p^{d/2})$ and

$$\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) = \mathbb{E}_{\boldsymbol{\mathcal{A}}}(\xi_1\xi_2) = \lambda,$$

thus ξ_1, ξ_2 follow the joint normal distribution $(\xi_1, \xi_2)^\top \sim N\left(0, \begin{bmatrix} \lambda^2 + p^{m/2} & \lambda \\ \lambda & p^{d/2} \end{bmatrix}\right)$. So we can write $\xi_1 = \lambda\xi_2/p^{d/2} + Z$ where $Z \sim N(0, p^{m/2} + \lambda^2 - \lambda^2/p^{d/2})$. Thus

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mathcal{A}}}(\varphi_6^2(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) &= \mathbb{E}(\xi_1^2\xi_2^2) = \mathbb{E}((\lambda\xi_2/p^{d/2} + Z)^2\xi_2^2) = 2\lambda^2 + p^{d/2}\lambda^2 + p^{(d+m)/2},\\
\mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(\varphi_6^2(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) &= \mathbb{E}_{\boldsymbol{\mathcal{A}}}(\varphi_6^2(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) - (\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))^2 = \lambda^2 + p^{d/2}\lambda^2 + p^{(d+m)/2}.
\end{aligned}$$

Take $B := \max(2\lambda^2 + p^{d/2}\lambda^2 + p^{(d+m)/2}, 3p^{2m} + 2p^m\lambda^2 + 4\lambda^2 + 3\lambda^4)$, by Lemma 37, we have if

$$n \ge C\frac{\max(\mathbb{V}ar_0(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)), \mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))}{(\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) - \mathbb{E}_0(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))^2}\log^2 n,$$

i.e., $n \ge C\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4}\log^2 n$, and set $\xi = (\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) - \mathbb{E}_0(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))/4$, we have the test $\phi_5 = 1(T_5(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) \ge p^m + \lambda^2/2)$ can detect by calling $O(\log(nB/\lambda^4))$ number of queries to VSTAT(n). Similarly, if

$$n \ge C\frac{\max(\mathbb{V}ar_0(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)), \mathbb{V}ar_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))}{(\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) - \mathbb{E}_0(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))^2}\log^2 n, ,$$

i.e., $n \ge C\frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2}\log^2 n$, and set $\xi = (\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) - \mathbb{E}_0(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)))/4$, we have the test $\phi_6 = 1(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1) \ge \lambda/2)$ can detect by calling $O(\log(nB/\lambda^2))$ number of queries to VSTAT(n).

In summary, if $n \ge C\left((\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4} \wedge \frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2}) \vee 1\right)\log^2 n$, the test $\phi_5 \vee \phi_6$ achieves reliable detection by calling $O(\log(nB/(\lambda^2 \wedge \lambda^4)))$ number of queries to VSTAT(n).

Finally,

- if $\lambda \le 1$, $\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4} \wedge \frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2} \asymp \frac{p^m}{\lambda^4} \wedge \frac{p^{(d+m)/2}}{\lambda^2}$;
- if $1 \le \lambda \le p^{m/4}$, $\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4} \wedge \frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2} \asymp \frac{p^m}{\lambda^4} \wedge \frac{p^{(d+m)/2}}{\lambda^2}$;
- if $\lambda \ge p^{m/4}$, $\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4} \wedge \frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2} \le 1$;

thus $\left(\frac{2p^m+4\lambda^2+2\lambda^4}{\lambda^4} \wedge \frac{\lambda^2+p^{d/2}\lambda^2+p^{(d+m)/2}}{\lambda^2}\right) \vee 1 \asymp \left(\frac{p^m}{\lambda^4} \wedge \frac{p^{(d+m)/2}}{\lambda^2}\right) \vee 1$ for all ranges of λ and this finishes the proof of this theorem.

E.3. An Equivalence Formulation For Scalar-on-tensor Regression

In this section, we show without loss of generality, we can assume $\boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ with $0 \leq \sigma^2 < 1$, $\boldsymbol{\mathcal{X}}_i \overset{i.i.d.}{\sim} N(0,1)$ and $\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2 = 1$ in establishing the computational lower bound for the testing problem (5). Consider $y_i = \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{X}}_i \rangle + \varepsilon_i$ and suppose we are in a simpler setting that $\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}}$ and σ are known. Then we can rescale the problem by multiplying $\sqrt{1/(\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2)}$ on both sides of the equation and get

$$y_i' = \langle \boldsymbol{\mathcal{A}}', \boldsymbol{\mathcal{X}}_i' \rangle + \varepsilon_i', \quad i = 1, \ldots, n, \tag{42}$$

where $y_i' = y_i/\sqrt{\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2} \sim N(0,1)$, $\boldsymbol{\mathcal{A}}' = \boldsymbol{\mathcal{A}}$ has i.i.d. $N(0,1)$ entries and $\boldsymbol{\mathcal{A}}' = \boldsymbol{\mathcal{A}}/\sqrt{\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2}$, ε_i' follows i.i.d. $N(0, \sigma^2/\sqrt{\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2})$ satisfying $\|\boldsymbol{\mathcal{A}}'\|_{\mathbf{F}}^2 + \mathbb{V}ar(\varepsilon_i') = 1$. Then the new null and alternative are exactly the ones in (5). Thus, without loss of generality, we can assume $\boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ with $0 \leq \sigma^2 < 1$, $\boldsymbol{\mathcal{X}}_i \overset{i.i.d.}{\sim} N(0,1)$ and $\|\boldsymbol{\mathcal{A}}\|_{\mathbf{F}} + \sigma^2 = 1$.

E.4. Proof of Theorem 25

We will first show the claim: for any $0 < \epsilon < \frac{1}{2}$, the pairwise-correlation-based statistical dimension (see Definition in (34)) with pairwise correlation $(O((1-\sigma^2)p^{(\epsilon - \frac{1}{2})d}), \frac{1}{\sigma^2} - 1)$ of the testing problem (5) is at least $2^{\Omega(p^\epsilon)}$. The Theorem 25 follows from Lemma 35 by picking $\gamma' = \gamma$.

Next we prove the claim. Take the distribution of $(\boldsymbol{\mathcal{X}}, y)$ generated from H_0 as the reference distribution and denote it by D_0. Denote the distribution of $(\boldsymbol{\mathcal{X}}, y)$ generated from H_1 with tensor parameter $\boldsymbol{\mathcal{A}} = \mathbf{a}^{\otimes d}$ as $D_{\mathbf{a}}$.

First, it is easy to check for $(\boldsymbol{\mathcal{X}}, y) \sim D_{\mathbf{a}}$, we have $(\text{vec}(\boldsymbol{\mathcal{X}}), y) \sim N(0, \boldsymbol{\Sigma}_{\mathbf{a}})$ with $\boldsymbol{\Sigma}_{\mathbf{a}} = \mathbf{I}_{p^d+1} + \mathbf{e}\bar{\mathbf{a}}^\top + \bar{\mathbf{a}}\mathbf{e}^\top = \begin{bmatrix} \mathbf{I}_{p^d} & \text{vec}(\mathbf{a}^{\otimes d}) \\ \text{vec}(\mathbf{a}^{\otimes d})^\top & 1 \end{bmatrix}$, where $\text{vec}(\cdot)$ denotes the vectorization of the input tensor, $\mathbf{e} = \begin{bmatrix} \mathbf{0}_{p^d} \\ 1 \end{bmatrix}$ and $\bar{\mathbf{a}} = \begin{bmatrix} \text{vec}(\mathbf{a}^{\otimes d}) \\ 0 \end{bmatrix}$. Notice

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{a}} + \boldsymbol{\Sigma}_{\mathbf{b}} - \boldsymbol{\Sigma}_{\mathbf{a}}\boldsymbol{\Sigma}_{\mathbf{b}} &= \mathbf{I} + \mathbf{e}\bar{\mathbf{a}}^\top + \bar{\mathbf{a}}\mathbf{e}^\top + \mathbf{I} + \mathbf{e}\bar{\mathbf{b}}^\top + \bar{\mathbf{b}}\mathbf{e}^\top - (\mathbf{I} + \mathbf{e}\bar{\mathbf{a}}^\top + \bar{\mathbf{a}}\mathbf{e}^\top)(\mathbf{I} + \mathbf{e}\bar{\mathbf{b}}^\top + \bar{\mathbf{b}}\mathbf{e}^\top) \\ &= \mathbf{I} + \mathbf{e}\bar{\mathbf{b}}^\top + \bar{\mathbf{b}}\mathbf{e}^\top - \mathbf{e}\bar{\mathbf{b}}^\top - \bar{\mathbf{b}}\mathbf{e}^\top - \bar{\mathbf{a}}\bar{\mathbf{b}}^\top - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}} \rangle \mathbf{e}\mathbf{e}^\top \\ &= \mathbf{I} - \bar{\mathbf{a}}\bar{\mathbf{b}}^\top - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}} \rangle \mathbf{e}\mathbf{e}^\top, \end{aligned} \tag{43}$$

and

$$\begin{aligned} \det(\mathbf{I} - \bar{\mathbf{a}}\bar{\mathbf{b}}^\top - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}} \rangle \mathbf{e}\mathbf{e}^\top) &= (1 - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}} \rangle) \cdot \det(\mathbf{I} - \text{vec}(\mathbf{a}^{\otimes d})\text{vec}(\mathbf{b}^{\otimes d})^\top) \\ &\overset{(Meyer,\ 2000, \text{Eq.6.2.2})}{=} (1 - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}} \rangle)^2. \end{aligned} \tag{44}$$

At the same time, we have $0 \prec \boldsymbol{\Sigma}_{\mathbf{a}}, \boldsymbol{\Sigma}_{\mathbf{b}} \prec 2\mathbf{I}_{p^d+1}$, the first "$\prec$" is because $\boldsymbol{\Sigma}_{\mathbf{a}}$ and $\boldsymbol{\Sigma}_{\mathbf{b}}$ are covariance matrices and they are full rank as $\sigma^2 < 1$; the second "$\prec$" is because $2\mathbf{I}_{p^d+1} - \boldsymbol{\Sigma}_{\mathbf{a}}$ is also a covariance matrix of $(\text{vec}(\boldsymbol{\mathcal{X}}), y)$ where their correlation is specified by $-\text{vec}(\mathbf{a}^{\otimes})$, similarly for $2\mathbf{I}_{p^d+1} - \mathbf{\Sigma_b}$. Then by (Diakonikolas and Kane, 2021, Lemma 4),

$$\begin{aligned}\chi_{D_0}(D_\mathbf{a}, D_\mathbf{b}) &= \chi_{N(0,\mathbf{I})}(N(0,\mathbf{\Sigma_a}), N(0,\mathbf{\Sigma_b})) \\ &\overset{\text{Lemma 4 of } DiakonikolasandKane\ (2021)}{=} (\det(\mathbf{\Sigma_a} + \mathbf{\Sigma_b} - \mathbf{\Sigma_a}\mathbf{\Sigma_b}))^{-1/2} - 1 \\ &\overset{(43)}{=} (\det(\mathbf{I} - \bar{\mathbf{a}}\bar{\mathbf{b}}^\top - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}}\rangle \mathbf{e}\mathbf{e}^\top))^{-1/2} - 1 \\ &\overset{(44)}{=} (1 - \langle \bar{\mathbf{a}}, \bar{\mathbf{b}}\rangle)^{-1} - 1 = (1 - (\mathbf{a}^\top \mathbf{b})^d)^{-1} - 1.\end{aligned}$$

This implies $\chi_{D_0}(D_\mathbf{a}, D_\mathbf{a}) = 1/\sigma^2 - 1$ since $\|\mathbf{a}^{\otimes d}\|_\mathbf{F}^2 = \|\mathbf{a}\|_2^{2d} = 1-\sigma^2$. Moreover, by Lemma 45, we can construct $N = 2^{\Omega(p^\epsilon)}$ vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ such that $\|\mathbf{a}_i\|_2^d = 1 - \sigma^2, \forall i \in [N]$ and for any distinct $i, j \in [N]$, $|\mathbf{a}_i^\top \mathbf{a}_j| \leq O((1-\sigma^2)p^{\epsilon - 1/2})$. Thus for distinct $i, j \in [m]$,

$$|\chi_{D_0}(D_{\mathbf{a}_i}, D_{\mathbf{a}_j})| \leq \frac{|(\mathbf{a}_i^\top \mathbf{a}_j)|^d}{1 - (\mathbf{a}_i^\top \mathbf{a}_j)^d} \leq O((1-\sigma^2)p^{(\epsilon - 1/2)d}) \tag{45}$$

E.5. Proof of Theorem 26

To estimate the quantity $T_6(y_1, \boldsymbol{\mathcal{X}}_1)$ via SQ, we need to compute the mean and second moment of this quantity under the null and alternative hypothesis. First, under H_0, we have $\mathbb{E}[T_6(y_1, \boldsymbol{\mathcal{X}}_1)] = 0$ and

$$\mathbb{E}[T_6^2(y_1, \boldsymbol{\mathcal{X}}_1)] = \mathbb{E}\left(\sum_{j_1,\ldots,j_{d/2}\in[p]} (y_1\boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}\right)^2 = \sum_{j_1,\ldots,j_{d/2}\in[p]} \mathbb{E}(y_1^2(\boldsymbol{\mathcal{X}}_1)^2_{[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}) = p^{d/2}.$$

Under the alternative, we have

$$\begin{aligned}\mathbb{E}[T_6(y_1, \boldsymbol{\mathcal{X}}_1)] &= \mathbb{E}\left(\sum_{j_1,\ldots,j_{d/2}\in[p]} (y_1\boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}\right) \\ &= \sum_{j_1,\ldots,j_{d/2}\in[p]} \mathbb{E}\left((y_1\boldsymbol{\mathcal{X}}_1)_{[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}\right) \\ &= \sum_{j_1,\ldots,j_{d/2}\in[p]} \mathbb{E}\left((\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{X}}_1\rangle + \varepsilon)\, \boldsymbol{\mathcal{X}}_{1[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}\right) \\ &= \sum_{j_1,\ldots,j_{d/2}\in[p]} \boldsymbol{\mathcal{A}}_{[j_1,j_1,\ldots,j_{d/2},j_{d/2}]} \\ &= \sum_{j_1,\ldots,j_{d/2}\in[p]} \mathbf{a}_{j_1}^2 \cdots \mathbf{a}_{j_{d/2}}^2 = \|\mathbf{a}\|_2^d = \|\boldsymbol{\mathcal{A}}\|_\mathbf{F} = \sqrt{1-\sigma^2},\end{aligned}$$

and $T_6 = y_1\xi_2$, where $\xi_2 = \sum_{j_1,\ldots,j_{d/2}=1}^p \boldsymbol{\mathcal{X}}_{1[j_1,j_1,\ldots,j_{d/2},j_{d/2}]}$. Notice $y_1 \sim N(0,1)$, $\xi_2 \sim N(0, p^{d/2})$ and $\mathbb{E}_{\boldsymbol{\mathcal{A}}}(T_6(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{X}}_1)) = \mathbb{E}_{\boldsymbol{\mathcal{A}}}(\xi_1\xi_2) = \sqrt{1-\sigma^2}$, thus y_1, ξ_2 follow the joint normal distribution

$$(y_1, \xi_2)^\top \sim N\left(0, \begin{bmatrix} 1 & \sqrt{1-\sigma^2} \\ \sqrt{1-\sigma^2} & p^{d/2}\end{bmatrix}\right).$$

So we can write $y_1 = \sqrt{1-\sigma^2}\xi_2/p^{d/2} + Z$ where $Z \sim N(0, 1-(1-\sigma^2)/p^{d/2})$ and is independent of ξ_2. Thus $\mathbb{E}[T_6^2(y_1, \boldsymbol{\mathcal{X}}_1)] = \mathbb{E}(y_1^2\xi_2^2) = \mathbb{E}((\sqrt{1-\sigma^2}\xi_2/p^{d/2} + Z)^2\xi_2^2) = p^{d/2} + 2(1-\sigma^2)$.

By setting $B = p^{d/2} + 2(1-\sigma^2)$ and $\xi = 1/\log(n)$ in Lemma 37, we have when $n(1-\sigma^2)/(p^{d/2}\log^2(n)) \to \infty$, with $O(\log(4nB/\xi^2))$ queries to $\mathrm{VSTAT}(n)$, there exists an SQ algorithm such that $\left|\hat{\mathbb{E}}_{x\sim D}[T_6(y_1, \boldsymbol{\mathcal{X}}_1)] - \mathbb{E}_{x\sim D}[T_6(y_1, \boldsymbol{\mathcal{X}}_1)]\right| / \sqrt{1-\sigma^2} \to 0$ as $n \to \infty$ under both null and alternative hypothesis. So the test $1(\hat{\mathbb{E}}_{(y,\boldsymbol{\mathcal{X}})\sim D}(T_6(y_1, \boldsymbol{\mathcal{X}}_1)) > \sqrt{1-\sigma^2}/2)$ can achieve vanishing type I+II error in solving the hypothesis testing problem (5).

Appendix F. Technical Lemmas

In this section, we collect a series of technical lemmas to be used in the main technical proofs of this paper.

Lemma 38 (χ^2-divergence and pairwise correlation) *Consider the tensor-on-tensor regression model (1) with $\boldsymbol{\mathcal{A}} = \lambda \cdot \mathbf{u}^{\otimes(d+m)}$, i.e., $\boldsymbol{\mathcal{Y}}_i = \langle \boldsymbol{\mathcal{X}}_i, \lambda \cdot \mathbf{u}^{\otimes(d+m)}\rangle_* + \boldsymbol{\mathcal{E}}_i$, where $\mathbf{u} \in \mathcal{S}_{p-1}$. Suppose $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{E}}_i$ are independent and have i.i.d. $N(0,1)$ entries. Denote the joint density of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ as $D_{\mathbf{u}}$ and the joint distribution of $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)$ with $\lambda = 0$ as D_0. We also use $D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)$ and $D_0(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)$ to denote the conditional density of $\boldsymbol{\mathcal{Y}}_i$ given $\boldsymbol{\mathcal{X}}_i$ when $(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i) \sim D_{\mathbf{u}}$ or D_0, respectively. Then given any $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{p-1}$, we have*

$$\begin{aligned} \chi^2(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i), D_0(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)) + 1 &= \exp(\lambda^2\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle^2), \\ \chi_{D_0}(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i), D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)) + 1 &= \exp(\lambda^2 t^m \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle\langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle), \end{aligned} \tag{46}$$

where $t := \mathbf{u}^\top \mathbf{v}$. Thus

$$\begin{aligned} \chi^2(D_{\mathbf{u}}, D_0) + 1 &= \mathbb{E}_{\boldsymbol{\mathcal{X}}_i}\left(\exp(\lambda^2\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle^2)\right), \\ \chi_{D_0}(D_{\mathbf{u}}, D_{\mathbf{v}}) + 1 &= \mathbb{E}_{\boldsymbol{\mathcal{X}}_i}\left(\exp(\lambda^2 t^m \langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle\langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}^{\otimes d}\rangle)\right), \end{aligned} \tag{47}$$

In particular, if $\lambda < 1/2$, we have

$$\chi^2(D_{\mathbf{u}}, D_0) + 1 = (1 - 2\lambda^2)^{-1/2}. \tag{48}$$

If $|(1-t^{2d})\lambda^2 t^m + t^d| < 1$, we have

$$\begin{aligned} \chi_{D_0}(D_{\mathbf{u}}, D_{\mathbf{v}}) + 1 &= \frac{\sqrt{1-t^{2d}}}{\sqrt{1-(1-t^{2d})^2(\lambda^2 t^m + t^d/(1-t^{2d}))^2}} \\ &= \left(1 - t^{d+m}\lambda^2\right)^{-1/2} \cdot \left(1 - \lambda^2 t^{d+m} - \frac{\lambda^4 t^{2m}}{1-\lambda^2 t^{d+m}}\right)^{-1/2}. \end{aligned} \tag{49}$$

Proof of Lemma 38. Let us first prove the result for $\chi_{D_0}(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i), D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)) + 1$ in (46), the expression for $\chi^2(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i), D_0(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_i)) + 1$ can be obtained by setting $\mathbf{u} = \mathbf{v}$. By Lemma 44 (53), we know $\boldsymbol{\mathcal{Y}}_i = \lambda\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle \mathbf{u}^{\otimes m} + \boldsymbol{\mathcal{E}}_i$. Thus, $\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i) \sim N(\mathrm{vec}(\lambda\langle \boldsymbol{\mathcal{X}}_i, \mathbf{u}^{\otimes d}\rangle \mathbf{u}^{\otimes m}), \mathbf{I}_{p^m})$.

Then

$$\begin{aligned}
&\chi_{D_0}(D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i,\boldsymbol{\mathcal{X}}_i),D_{\mathbf{v}}(\boldsymbol{\mathcal{Y}}_i,\boldsymbol{\mathcal{X}}_i))+1\\
&=\int\frac{D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)D_{\mathbf{u}}(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)}{D_0(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{X}}_i)}d\boldsymbol{\mathcal{Y}}_i\\
&=\int\frac{1}{(2\pi)^{p^m/2}}\frac{\exp(-\frac{1}{2}\|\boldsymbol{\mathcal{Y}}_i-\lambda\langle\boldsymbol{\mathcal{X}}_i,\mathbf{u}^{\otimes d}\rangle\mathbf{u}^{\otimes m}\|_{\mathbf{F}}^2-\frac{1}{2}\|\boldsymbol{\mathcal{Y}}_i-\lambda\langle\boldsymbol{\mathcal{X}}_i,\mathbf{v}^{\otimes d}\rangle\mathbf{v}^{\otimes m}\|_{\mathbf{F}}^2)}{\exp(-\frac{1}{2}\|\boldsymbol{\mathcal{Y}}_i\|_{\mathbf{F}}^2)}d\boldsymbol{\mathcal{Y}}_i\\
&=\exp(\lambda^2\langle\mathbf{u},\mathbf{v}\rangle^m\langle\boldsymbol{\mathcal{X}}_i,\mathbf{v}^{\otimes d}\rangle\langle\boldsymbol{\mathcal{X}}_i,\mathbf{u}^{\otimes d}\rangle).
\end{aligned}\tag{50}$$

Next let us prove (48). Let us do a change of variable, set

$$\mathrm{vec}(\widetilde{\boldsymbol{\mathcal{X}}_i})=[\mathbf{v}_1,\mathbf{v}_{1\perp}]^\top\mathrm{vec}(\boldsymbol{\mathcal{X}}_i),$$

where $\mathbf{v}_1=\mathrm{vec}(\mathbf{u}^{\otimes d})\in\mathbb{R}^{p^d}$ and $\mathbf{v}_{1\perp}\in\mathbb{R}^{p^d\times(p^d-1)}$ is the orthogonal complement of $\mathbf{v}_1$. By doing so, we have

$$\begin{aligned}
\chi^2(D_{\mathbf{u}},D_0)+1&=\mathbb{E}_{\boldsymbol{\mathcal{X}}_i}\left(\exp(\lambda^2\langle\boldsymbol{\mathcal{X}}_i,\mathbf{u}^{\otimes d}\rangle^2)\right)\\
&=\int\frac{1}{\sqrt{2\pi}}\exp(\lambda^2\widetilde{x}_1^2-\widetilde{x}_1^2/2)d\widetilde{x}_1,\\
&=(1-2\lambda^2)^{-1/2},
\end{aligned}$$

where $\widetilde{x}_1$ is the first element in $\mathrm{vec}(\widetilde{\boldsymbol{\mathcal{X}}_i})$ and in the last inequality, we use the fact $\lambda^2<1/2$ and the integral of a Gaussian random variable.

Finally, we compute $\chi_{D_0}(D_{\mathbf{u}},D_{\mathbf{v}})+1$. Notice that $|(1-t^{2d})\lambda^2t^m+t^d|<1$ implies $t<1$. Again we perform the change of variables in the integration

$$\mathrm{vec}(\widetilde{\boldsymbol{\mathcal{X}}_i})=[\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_{12\perp}]^\top\mathrm{vec}(\boldsymbol{\mathcal{X}}_i),$$

where $\mathbf{v}_1=\mathrm{vec}(\mathbf{u}^{\otimes d})\in\mathbb{R}^{p^d},\mathbf{v}_2=\mathrm{vec}(\mathbf{v}^{\otimes d})\in\mathbb{R}^{p^d}$ and $\mathbf{v}_{12\perp}\in\mathbb{R}^{p^d\times(p^d-2)}$ is the orthogonal complement of $[\mathbf{v}_1,\mathbf{v}_2]$. Then

$$\begin{aligned}
\chi_{D_0}(D_{\mathbf{u}},D_{\mathbf{v}})+1&=\mathbb{E}_{\boldsymbol{\mathcal{X}}_i}\left(\exp(\lambda^2t^m\langle\boldsymbol{\mathcal{X}}_i,\mathbf{u}^{\otimes d}\rangle\langle\boldsymbol{\mathcal{X}}_i,\mathbf{v}^{\otimes d}\rangle)\right)\\
&=\int\exp(\lambda^2t^m\widetilde{x}_1\widetilde{x}_2)\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp(-\frac{1}{2}(\widetilde{x}_1,\widetilde{x}_2)\boldsymbol{\Sigma}^{-1}(\widetilde{x}_1,\widetilde{x}_2)^\top)d\widetilde{x}_1d\widetilde{x}_2\\
&=\int\frac{1}{\sqrt{1-t^{2d}}}\exp\left(-\frac{1}{2}(\widetilde{x}_1,\widetilde{x}_2)\widetilde{\boldsymbol{\Sigma}}^{-1}(\widetilde{x}_1,\widetilde{x}_2)^\top\right)d\widetilde{x}_1d\widetilde{x}_2\\
&\overset{(a)}{=}\frac{|\widetilde{\boldsymbol{\Sigma}}|^{1/2}}{\sqrt{1-t^{2d}}}=\frac{\sqrt{1-t^{2d}}}{\sqrt{1-(1-t^{2d})^2(\lambda^2t^m+t^d/(1-t^{2d}))^2}}\\
&=\left(1-t^{d+m}\lambda^2\right)^{-1/2}\cdot\left(1-\lambda^2t^{d+m}-\frac{\lambda^4t^{2m}}{1-\lambda^2t^{d+m}}\right)^{-1/2}
\end{aligned}$$

here $|\cdot|$ denotes the determinant of a given matrix, $\boldsymbol{\Sigma}=\begin{bmatrix}1&t^d\\t^d&1\end{bmatrix}$ is the covariant matrix between $\widetilde{x}_1$ and $\widetilde{x}_2$ and

$$\widetilde{\boldsymbol{\Sigma}}^{-1}=\begin{bmatrix}1&-(1-t^{2d})(\lambda^2t^m+t^d/(1-t^{2d}))\\-(1-t^{2d})(\lambda^2t^m+t^d/(1-t^{2d}))&1\end{bmatrix}/(1-t^{2d}).$$

In (a), we use the fact that $\widetilde{\mathbf{\Sigma}}$ is a positive definite matrix when $|(1-t^{2d})\lambda^2 t^m + t^d| < 1$ and the integration for the multivariate Gaussian random variable. This finishes the proof of this lemma. $\square$

Lemma 39 (Lemma 11 of Carpentier et al. (2019)) *Assume that the matrix* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *has i.i.d.* $N(0,1)$ *entries. Then for any* $n \geq 1$*, there exists a universal* $c_0 > 0$ *such that for all* $\theta, \theta' \in \mathbb{R}^p$ *with* $\|\theta\|_2, \|\theta'\|_2 \leq cn^{-1/4}$ *and* $c \in (0, c_0)$ *we have* $\mathbb{E}_{\mathbf{X}} \exp(\langle \mathbf{X}\theta, \mathbf{X}\theta'\rangle) \leq \exp(n\langle\theta,\theta'\rangle)(1+C_0c^2)$*, where* $C_0 > 0$ *is also a universal constant.*

Lemma 40 (Lemma 8 of Luo and Zhang (2022a)) *Let* $\mathcal{C}$ *denotes the set of all unit vectors in* $\mathbb{R}^p$ *with* $\{1/\sqrt{p}, -1/\sqrt{p}\}$ *entries. Then for* $d \geq 2$*, there exits a function* $g : (0, c_0) \to (1, \infty)$ *with* $c_0 > 0$ *is a fixed small constant and the left limit of* g *at* 0 *is* 1*, i.e.,* $g(0+) = 1$*, such that for any* $c < c_0$*, we have* $\mathbb{E}_{\mathbf{u},\mathbf{v}\in \mathrm{Unif}(\mathcal{C})} \exp(h\langle \mathbf{u}, \mathbf{v}\rangle^d) \leq g(c)$ *given* $h = cp$*.*

Lemma 41 *For any* $0 < \delta < 1$*, then*

$$\frac{1}{1-x} \leq \exp\left(\sqrt{\frac{2}{1-\delta}}x\right), \quad \textit{and} \quad \frac{1}{1+x} \leq \exp(-\frac{x}{1+\delta}) \quad \forall\, 0 \leq x \leq \delta.$$

Proof of Lemma 41 . We first prove the first inequality. By Taylor expansion, we have

$$\begin{aligned} \frac{1}{1-x} &= 1 + x + x^2 + x^3 + x^4 + \cdots \\ &\leq 1 + x + \frac{x^2}{1-x} \overset{(a)}{\leq} 1 + x + \frac{x^2}{1-\delta}. \end{aligned}$$

Here (a) is because $0 \leq x \leq \delta$. At the same time,

$$\exp\left(\sqrt{\frac{2}{1-\delta}}x\right) = 1 + \sqrt{\frac{2}{1-\delta}}x + \frac{(\sqrt{\frac{2}{1-\delta}}x)^2}{2!} + \cdots \geq 1 + \sqrt{\frac{2}{1-\delta}}x + \frac{x^2}{1-\delta} \geq \frac{1}{1-x}$$

To show the second inequality, we just need to show $1 + x \geq \exp(\frac{x}{1+\delta})$. By Taylor expansion, we have

$$\begin{aligned} \exp(\frac{x}{1+\delta}) = 1 + \frac{x}{1+\delta} + (\frac{x}{1+\delta})^2/2! + \cdots &\leq 1 + \frac{x}{1+\delta} + (\frac{x}{1+\delta})^2 + (\frac{x}{1+\delta})^3 + \cdots \\ &\leq 1 + (\frac{x}{1+\delta})/(1 - \frac{x}{1+\delta}) \leq 1 + (\frac{x}{1+\delta})/(1 - \frac{\delta}{1+\delta}) \\ &= 1 + x. \end{aligned}$$

$\square$

Lemma 42 *Suppose* $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{p_1\times\cdots\times p_d}$ *for* $i = 1, \ldots, n$ *are independent and has i.i.d.* $N(0,1)$ *entries. Then for any fixed* $\mathbf{v}_j \in \mathcal{S}_{p_j-1}$ *for* $j = 1, \ldots, d$*, we have* $\mathbb{P}(|\sum_{i=1}^n \langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_{d+m}\rangle^2 - n| \geq n/2) \leq 2\exp(-cn)$ *for some* $c > 0$ *and every positive integer* n*.*

Proof Notice that $\{\langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_{d+m} \rangle\}$ are i.i.d. with $N(0,1)$. So $\langle \boldsymbol{\mathcal{X}}_i, \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_{d+m} \rangle^2 - 1$ for $i = 1, \ldots, n$ are i.i.d. centered sub-exponential random variables. Then the results follow by applying Bernstein's inequality for the concentration of the sum of sub-exponential random variables, see (Vershynin, 2010, Proposition 5.16). ■

Lemma 43 (Concentration for Random Tensors) *(a) Suppose $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$ is an order-d random tensor with i.i.d. $N(0,1)$ entries. Then*

$$\mathbb{P}\left(\sup_{\mathbf{v}_j \in \mathcal{S}_{p_j - 1}, j=1,\ldots,d} \boldsymbol{\mathcal{E}} \times_{j=1}^{d} \mathbf{v}_j \geq C \sqrt{\sum_{j=1}^{d} p_j} \right) \leq \exp(-C'(\sum_{j=1}^{d} p_j)).$$

(b) Suppose $\mathscr{X}^*$ is the adjoint operator of the linear operator $\mathscr{X}$ defined in (2) and $\boldsymbol{\mathcal{X}}_i$ for $i = 1, \ldots, n$ has i.i.d. $N(0,1)$ entries. If $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{p_{d+1} \times \cdots \times p_{d+m} \times n}$ has i.i.d. $N(0,1)$ entries and $n \geq C_{d,m}(\sum_{j=1}^{d+m} p_j)$ for some sufficiently large $C_{d,m}$, then

$$\mathbb{P}\left(\sup_{\mathbf{v}_j, j=1,\ldots,d+m} \mathscr{X}^*(\boldsymbol{\mathcal{E}}) \times_{j=1}^{d+m} \mathbf{v}_j \geq c \sqrt{n \left(\sum_{j=1}^{d+m} p_j \right)} \right) \leq \exp(-c'(\sum_{j=1}^{d+m} p_j)).$$

Proof Part (a) can be found in (Zhang and Xia, 2018, Lemma 5) and part (b) is proved in (Luo and Zhang, 2022b, Theorem 4, part 1). ■

The next lemma 44, which reveals a few useful properties of the contracted tensor inner product defined in (1).

Lemma 44 (Properties of Contracted Tensor Inner Product (Lemma 12 of Luo and Zhang (2022b))) *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{p_1 \times \cdots \times p_d \times p_{d+1} \times \cdots \times p_{d+m}}$, $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$, $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{p_1 \times \cdots \times p_{k-1} \times q_k \times p_{k+1} \times \cdots \times p_d}$ be tensors with $d \geq k \geq 1$, $m \geq 0$. For any $\mathbf{A} \in \mathbb{R}^{q_k \times p_k}$, we have*

$$\langle \boldsymbol{\mathcal{X}} \times_k \mathbf{A}, \boldsymbol{\mathcal{W}} \rangle_* = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{W}} \times_k \mathbf{A}^\top \rangle_*. \tag{51}$$

For any $\mathbf{B} \in \mathbb{R}^{q_{d+j} \times p_{d+j}}$ with $1 \leq j \leq m$, we have

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \rangle_* \times_j \mathbf{B} = \langle \boldsymbol{\mathcal{X}} \times_{d+j} \mathbf{B}, \boldsymbol{\mathcal{Z}} \rangle_*. \tag{52}$$

In particular, given $\mathbf{u}_j \in \mathbb{R}^{p_j}$ for $1 \leq j \leq d+m$, we have

$$\langle \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d, \boldsymbol{\mathcal{Z}} \rangle \cdot \mathbf{u}_{d+1} \otimes \cdots \otimes \mathbf{u}_{d+m} = \langle \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d, \boldsymbol{\mathcal{Z}} \rangle_* \times_{j=d+1}^{d+m} \mathbf{u}_j = \langle \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d+m}, \boldsymbol{\mathcal{Z}} \rangle_*. \tag{53}$$

The following lemma is important in showing the SQ hardness for tensor-on-tensor association detection.

Lemma 45 ((Diakonikolas et al., 2017, Lemma 3.7)) *For any $0 < c < 1/2$, there is a set S' of at least $2^{\Omega(p^c)}$ unit vectors in $\mathbb{R}^p$ such that for each pair of distinct $\mathbf{u}, \mathbf{u}' \in S'$, it holds $|\mathbf{u}^\top \mathbf{u}'| \leq O(p^{c-1/2})$.*

The next lemma provides the distribution for the inner of two unit vectors drawn uniformly from the sphere.

Lemma 46 *Let* $\mathbf{u}, \mathbf{v}$ *be two random vectors drawn uniformly at random from* $\mathcal{S}_{p-1}$*. Then* $(\mathbf{u}^\top \mathbf{v} + 1)/2$ *follows Beta* $\left(\frac{p-1}{2}, \frac{p-1}{2}\right)$ *and* $(\mathbf{u}^\top \mathbf{v})^2$ *follows Beta* $\left(\frac{1}{2}, \frac{p-1}{2}\right)$.

Proof Let $Z = \mathbf{u}^\top \mathbf{v}$, then by Cho (2009) Theorem 1, we have $f_Z(z) \propto (1 - z^2)^{(n-3)/2}$, this implies that $f_Y(y) \propto (y - y^2)^{(n-3)/2}$ for $Y = (\mathbf{u}^\top \mathbf{v} + 1)/2$. This is the density function for Beta $\left(\frac{p-1}{2}, \frac{p-1}{2}\right)$ up to the normalization constant. This proves the first statement.

For the second statement, since the distribution of $\mathbf{u}^\top \mathbf{v}$ does not change if we rotate them by the same orthogonal matrix. Without loss of generality we can assume $\mathbf{v} = (1, 0, \dots, 0)^\top$. Moreover, it is well known that $\mathbf{u}$ follows the same distribution as $Z_1/\sqrt{(\sum_{i=1}^p Z_i^2)}$ where $Z_i \overset{i.i.d.}{\sim} N(0, 1)$. Thus $(\mathbf{u}^\top \mathbf{v})^2 \sim Z_1^2/(\sum_{i=1}^p Z_i^2)$ and this is Beta $\left(\frac{1}{2}, \frac{p-1}{2}\right)$ as $Z_1^2/(\sum_{i=1}^p Z_i^2)$ can be written as $X/(X + Y)$ where $X \sim \chi_1^2, Y \sim \chi_{p-1}^2$ and X is independent of Y.

■