

On the estimation of Gaussian moment tensors*

Omar Al-Ghattas[†] Jiaheng Chen[‡] Daniel Sanz-Alonso[§]

Abstract

This paper studies two estimators for Gaussian moment tensors: the standard sample moment estimator and a plug-in estimator based on Isserlis’s theorem. We establish dimension-free, non-asymptotic error bounds that demonstrate and quantify the advantage of Isserlis’s estimator for tensors of even order $p > 2$. Our bounds hold in operator and entrywise maximum norms, and apply to symmetric and asymmetric tensors.

Keywords: moment tensor estimation; Isserlis’s theorem; concentration inequalities.

MSC2020 subject classifications: 62H12; 60E15; 15A69.

Submitted to ECP on July 8, 2025, final version accepted on October 28, 2025.

1 Introduction

This paper develops dimension-free, non-asymptotic theory for estimating Gaussian moment tensors of arbitrary order. We study two estimators: the sample moment estimator and a plug-in estimator based on Isserlis’s formula. Our non-asymptotic bounds quantify the advantage of Isserlis’s estimator for tensors of even order $p > 2$, both in operator and entrywise maximum norms. By studying the sample moment estimator, this work contributes to the emerging literature on concentration inequalities for random tensors [27, 30, 34, 33, 9, 11, 4, 12]; and by studying Isserlis’s estimator, it contributes to the literature on estimating functionals of high and infinite-dimensional parameters [16, 17, 20, 18]. More broadly, this work is motivated by the growing adoption of tensor methods in statistics and machine learning [23, 21, 8, 10, 7, 13, 25], which calls for new theory for high-dimensional tensor estimation.

The concentration of the sample covariance (the sample moment tensor of order $p = 2$) has been extensively studied. Most pertinent to our work, the papers [28, 19] identify a notion of *effective dimension* that determines the sample complexity. In contrast to our results for even order $p > 2$, in the case $p = 2$ the sample covariance matrix is known to be minimax optimal under the operator norm (see [22, Theorem 2] and [19, Theorem 4]). Sharp operator norm concentration inequalities for simple random tensors of arbitrary order were recently established in [2, 12] using the effective dimension from [28, 19]. In this paper, we prove a new sharp concentration inequality in

*The work of DSA was partly funded by NSF CAREER DMS-2237628. The work of OAG was supported in part by funding from the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard.

[†]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, United States of America. E-mail: oag@broadinstitute.org

[‡]University of Chicago, United States of America. E-mail: jiaheng@uchicago.edu

[§]University of Chicago, United States of America. E-mail: sanzalonso@uchicago.edu

entrywise maximum norm using a notion of effective dimension introduced in [3, 4] to analyze thresholding estimators for structured covariance operators.

Isserlis's theorem expresses higher-order moments of a Gaussian distribution as functionals of its covariance. Consequently, our problem falls within the general framework of estimating functionals $f(\Sigma)$ of a covariance operator Σ [16, 17], where plug-in estimators $f(\hat{\Sigma})$ based on the sample covariance $\hat{\Sigma}$ are often severely biased and suboptimal [20, 18]. Bias reduction techniques —such as jackknife, bootstrap, and Taylor expansions— are commonly used to improve efficiency. Our new dimension-free, non-asymptotic error bounds for Isserlis's estimator show that, in the context of estimating Gaussian moment tensors, plug-in estimation via Isserlis's formula performs well and strictly outperforms the sample moment estimator.

2 Preliminaries

Given positive integers d_1, \dots, d_p and tensors $T, T' \in \mathbb{R}^{d_1 \times \dots \times d_p}$, their Frobenius inner product is defined by

$$\langle T, T' \rangle := \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \cdots \sum_{i_p=1}^{d_p} T_{i_1, \dots, i_p} T'_{i_1, \dots, i_p}.$$

For vectors $v_k \in \mathbb{R}^{d_k}$, $1 \leq k \leq p$, let $v_1 \otimes \dots \otimes v_p \in \mathbb{R}^{d_1 \times \dots \times d_p}$ denote their outer product, defined entrywise by

$$(v_1 \otimes \dots \otimes v_p)_{\ell_1, \dots, \ell_p} := v_{1, \ell_1} \cdots v_{p, \ell_p},$$

where $v_{i, \ell}$ denotes the ℓ -th coordinate of v_i . The operator norm of $T \in \mathbb{R}^{d_1 \times \dots \times d_p}$ is

$$\|T\| := \sup_{v_k \in \mathcal{S}^{d_k}, 1 \leq k \leq p} \langle T, v_1 \otimes \dots \otimes v_p \rangle,$$

where $\mathcal{S}^{d_k} := \{v \in \mathbb{R}^{d_k} : |v| = 1\}$ denotes the unit Euclidean sphere in \mathbb{R}^{d_k} . Similarly, the entrywise maximum norm is defined by

$$\|T\|_{\max} := \sup_{v_k \in \mathcal{E}^{d_k}, 1 \leq k \leq p} |\langle T, v_1 \otimes \dots \otimes v_p \rangle|,$$

where $\mathcal{E}^{d_k} := \{e_i\}_{i=1}^{d_k}$ is the standard basis of \mathbb{R}^{d_k} .

We will consider two notions of effective dimension of a covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. The first is

$$r_2(\Sigma) := \frac{\text{Tr}(\Sigma)}{\|\Sigma\|},$$

which will arise in our bounds in operator norm (see e.g. [19, 28]). The second is

$$r_{\max}(\Sigma) := \frac{(\mathbb{E}_{X \sim \mathcal{N}(0, \Sigma)} \|X\|_{\infty})^2}{\|\Sigma\|_{\max}},$$

which arises in our bounds in entrywise maximum norm (see e.g. [5, 3, 4]).

For positive sequences $\{a_n\}, \{b_n\}$, we write $a_n \lesssim b_n$ to denote that, for some constant $c > 0$, $a_n \leq cb_n$. If both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, we write $a_n \asymp b_n$. We write \lesssim_p and \asymp_p to indicate that the implicit constant may depend on the parameter p .

3 Main results

This section contains our main results. In Subsection 3.1, we study the *symmetric case*: moment tensors defined by expectation of the p -fold outer product of a single random vector. In Subsection 3.2, we consider the *asymmetric case*: moment tensors defined by expectation of the outer product of p random vectors of possibly different dimensions. While the asymmetric case subsumes the symmetric one, we present first the symmetric case due to its simplicity and its central role in theory and applications.

3.1 Symmetric case

Let $X \sim \mathcal{N}(0, \Sigma)$ be a zero-mean Gaussian random vector in \mathbb{R}^d with covariance matrix Σ , and let X_1, \dots, X_N be i.i.d. copies of X . Let p be an even integer. Our goal is to estimate the p -th order moment tensor

$$T := \mathbb{E}X^{\otimes p} := \mathbb{E}X \otimes \dots \otimes X, \quad (3.1)$$

which is the expectation of the p -fold tensor product of X .

A natural estimator is the sample moment estimator \hat{T}_S , defined by

$$\hat{T}_S := \frac{1}{N} \sum_{i=1}^N X_i^{\otimes p}. \quad (3.2)$$

Entrywise, this corresponds to

$$(\hat{T}_S)_{\ell_1, \dots, \ell_p} := \frac{1}{N} \sum_{i=1}^N X_{i, \ell_1} \dots X_{i, \ell_p}.$$

Alternatively, Isserlis's theorem [15] (also known as Wick's probability theorem [31]) provides an exact expression for Gaussian moment tensors in terms of second-order moments. Specifically, for any multi-index (ℓ_1, \dots, ℓ_p) ,

$$(\mathbb{E}X^{\otimes p})_{\ell_1, \dots, \ell_p} = \mathbb{E}[X_{\ell_1} \dots X_{\ell_p}] = \sum_{\pi \in \Pi_p^2} \prod_{(j, k) \in \pi} \Sigma_{\ell_j, \ell_k},$$

where Π_p^2 is the set of all pairwise partitions of $\{1, \dots, p\}$. This motivates Isserlis's estimator, which substitutes the sample covariance

$$\hat{\Sigma} := \frac{1}{N} \sum_{i=1}^N X_i X_i^\top$$

into the same expression. The resulting estimator is given entrywise by

$$(\hat{T}_I)_{\ell_1, \dots, \ell_p} := \sum_{\pi \in \Pi_p^2} \prod_{(j, k) \in \pi} \hat{\Sigma}_{\ell_j, \ell_k}. \quad (3.3)$$

Using the notion of induced likelihood [32], the plug-in estimator \hat{T}_I can be interpreted as the maximum likelihood estimator of T .

Theorem 3.1 compares the performance of the two estimators, \hat{T}_S and \hat{T}_I , under both the operator norm and the entrywise maximum norm.

Theorem 3.1. *Let $X \sim \mathcal{N}(0, \Sigma)$ be a zero-mean Gaussian random vector in \mathbb{R}^d with covariance matrix Σ , and let X_1, \dots, X_N be i.i.d. copies of X . For any even integer p , let T , \hat{T}_S , and \hat{T}_I be as defined in equations (3.1), (3.2), and (3.3), respectively. Then the following bounds hold.*

(i) *Operator norm bounds:*

$$\begin{aligned}\mathbb{E}\|\hat{T}_S - T\| &\asymp_p \|\Sigma\|^{p/2} \left(\sqrt{\frac{r_2(\Sigma)}{N}} + \frac{r_2(\Sigma)^{p/2}}{N} \right), \\ \mathbb{E}\|\hat{T}_I - T\| &\lesssim_p \|\Sigma\|^{p/2} \left(\sqrt{\frac{r_2(\Sigma)}{N}} + \left(\frac{r_2(\Sigma)}{N} \right)^{p/2} \right).\end{aligned}$$

(ii) *Entrywise maximum norm bounds:*

$$\begin{aligned}\mathbb{E}\|\hat{T}_S - T\|_{\max} &\asymp_p \|\Sigma\|_{\max}^{p/2} \left(\sqrt{\frac{r_{\max}(\Sigma)}{N}} + \frac{r_{\max}(\Sigma)^{p/2}}{N} \right), \\ \mathbb{E}\|\hat{T}_I - T\|_{\max} &\lesssim_p \|\Sigma\|_{\max}^{p/2} \left(\sqrt{\frac{r_{\max}(\Sigma)}{N}} + \left(\frac{r_{\max}(\Sigma)}{N} \right)^{p/2} \right).\end{aligned}$$

Proof. (i) *Operator norm bounds.* The sample moment estimator bound follows by [2, Theorem 2.1] (see also [1]). For \hat{T}_I , by Corollary 3.9, we have

$$\frac{\|\hat{T}_I - T\|}{\|T\|} \leq \frac{p}{2} \cdot \frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \left(1 + \frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \right)^{p/2-1} \lesssim_p \frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} + \left(\frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \right)^{p/2}.$$

Further, by [19, Theorem 4], $\mathbb{E}\|\hat{\Sigma} - \Sigma\|/\|\Sigma\| \asymp \sqrt{\frac{r_2(\Sigma)}{N}} + \frac{r_2(\Sigma)}{N}$, which can be plugged into the expression in the previous display to give the result.

(ii) *Entrywise maximum norm bounds.* For the sample moment estimator, the upper bound is a direct corollary of Theorem 3.6. The lower bound follows by an analogous argument to the one used in the proof of [2, Theorem 2.1] and is omitted for brevity. For Isserlis's estimator, the bound follows from Corollary 3.9. \square

Remark 3.2. Theorem 3.1 shows that consistent estimation of T using the sample moment estimator \hat{T}_S requires a sample size $N \gg (r_2(\Sigma))^{p/2}$ under the operator norm, or $N \gg (r_{\max}(\Sigma))^{p/2}$ under the entrywise maximum norm. In contrast, consistency of Isserlis's estimator \hat{T}_I only requires sample size $N \gg r_2(\Sigma)$ or $N \gg r_{\max}(\Sigma)$, leading to a significant reduction in sample complexity.

Remark 3.3. The upper bounds for the sample moment tensor in Theorem 3.1 hold for sub-Gaussian data. In contrast, our analysis of the Isserlis's estimator is limited to Gaussian data. An interesting question beyond the scope of this work is to leverage the generalization of Isserlis's theorem for isotropic distributions in [24, Theorem 3] to define and analyze more general Isserlis's-type estimators. Numerical results in [6] suggest that the advantage of Isserlis's-type estimators over the sample moment tensor may carry over to isotropic, sub-Gaussian data.

The following theorem establishes a lower bound for Isserlis's estimator.

Theorem 3.4. Let $X \sim \mathcal{N}(0, \Sigma)$ be a zero-mean Gaussian random vector in \mathbb{R}^d with covariance matrix Σ , and let X_1, \dots, X_N be i.i.d. copies of X . For any even integer p , let T and \hat{T}_I be as defined in equations (3.1) and (3.3), respectively. Then the following bounds hold.

(i) *Operator norm bound:*

$$\mathbb{E}\|\hat{T}_I - T\| \gtrsim_p \|\Sigma\|^{p/2} \left(\frac{1}{\kappa(\Sigma)^{p/2-1}} \sqrt{\frac{r_2(\Sigma)}{N}} + \left(\frac{r_2(\Sigma)}{N} \right)^{p/2} \right),$$

where $\kappa(\Sigma) := \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ is the condition number of Σ .

(ii) *Entrywise maximum norm bound:*

$$\mathbb{E}\|\hat{T}_I - T\|_{\max} \gtrsim_p \|\Sigma\|_{\max}^{p/2} \left(\frac{1}{\kappa(D(\Sigma))^{p/2-1}} \sqrt{\frac{r_{\max}(\Sigma)}{N}} + \left(\frac{r_{\max}(\Sigma)}{N} \right)^{p/2} \right),$$

where $D(\Sigma)$ is the diagonal matrix with the same diagonal entries as Σ , and $\kappa(D(\Sigma))$ denotes the condition number of $D(\Sigma)$.

Proof. (i) *Operator norm bound.* By Proposition 3.5 below, we have

$$\frac{\|\hat{T}_I - T\|}{\|T\|} \geq \max \left\{ \frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \left(\frac{1}{\kappa(\Sigma)} \right)^{p/2-1}, \left(\frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \right)^{p/2} \right\}.$$

Taking expectations on both sides and substituting the bound $\mathbb{E}\|\hat{\Sigma} - \Sigma\|/\|\Sigma\| \asymp \sqrt{\frac{r_2(\Sigma)}{N}} + \frac{r_2(\Sigma)}{N}$ from [19, Theorem 4] yields the desired result.

(ii) *Entrywise maximum norm bound.* By Proposition 3.5 below,

$$\frac{\|\hat{T}_I - T\|_{\max}}{\|T\|_{\max}} \geq \max \left\{ \frac{\|D(\hat{\Sigma} - \Sigma)\|_{\max}}{\|D(\Sigma)\|_{\max}} \left(\frac{1}{\kappa(D(\Sigma))} \right)^{p/2-1}, \left(\frac{\|D(\hat{\Sigma} - \Sigma)\|_{\max}}{\|D(\Sigma)\|_{\max}} \right)^{p/2} \right\}.$$

A straightforward adaptation of the argument in [12, Proposition 3.1] and [2, Proposition 3.1] —replacing the operator norm with the entrywise maximum norm— yields

$$\mathbb{E}\|D(\hat{\Sigma} - \Sigma)\|_{\max} = \mathbb{E} \sup_{v \in \mathcal{E}^d} |\langle (\hat{\Sigma} - \Sigma)v, v \rangle| \gtrsim \|\Sigma\|_{\max} \left(\sqrt{\frac{r_{\max}(\Sigma)}{N}} + \frac{r_{\max}(\Sigma)}{N} \right).$$

Taking expectations in the previous inequality, substituting this bound, and noting that $\|D(\Sigma)\|_{\max} = \|\Sigma\|_{\max}$ completes the proof. \square

The following proposition, proved in Section 4, was used in the proof of Theorem 3.4.

Proposition 3.5. *Let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$, and let $T_X := \mathbb{E}X^{\otimes p}$, $T_Y := \mathbb{E}Y^{\otimes p}$. Then,*

$$\frac{\|T_X - T_Y\|}{\|T_Y\|} \geq \max \left\{ \frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \left(\frac{1}{\kappa(\Sigma_Y)} \right)^{p/2-1}, \left(\frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \right)^{p/2} \right\},$$

where $\kappa(\Sigma_Y) := \lambda_{\max}(\Sigma_Y)/\lambda_{\min}(\Sigma_Y)$ is the condition number of Σ_Y . Similarly, under the entrywise maximum norm,

$$\frac{\|T_X - T_Y\|_{\max}}{\|T_Y\|_{\max}} \geq \max \left\{ \frac{\|D(\Sigma_X - \Sigma_Y)\|_{\max}}{\|D(\Sigma_Y)\|_{\max}} \left(\frac{1}{\kappa(D(\Sigma_Y))} \right)^{p/2-1}, \left(\frac{\|D(\Sigma_X - \Sigma_Y)\|_{\max}}{\|D(\Sigma_Y)\|_{\max}} \right)^{p/2} \right\},$$

where $D(\Sigma_X - \Sigma_Y)$ and $D(\Sigma_Y)$ denote the diagonal matrices with the same diagonal entries as $\Sigma_X - \Sigma_Y$ and Σ_Y , respectively.

3.2 Asymmetric case

Let p be an even integer, and let $X = (X^{(1)}, \dots, X^{(p)}) \in \mathbb{R}^d$ be a zero-mean Gaussian random vector with covariance matrix Σ , where each block $X^{(k)} \in \mathbb{R}^{d_k}$ and $\sum_{k=1}^p d_k = d$. For each $k \in \{1, \dots, p\}$, denote the marginal covariance by $\Sigma^{(k)} := \mathbb{E}X^{(k)} \otimes X^{(k)}$, and the cross-covariance by $\Sigma^{(j,k)} := \mathbb{E}X^{(j)} \otimes X^{(k)}$. Let X_1, \dots, X_N be i.i.d. copies of X , where $X_i = (X_i^{(1)}, \dots, X_i^{(p)})$. Our goal is to estimate the moment tensor

$$T := \mathbb{E}X^{(1)} \otimes \dots \otimes X^{(p)}, \quad (3.4)$$

formed by taking the tensor product over the blocks of X .

A natural estimator is the sample moment estimator, defined as

$$\hat{T}_S := \frac{1}{N} \sum_{i=1}^N X_i^{(1)} \otimes \cdots \otimes X_i^{(p)}. \quad (3.5)$$

Entrywise, this corresponds to

$$(\hat{T}_S)_{\ell_1, \dots, \ell_p} := \frac{1}{N} \sum_{i=1}^N X_{i, \ell_1}^{(1)} \cdots X_{i, \ell_p}^{(p)}.$$

Alternatively, by Isserlis's theorem, for any multi-index (ℓ_1, \dots, ℓ_p) ,

$$(\mathbb{E}X^{(1)} \otimes \cdots \otimes X^{(p)})_{\ell_1, \dots, \ell_p} = \mathbb{E}[X_{\ell_1}^{(1)} \cdots X_{\ell_p}^{(p)}] = \sum_{\pi \in \Pi_p^2} \prod_{(j,k) \in \pi} \Sigma_{\ell_j, \ell_k}^{(j,k)},$$

where Π_p^2 denotes the set of pairwise partitions of $\{1, \dots, p\}$. This motivates Isserlis's estimator, which substitutes the sample covariances

$$\hat{\Sigma}^{(j,k)} := \frac{1}{N} \sum_{i=1}^N X_i^{(j)} (X_i^{(k)})^\top$$

into the same expression. The resulting estimator is given entrywise by

$$(\hat{T}_I)_{\ell_1, \dots, \ell_p} := \sum_{\pi \in \Pi_p^2} \prod_{(j,k) \in \pi} \hat{\Sigma}_{\ell_j, \ell_k}^{(j,k)}. \quad (3.6)$$

Theorem 3.6 compares the performance of the two estimators, \hat{T}_S and \hat{T}_I , under both the operator norm and the entrywise maximum norm.

Theorem 3.6. *Let p be an even integer, and let $X = (X^{(1)}, \dots, X^{(p)}) \in \mathbb{R}^d$ be a zero-mean Gaussian random vector with covariance matrix Σ . For each $k \in \{1, \dots, p\}$, let $\Sigma^{(k)} := \mathbb{E}X^{(k)} \otimes X^{(k)}$, and for $j, k \in \{1, \dots, p\}$, let $\Sigma^{(j,k)} := \mathbb{E}X^{(j)} \otimes X^{(k)}$. Let X_1, \dots, X_N be i.i.d. copies of X , where $X_i = (X_i^{(1)}, \dots, X_i^{(p)})$. Let T, \hat{T}_S , and \hat{T}_I be as defined in equations (3.4), (3.5), and (3.6), respectively. Then the following bounds hold.*

(i) *Operator norm bounds:*

$$\begin{aligned} \mathbb{E}\|\hat{T}_S - T\| &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) \left(\left(\frac{\sum_{k=1}^p r_2(\Sigma^{(k)})}{N} \right)^{1/2} + \frac{\prod_{k=1}^p (r_2(\Sigma^{(k)}) + \log N)^{1/2}}{N} \right), \\ \mathbb{E}\|\hat{T}_I - T\| &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{1/2}, \text{ if } N \geq \max_{1 \leq k \leq p} r_2(\Sigma^{(k)}). \end{aligned}$$

(ii) *Entrywise maximum norm bounds:*

$$\begin{aligned} \mathbb{E}\|\hat{T}_S - T\|_{\max} &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|_{\max}^{1/2} \right) \left(\left(\frac{\sum_{k=1}^p r_{\max}(\Sigma^{(k)})}{N} \right)^{1/2} + \frac{\prod_{k=1}^p (r_{\max}(\Sigma^{(k)}) + \log N)^{1/2}}{N} \right), \\ \mathbb{E}\|\hat{T}_I - T\|_{\max} &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|_{\max}^{1/2} \right) \left(\frac{\max_{1 \leq k \leq p} r_{\max}(\Sigma^{(k)})}{N} \right)^{1/2}, \text{ if } N \geq \max_{1 \leq k \leq p} r_{\max}(\Sigma^{(k)}). \end{aligned}$$

Remark 3.7. The upper bounds on the deviation of sample moment estimator \hat{T}_S from its expectation are sharp under both the operator norm and entrywise maximum norm when the vectors $X^{(1)}, \dots, X^{(p)}$, together with their samples $(X_i^{(1)})_{i=1}^N, \dots, (X_i^{(p)})_{i=1}^N$, are

mutually independent; see [12, Theorem 2.1] and Theorem 6.1. Consistent estimation of T using the sample moment estimator \hat{T}_S requires a sample size satisfying

$$N \gg \prod_{k=1}^p (r_2(\Sigma^{(k)}) + \log N)^{1/2} \quad \text{or} \quad N \gg \prod_{k=1}^p (r_{\max}(\Sigma^{(k)}) + \log N)^{1/2}$$

under the operator norm and the entrywise maximum norm, respectively. In contrast, Isserlis's estimator \hat{T}_I only requires sample size

$$N \gg \max_{1 \leq k \leq p} r_2(\Sigma^{(k)}) \quad \text{or} \quad N \gg \max_{1 \leq k \leq p} r_{\max}(\Sigma^{(k)})$$

for consistency under the respective norms, leading to a significant reduction in sample complexity.

Proof of Theorem 3.6. First, for the sample moment estimator \hat{T}_S , the upper bound on its deviation from T under the operator norm follows directly from [12, Theorem 2.1], while the corresponding bound under the entrywise maximum norm is given by Theorem 6.1 in this paper. It remains to analyze Isserlis's estimator \hat{T}_I .

We apply the upper bound in Proposition 3.8, which yields

$$\|\hat{T}_I - T\| \leq \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) (p-1)!! \cdot \frac{p}{2} \cdot \varepsilon_* (1 + \varepsilon_*)^{p/2-1},$$

where

$$\varepsilon_* := \max_{j \neq k} \frac{\|\hat{\Sigma}^{(j,k)} - \Sigma^{(j,k)}\|}{(\|\Sigma^{(j,j)}\| \|\Sigma^{(k,k)}\|)^{1/2}}.$$

To control ε_* , we use the bound from [12, Remark 2.1 and page 21] for the sample cross-covariance. For $j, k \in \{1, \dots, p\}$ and any $u \geq 0$, it holds with probability at least $1 - \exp(-u^2)$ that

$$\frac{\|\hat{\Sigma}^{(j,k)} - \Sigma^{(j,k)}\|}{(\|\Sigma^{(j,j)}\| \|\Sigma^{(k,k)}\|)^{1/2}} \lesssim u \left(\frac{r_2(\Sigma^{(j)}) + r_2(\Sigma^{(k)})}{N} \right)^{1/2} + u^2 \frac{(r_2(\Sigma^{(j)}) r_2(\Sigma^{(k)}))^{1/2}}{N}.$$

Applying a union bound over all $j \neq k$ yields, with probability at least $1 - p^2 \exp(-u^2)$,

$$\begin{aligned} \varepsilon_* &= \max_{j \neq k} \frac{\|\hat{\Sigma}^{(j,k)} - \Sigma^{(j,k)}\|}{(\|\Sigma^{(j,j)}\| \|\Sigma^{(k,k)}\|)^{1/2}} \\ &\lesssim \max_{j \neq k} \left\{ u \left(\frac{r_2(\Sigma^{(j)}) + r_2(\Sigma^{(k)})}{N} \right)^{1/2} + u^2 \frac{(r_2(\Sigma^{(j)}) r_2(\Sigma^{(k)}))^{1/2}}{N} \right\} \\ &\lesssim u \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{1/2} + u^2 \frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N}. \end{aligned}$$

Substituting this bound into the inequality for $\|\hat{T}_I - T\|$, we obtain, with probability at least $1 - p^2 \exp(-u^2)$:

$$\begin{aligned} \|\hat{T}_I - T\| &\leq \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) (p-1)!! \cdot \frac{p}{2} \cdot \varepsilon_* (1 + \varepsilon_*)^{p/2-1} \\ &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) (\varepsilon_* + \varepsilon_*^{p/2}) \end{aligned}$$

$$\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) \left(u \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{1/2} + u^p \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{p/2} \right).$$

Integrating the tail bound yields the following expectation bound:

$$\begin{aligned} \mathbb{E} \|\widehat{T}_I - T\| &\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) \left(\left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{1/2} + \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{p/2} \right) \\ &\asymp_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2} \right) \left(\frac{\max_{1 \leq k \leq p} r_2(\Sigma^{(k)})}{N} \right)^{1/2}, \quad \text{if } N \geq \max_{1 \leq k \leq p} r_2(\Sigma^{(k)}). \end{aligned}$$

An analogous argument yields the corresponding bound under the entrywise maximum norm. This completes the proof. \square

The following proposition is used in the proof of Theorem 3.6, and its proof is deferred to Section 5.

Proposition 3.8. *Let p be an even integer, and let $X = (X^{(1)}, \dots, X^{(p)})$ and $Y = (Y^{(1)}, \dots, Y^{(p)})$ be zero-mean Gaussian random vectors in \mathbb{R}^d , where each $X^{(k)}, Y^{(k)} \in \mathbb{R}^{d_k}$ and $d = \sum_{k=1}^p d_k$. For $j, k \in \{1, \dots, p\}$, denote the cross-covariance blocks by*

$$\Sigma_X^{(j,k)} = \mathbb{E} X^{(j)} \otimes X^{(k)}, \quad \Sigma_Y^{(j,k)} = \mathbb{E} Y^{(j)} \otimes Y^{(k)}.$$

Let

$$T_X := \mathbb{E} X^{(1)} \otimes \dots \otimes X^{(p)}, \quad T_Y := \mathbb{E} Y^{(1)} \otimes \dots \otimes Y^{(p)},$$

denote the moment tensors formed by taking tensor products over the blocks of X and Y , respectively. Then,

$$\|T_X - T_Y\| \leq \left(\prod_{k=1}^p \|\Sigma_Y^{(k,k)}\|^{1/2} \right) (p-1)!! \cdot \frac{p}{2} \cdot \varepsilon_* (1 + \varepsilon_*)^{p/2-1},$$

where

$$\varepsilon_* := \max_{j \neq k} \frac{\|\Sigma_X^{(j,k)} - \Sigma_Y^{(j,k)}\|}{(\|\Sigma_Y^{(j,j)}\| \|\Sigma_Y^{(k,k)}\|)^{1/2}}.$$

For the entrywise maximum norm, it similarly holds that

$$\|T_X - T_Y\|_{\max} \leq \left(\prod_{k=1}^p \|\Sigma_Y^{(k,k)}\|_{\max}^{1/2} \right) (p-1)!! \cdot \frac{p}{2} \cdot \bar{\varepsilon}_* (1 + \bar{\varepsilon}_*)^{p/2-1},$$

where

$$\bar{\varepsilon}_* := \max_{j \neq k} \frac{\|\Sigma_X^{(j,k)} - \Sigma_Y^{(j,k)}\|_{\max}}{(\|\Sigma_Y^{(j,j)}\|_{\max} \|\Sigma_Y^{(k,k)}\|_{\max})^{1/2}}.$$

Corollary 3.9. *Let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$, and let $T_X := \mathbb{E} X^{\otimes p}$, $T_Y := \mathbb{E} Y^{\otimes p}$. Then,*

$$\frac{\|T_X - T_Y\|}{\|T_Y\|} \leq \frac{p}{2} \cdot \frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \left(1 + \frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \right)^{p/2-1}.$$

Similarly, under the entrywise maximum norm,

$$\frac{\|T_X - T_Y\|_{\max}}{\|T_Y\|_{\max}} \leq \frac{p}{2} \cdot \frac{\|\Sigma_X - \Sigma_Y\|_{\max}}{\|\Sigma_Y\|_{\max}} \left(1 + \frac{\|\Sigma_X - \Sigma_Y\|_{\max}}{\|\Sigma_Y\|_{\max}} \right)^{p/2-1}.$$

Proof of Corollary 3.9. Apply Proposition 3.8 with $X^{(1)} = \dots = X^{(p)} =: X$ and $Y^{(1)} = \dots = Y^{(p)} =: Y$, so that $\Sigma_X^{(j,k)} =: \Sigma_X$ and $\Sigma_Y^{(j,k)} =: \Sigma_Y$ for all $j, k \in \{1, \dots, p\}$. The result then follows from computing the norm of the reference tensor:

$$\|T_Y\| = \|\mathbb{E}Y^{\otimes p}\| = \sup_{v \in \mathcal{S}^d} \mathbb{E}\langle Y, v \rangle^p = (p-1)!! \cdot \|\Sigma_Y\|^{p/2},$$

where the last equality uses the moment formula for centered Gaussian variables. The bound under the entrywise maximum norm follows analogously. \square

4 Proof of Proposition 3.5

Proof of Proposition 3.5. We begin by observing that

$$\begin{aligned} \|T_X - T_Y\| &= \sup_{v \in \mathcal{S}^d} |\langle \mathbb{E}X^{\otimes p} - \mathbb{E}Y^{\otimes p}, v^{\otimes p} \rangle| \\ &= \sup_{v \in \mathcal{S}^d} |\mathbb{E}\langle X, v \rangle^p - \mathbb{E}\langle Y, v \rangle^p| \\ &= (p-1)!! \sup_{v \in \mathcal{S}^d} \left| \langle \Sigma_X v, v \rangle^{p/2} - \langle \Sigma_Y v, v \rangle^{p/2} \right| \\ &= (p-1)!! \sup_{v \in \mathcal{S}^d} \left| \langle (\Sigma_X - \Sigma_Y)v, v \rangle \right| \left| \sum_{\ell=0}^{p/2-1} \langle \Sigma_X v, v \rangle^\ell \langle \Sigma_Y v, v \rangle^{p/2-1-\ell} \right|, \end{aligned}$$

where the third equality uses the moment formula for centered Gaussian variables. Taking the supremum over the first term and the infimum over the second yields

$$\begin{aligned} \|T_X - T_Y\| &\geq (p-1)!! \sup_{v \in \mathcal{S}^d} |\langle (\Sigma_X - \Sigma_Y)v, v \rangle| \cdot \inf_{v \in \mathcal{S}^d} \left| \sum_{\ell=0}^{p/2-1} \langle \Sigma_X v, v \rangle^\ell \langle \Sigma_Y v, v \rangle^{p/2-1-\ell} \right| \\ &\geq (p-1)!! \|\Sigma_X - \Sigma_Y\| \cdot \lambda_{\min}(\Sigma_Y)^{p/2-1}. \end{aligned}$$

Since $\|T_Y\| = (p-1)!! \|\Sigma_Y\|^{p/2}$, it follows that

$$\frac{\|T_X - T_Y\|}{\|T_Y\|} \geq \frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \cdot \left(\frac{1}{\kappa(\Sigma_Y)} \right)^{p/2-1}, \quad (4.1)$$

where $\kappa(\Sigma_Y) := \lambda_{\max}(\Sigma_Y)/\lambda_{\min}(\Sigma_Y)$ is the condition number of Σ_Y .

Next, let v be a unit eigenvector associated with the eigenvalue of $\Sigma_X - \Sigma_Y$ of largest magnitude, so that $\langle (\Sigma_X - \Sigma_Y)v, v \rangle = \pm \|\Sigma_X - \Sigma_Y\|$. If the sign is positive, we have

$$\begin{aligned} \|T_X - T_Y\| &\geq (p-1)!! \left| \langle \Sigma_X v, v \rangle^{p/2} - \langle \Sigma_Y v, v \rangle^{p/2} \right| \\ &= (p-1)!! \left| \left(\langle (\Sigma_X - \Sigma_Y)v, v \rangle + \langle \Sigma_Y v, v \rangle \right)^{p/2} - \langle \Sigma_Y v, v \rangle^{p/2} \right| \\ &= (p-1)!! \left(\left(\|\Sigma_X - \Sigma_Y\| + \langle \Sigma_Y v, v \rangle \right)^{p/2} - \langle \Sigma_Y v, v \rangle^{p/2} \right) \\ &\geq (p-1)!! \|\Sigma_X - \Sigma_Y\|^{p/2}. \end{aligned}$$

If the sign is negative, a symmetric argument gives the same bound with Σ_X and Σ_Y interchanged. Therefore,

$$\frac{\|T_X - T_Y\|}{\|T_Y\|} \geq \frac{(p-1)!! \|\Sigma_X - \Sigma_Y\|^{p/2}}{(p-1)!! \|\Sigma_Y\|^{p/2}} = \left(\frac{\|\Sigma_X - \Sigma_Y\|}{\|\Sigma_Y\|} \right)^{p/2}. \quad (4.2)$$

Combining (4.1) and (4.2) yields the desired lower bound in operator norm.

For the entrywise maximum norm, we first obtain a lower bound of the same form as in the operator norm case, except that the supremum is now taken over the standard basis vectors. Specifically,

$$\begin{aligned}\|T_X - T_Y\|_{\max} &= \sup_{v_k \in \mathcal{E}^d, 1 \leq k \leq p} |\langle T_X - T_Y, v_1 \otimes \cdots \otimes v_p \rangle| \\ &\geq \sup_{v \in \mathcal{E}^d} |\langle T_X - T_Y, v^{\otimes p} \rangle| \\ &= (p-1)!! \sup_{v \in \mathcal{E}^d} \left| \langle \Sigma_X v, v \rangle^{p/2} - \langle \Sigma_Y v, v \rangle^{p/2} \right|,\end{aligned}$$

where $\mathcal{E}^d = \{e_i\}_{i=1}^d$ denotes the standard basis of \mathbb{R}^d . Then, the desired lower bound for the entrywise maximum norm follows from the same argument as for the operator norm case, with only minor modifications. We omit the details for brevity. \square

5 Proof of Proposition 3.8

Proof of Proposition 3.8. For arbitrary vectors $v_1 \in \mathbb{R}^{d_1}, \dots, v_p \in \mathbb{R}^{d_p}$, we have

$$\begin{aligned}\langle T_X - T_Y, v_1 \otimes \cdots \otimes v_p \rangle &= \left\langle \mathbb{E}X^{(1)} \otimes \cdots \otimes X^{(p)} - \mathbb{E}Y^{(1)} \otimes \cdots \otimes Y^{(p)}, v_1 \otimes \cdots \otimes v_p \right\rangle \\ &= \mathbb{E} \prod_{k=1}^p \langle X^{(k)}, v_k \rangle - \mathbb{E} \prod_{k=1}^p \langle Y^{(k)}, v_k \rangle \\ &\stackrel{(i)}{=} \sum_{\pi \in \Pi_p^2} \prod_{(j,k) \in \pi} \mathbb{E} \langle X^{(j)}, v_j \rangle \langle X^{(k)}, v_k \rangle - \sum_{\pi \in \Pi_p^2} \prod_{(j,k) \in \pi} \mathbb{E} \langle Y^{(j)}, v_j \rangle \langle Y^{(k)}, v_k \rangle \\ &= \sum_{\pi \in \Pi_p^2} \left(\prod_{(j,k) \in \pi} \langle \Sigma_X^{(j,k)} v_j, v_k \rangle - \prod_{(j,k) \in \pi} \langle \Sigma_Y^{(j,k)} v_j, v_k \rangle \right).\end{aligned}$$

Here, (i) follows from Isserlis's theorem, and Π_p^2 denotes the set of pairwise partitions of $\{1, \dots, p\}$. Denoting by $\pi = (\pi(1), \dots, \pi(p))$ a fixed ordering of the indices in the pairing π , and applying the telescoping identity

$$a_1 \cdots a_{p/2} - b_1 \cdots b_{p/2} = \sum_{\ell=1}^{p/2} a_1 \cdots a_{\ell-1} (a_\ell - b_\ell) b_{\ell+1} \cdots b_{p/2},$$

we obtain

$$\begin{aligned}\langle T_X - T_Y, v_1 \otimes \cdots \otimes v_p \rangle &= \sum_{\pi \in \Pi_p^2} \left(\prod_{(j,k) \in \pi} \langle \Sigma_X^{(j,k)} v_j, v_k \rangle - \prod_{(j,k) \in \pi} \langle \Sigma_Y^{(j,k)} v_j, v_k \rangle \right) \\ &= \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left[\left(\prod_{s=1}^{\ell-1} \langle \Sigma_X^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \rangle \right) \right. \\ &\quad \times \left. \left(\langle \Sigma_X^{(\pi(2\ell-1), \pi(2\ell))} v_{\pi(2\ell-1)}, v_{\pi(2\ell)} \rangle - \langle \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell))} v_{\pi(2\ell-1)}, v_{\pi(2\ell)} \rangle \right) \left(\prod_{s=\ell+1}^{p/2} \langle \Sigma_Y^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \rangle \right) \right].\end{aligned}$$

To bound the operator norm, we take the supremum over $v_1 \in \mathcal{S}^{d_1}, \dots, v_p \in \mathcal{S}^{d_p}$,

$$\begin{aligned}\|T_X - T_Y\| &= \sup_{v_k \in \mathcal{S}^{d_k}, 1 \leq k \leq p} |\langle T_X - T_Y, v_1 \otimes \cdots \otimes v_p \rangle| \\ &= \sup_{v_k \in \mathcal{S}^{d_k}, 1 \leq k \leq p} \left| \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left(\prod_{s=1}^{\ell-1} \langle \Sigma_X^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \rangle \right) \right.\end{aligned}$$

$$\begin{aligned}
 & \times \left\langle \left(\Sigma_X^{(\pi(2\ell-1), \pi(2\ell))} - \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell))} \right) v_{\pi(2\ell-1)}, v_{\pi(2\ell)} \right\rangle \left(\prod_{s=\ell+1}^{p/2} \left\langle \Sigma_Y^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \right\rangle \right) \Bigg| \\
 & \leq \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left[\sup_{v_k \in \mathcal{S}^{d_k}, 1 \leq k \leq p} \left| \left(\prod_{s=1}^{\ell-1} \left\langle \Sigma_X^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \right\rangle \right) \right. \right. \\
 & \quad \times \left. \left. \left\langle \left(\Sigma_X^{(\pi(2\ell-1), \pi(2\ell))} - \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell))} \right) v_{\pi(2\ell-1)}, v_{\pi(2\ell)} \right\rangle \left(\prod_{s=\ell+1}^{p/2} \left\langle \Sigma_Y^{(\pi(2s-1), \pi(2s))} v_{\pi(2s-1)}, v_{\pi(2s)} \right\rangle \right) \right| \right] \\
 & = \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left[\left(\prod_{s=1}^{\ell-1} \left\| \Sigma_X^{(\pi(2s-1), \pi(2s))} \right\| \right) \right. \\
 & \quad \times \left. \left\| \Sigma_X^{(\pi(2\ell-1), \pi(2\ell))} - \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell))} \right\| \left(\prod_{s=\ell+1}^{p/2} \left\| \Sigma_Y^{(\pi(2s-1), \pi(2s))} \right\| \right) \right].
 \end{aligned}$$

For $j, k \in \{1, \dots, p\}$, we introduce the normalized deviation $\varepsilon^{(j,k)} := \frac{\|\Sigma_X^{(j,k)} - \Sigma_Y^{(j,k)}\|}{(\|\Sigma_Y^{(j,j)}\| \|\Sigma_Y^{(k,k)}\|)^{1/2}}$ and define $\varepsilon_* := \max_{j \neq k} \varepsilon^{(j,k)}$. Using that $\|\Sigma_Y^{(j,k)}\| \leq \|\Sigma_Y^{(j,j)}\|^{1/2} \|\Sigma_Y^{(k,k)}\|^{1/2}$, we obtain

$$\begin{aligned}
 \|T_X - T_Y\| & \leq \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left[\left(\prod_{s=1}^{\ell-1} \left\| \Sigma_X^{(\pi(2s-1), \pi(2s))} \right\| \right) \right. \\
 & \quad \times \left. \left\| \Sigma_X^{(\pi(2\ell-1), \pi(2\ell))} - \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell))} \right\| \left(\prod_{s=\ell+1}^{p/2} \left\| \Sigma_Y^{(\pi(2s-1), \pi(2s))} \right\| \right) \right] \\
 & \leq \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \left[\left(\prod_{s=1}^{\ell-1} \left\| \Sigma_Y^{(\pi(2s-1), \pi(2s-1))} \right\|^{1/2} \left\| \Sigma_Y^{(\pi(2s), \pi(2s))} \right\|^{1/2} (1 + \varepsilon^{(\pi(2s-1), \pi(2s))}) \right) \right. \\
 & \quad \times \left(\left\| \Sigma_Y^{(\pi(2\ell-1), \pi(2\ell-1))} \right\|^{1/2} \left\| \Sigma_Y^{(\pi(2\ell), \pi(2\ell))} \right\|^{1/2} \varepsilon^{(\pi(2\ell-1), \pi(2\ell))} \right) \\
 & \quad \times \left. \left(\prod_{s=\ell+1}^{p/2} \left\| \Sigma_Y^{(\pi(2s-1), \pi(2s-1))} \right\|^{1/2} \left\| \Sigma_Y^{(\pi(2s), \pi(2s))} \right\|^{1/2} \right) \right] \\
 & = \left(\prod_{k=1}^p \left\| \Sigma_Y^{(k,k)} \right\|^{1/2} \right) \sum_{\pi \in \Pi_p^2} \sum_{\ell=1}^{p/2} \varepsilon^{(\pi(2\ell-1), \pi(2\ell))} \prod_{s=1}^{\ell-1} (1 + \varepsilon^{(\pi(2s-1), \pi(2s))}).
 \end{aligned}$$

Using the identity

$$\sum_{\ell=1}^{p/2} a_\ell \prod_{s=1}^{\ell-1} (1 + a_s) = \prod_{\ell=1}^{p/2} (1 + a_\ell) - 1,$$

we conclude that

$$\begin{aligned}
 \|T_X - T_Y\| & \leq \left(\prod_{k=1}^p \left\| \Sigma_Y^{(k,k)} \right\|^{1/2} \right) \sum_{\pi \in \Pi_p^2} \left(\prod_{\ell=1}^{p/2} (1 + \varepsilon^{(\pi(2\ell-1), \pi(2\ell))}) - 1 \right) \\
 & \leq \left(\prod_{k=1}^p \left\| \Sigma_Y^{(k,k)} \right\|^{1/2} \right) (p-1)!! \left((1 + \varepsilon_*)^{p/2} - 1 \right) \\
 & \leq \left(\prod_{k=1}^p \left\| \Sigma_Y^{(k,k)} \right\|^{1/2} \right) (p-1)!! \cdot \frac{p}{2} \cdot \varepsilon_* (1 + \varepsilon_*)^{p/2-1},
 \end{aligned}$$

where we used the inequality $(1 + \varepsilon_*)^{p/2} - 1 \leq \frac{p}{2} \cdot \varepsilon_* (1 + \varepsilon_*)^{p/2-1}$.

The bound under the entrywise maximum norm follows from an analogous argument, replacing the supremum over the unit spheres $\mathcal{S}^{d_1}, \dots, \mathcal{S}^{d_p}$ with the supremum over the standard bases $\mathcal{E}^{d_k} = \{e_i\}_{i=1}^{d_k}$ for $1 \leq k \leq p$. This completes the proof. \square

6 Entrywise maximum norm bound for sample moment tensor

Theorem 6.1. For any integer $p \geq 2$ and $1 \leq k \leq p$, let $X^{(k)}, X_1^{(k)}, \dots, X_N^{(k)}$ be i.i.d. zero-mean Gaussian random vectors in \mathbb{R}^{d_k} with covariance matrix $\Sigma^{(k)}$. Then,

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{(1)} \otimes \dots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \dots \otimes X^{(p)} \right\|_{\max} \lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|_{\max}^{1/2} \right) \mathcal{E}_N((\Sigma^{(k)})_{k=1}^p),$$

where

$$\mathcal{E}_N((\Sigma^{(k)})_{k=1}^p) := \left(\frac{\sum_{k=1}^p r_{\max}(\Sigma^{(k)})}{N} \right)^{1/2} + \frac{1}{N} \prod_{k=1}^p \left(r_{\max}(\Sigma^{(k)}) + \log N \right)^{1/2}.$$

Moreover, the upper bound is sharp in the following two cases:

1. *Independent Components.* If $X^{(1)}, \dots, X^{(p)}, (X_i^{(1)})_{i=1}^N, \dots, (X_i^{(p)})_{i=1}^N$ are mutually independent, then

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{(1)} \otimes \dots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \dots \otimes X^{(p)} \right\|_{\max} \asymp_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|_{\max}^{1/2} \right) \mathcal{E}_N((\Sigma^{(k)})_{k=1}^p).$$

2. *Identical Components.* If $X^{(1)} = \dots = X^{(p)} = X$ and $X_i^{(1)} = \dots = X_i^{(p)} = X_i$ for all $1 \leq i \leq N$, $\Sigma^{(1)} = \dots = \Sigma^{(p)} = \Sigma$, then

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\|_{\max} \asymp_p \|\Sigma\|_{\max}^{p/2} \left(\sqrt{\frac{r_{\max}(\Sigma)}{N}} + \frac{(r_{\max}(\Sigma) + \log N)^{p/2}}{N} \right).$$

Furthermore, since $(\log N)^{p/2}/N \lesssim_p 1/\sqrt{N}$, this bound simplifies to

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\|_{\max} \asymp_p \|\Sigma\|_{\max}^{p/2} \left(\sqrt{\frac{r_{\max}(\Sigma)}{N}} + \frac{r_{\max}(\Sigma)^{p/2}}{N} \right).$$

Remark 6.2. The upper bound in Theorem 6.1 holds without requiring independence between the sequences $(X_i^{(k)})_{i=1}^N$ and $(X_i^{(k')})_{i=1}^N$ for $k \neq k'$; that is, no assumptions are made on the correlation structure across components. Moreover, the upper bound in Theorem 6.1 extends directly to sub-Gaussian settings. Dimension-dependent counterparts can also be derived by applying standard ε -net arguments in conjunction with the α -sub-exponential concentration inequality of [14].

Proof of Theorem 6.1. Upper bound By the definition of the entrywise maximum norm, we have

$$\begin{aligned} & \mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{(1)} \otimes \dots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \dots \otimes X^{(p)} \right\|_{\max} \\ &= \mathbb{E} \max_{v_k \in \mathcal{E}^{d_k}, 1 \leq k \leq p} \left| \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^p \langle X_i^{(k)}, v_k \rangle - \mathbb{E} \prod_{k=1}^p \langle X^{(k)}, v_k \rangle \right|, \end{aligned}$$

where $\mathcal{E}^{d_k} = \{e_i\}_{i=1}^{d_k}$ denotes the standard basis of \mathbb{R}^{d_k} . Let $\bar{\mathcal{E}}^{d_k} := \mathcal{E}^{d_k} \cup -\mathcal{E}^{d_k}$, and define $\mathcal{F}^{(k)} := \{\langle \cdot, v \rangle : v \in \bar{\mathcal{E}}^{d_k}\}$ for $1 \leq k \leq p$. Then the maximum can be upper bounded by

$$\mathbb{E} \sup_{f^{(k)} \in \mathcal{F}^{(k)}, 1 \leq k \leq p} \left| \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^p f^{(k)}(X_i^{(k)}) - \mathbb{E} \prod_{k=1}^p f^{(k)}(X^{(k)}) \right|.$$

Applying [12, Theorem 2.2], we obtain

$$\begin{aligned} & \mathbb{E} \sup_{f^{(k)} \in \mathcal{F}^{(k)}, 1 \leq k \leq p} \left| \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^p f^{(k)}(X_i^{(k)}) - \mathbb{E} \prod_{k=1}^p f^{(k)}(X^{(k)}) \right| \\ & \lesssim_p \left(\prod_{k=1}^p d_{\psi_2}(\mathcal{F}^{(k)}) \right) \left(\frac{\sum_{k=1}^p \bar{\gamma}(\mathcal{F}^{(k)}, \psi_2)}{\sqrt{N}} + \frac{\prod_{k=1}^p (\bar{\gamma}(\mathcal{F}^{(k)}, \psi_2) + (\log N)^{1/2})}{N} \right), \end{aligned} \quad (6.1)$$

where $\bar{\gamma}(\mathcal{F}^{(k)}, \psi_2) := \gamma(\mathcal{F}^{(k)}, \psi_2)/d_{\psi_2}(\mathcal{F}^{(k)})$. Here, $\gamma(\mathcal{F}, \psi_2)$ denotes Talagrand's generic chaining complexity of the function class \mathcal{F} [26, Definition 2.7.3], $d_{\psi_2}(\mathcal{F}) := \sup_{f \in \mathcal{F}} \|f\|_{\psi_2}$, and ψ_2 refers to the Orlicz norm with Orlicz function $\psi(x) = e^{x^2} - 1$, see e.g. [29, Definition 2.5.6].

Since $X^{(k)}$ is Gaussian with covariance $\Sigma^{(k)}$, the ψ_2 -norm of linear functionals is equivalent to the L_2 -norm. Hence,

$$d_{\psi_2}(\mathcal{F}^{(k)}) = \sup_{f^{(k)} \in \mathcal{F}^{(k)}} \|f^{(k)}\|_{\psi_2} \asymp \sup_{f^{(k)} \in \mathcal{F}^{(k)}} \|f^{(k)}\|_{L_2} = \sup_{v \in \bar{\mathcal{E}}^{d_k}} (\mathbb{E} \langle X^{(k)}, v \rangle^2)^{1/2} = \|\Sigma^{(k)}\|_{\max}^{1/2}.$$

For the generic chaining term, note that the canonical metric on $\bar{\mathcal{E}}^{d_k}$ is given by

$$d_{X^{(k)}}(u, v) := (\mathbb{E}(\langle X^{(k)}, u \rangle - \langle X^{(k)}, v \rangle)^2)^{1/2} = \langle u - v, \Sigma^{(k)}(u - v) \rangle^{1/2} = \|\langle \cdot, u \rangle - \langle \cdot, v \rangle\|_{L_2(\mu^{(k)})},$$

where $\mu^{(k)}$ is the law of $X^{(k)}$. By Talagrand's majorizing measure theorem [26, Theorem 2.10.1],

$$\gamma(\mathcal{F}^{(k)}, \psi_2) \asymp \gamma(\mathcal{F}^{(k)}, L_2) = \gamma(\bar{\mathcal{E}}^{d_k}, d_{X^{(k)}}) \asymp \mathbb{E} \sup_{u \in \bar{\mathcal{E}}^{d_k}} \langle X^{(k)}, u \rangle = \mathbb{E} \|X^{(k)}\|_{\infty}.$$

We conclude that

$$d_{\psi_2}(\mathcal{F}^{(k)}) \asymp \|\Sigma^{(k)}\|_{\max}^{1/2}, \quad \gamma(\mathcal{F}^{(k)}, \psi_2) \asymp \mathbb{E} \|X^{(k)}\|_{\infty},$$

and therefore,

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \cdots \otimes X^{(p)} \right\|_{\max} \lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|_{\max}^{1/2} \right) \mathcal{E}_N((\Sigma^{(k)})_{k=1}^p),$$

where

$$\mathcal{E}_N((\Sigma^{(k)})_{k=1}^p) := \left(\frac{\sum_{k=1}^p r_{\max}(\Sigma^{(k)})}{N} \right)^{1/2} + \frac{1}{N} \prod_{k=1}^p \left(r_{\max}(\Sigma^{(k)}) + \log N \right)^{1/2}$$

and $r_{\max}(\Sigma^{(k)}) := (\mathbb{E}_{X^{(k)} \sim \mathcal{N}(0, \Sigma^{(k)})} \|X^{(k)}\|_{\infty}^2) / \|\Sigma^{(k)}\|_{\max}$.

Lower bound The lower bounds follow from straightforward modifications of the argument in [12, Proposition 3.1] and [2, Proposition 3.1], respectively, by replacing the operator norm with the entrywise maximum norm. We omit the details for brevity. \square

References

- [1] P. Abdalla and R. Vershynin, *On the dimension-free concentration of simple tensors via matrix deviation*, arXiv preprint arXiv:2506.09333 (2025).
- [2] O. Al-Ghattas, J. Chen, and D. Sanz-Alonso, *Sharp concentration of simple random tensors*, arXiv preprint arXiv:2502.16916 (2025).
- [3] O. Al-Ghattas, J. Chen, D. Sanz-Alonso, and N. Waniorek, *Optimal estimation of structured covariance operators*, arXiv preprint arXiv:2408.02109 (2024). MR4890837
- [4] O. Al-Ghattas, J. Chen, D. Sanz-Alonso, and N. Waniorek, *Covariance operator estimation: Sparsity, lengthscale, and ensemble Kalman filters*, *Bernoulli* **31** (2025), no. 3, 2377–2402. MR4890837
- [5] O. Al-Ghattas and D. Sanz-Alonso, *Non-asymptotic analysis of ensemble Kalman updates: effective dimension and localization*, *Information and Inference: A Journal of the IMA* **13** (2024), no. 1, iaad043. MR4682756
- [6] O. Al-Ghattas and D. Sanz-Alonso, *Covariance operator estimation via adaptive thresholding*, *Stochastic Processes and their Applications* **189** (2025), 104705. MR4925087
- [7] A. Auddy, D. Xia, and M. Yuan, *Tensors in high-dimensional data analysis: Methodological opportunities and theoretical challenges*, *Annual Review of Statistics and Its Application* **12** (2025), 527–551. MR4879689
- [8] G. Ballard and T. G. Kolda, *Tensor Decompositions for Data Science*, Cambridge University Press, 2025. MR4905283
- [9] A. S. Bandeira, S. Gopi, H. Jiang, K. Lucca, and T. Rothvoss, *A geometric perspective on the injective norm of sums of random tensors*, arXiv preprint arXiv:2411.10633 (2024). MR4928476
- [10] X. Bi, X. Tang, Y. Yuan, Y. Zhang, and A. Qu, *Tensors in statistics*, *Annual Review of Statistics and Its Application* **8** (2021), 345–368. MR4243551
- [11] M. T. Boedihardjo, *Injective norm of random tensors with independent entries*, arXiv preprint arXiv:2412.21193 (2024).
- [12] J. Chen and D. Sanz-Alonso, *Sharp concentration of simple random tensors II: Asymmetry*, arXiv preprint arXiv:2505.24144 (2025).
- [13] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, *IEEE Signal Processing Magazine* **32** (2015), no. 2, 145–163.
- [14] F. Götze, H. Sambale, and A. Sinulis, *Concentration inequalities for polynomials in alpha-sub-exponential random variables*, *Electronic Journal of Probability* **26** (2021), 1–22. MR4247973
- [15] L. Isserlis, *On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables*, *Biometrika* **12** (1918), no. 1/2, 134–139.
- [16] V. Koltchinskii, *Asymptotic efficiency in high-dimensional covariance estimation*, *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, World Scientific, 2018, pp. 2903–2923. MR3966516
- [17] V. Koltchinskii, *Asymptotically efficient estimation of smooth functionals of covariance operators*, *Journal of the European Mathematical Society* **23** (2021), no. 3, 765–843. MR4210724
- [18] V. Koltchinskii, *Estimation of smooth functionals of covariance operators: Jackknife bias reduction and bounds in terms of effective rank*, *Annales de l’Institut Henri Poincaré (B) Probabilités et statistiques*, vol. 61, Institut Henri Poincaré, 2025, pp. 665–712. MR4863057
- [19] V. Koltchinskii and K. Lounici, *Concentration inequalities and moment bounds for sample covariance operators*, *Bernoulli* **23** (2017), no. 1, 110–133. MR3556768
- [20] V. Koltchinskii and M. Zhilova, *Estimation of smooth functionals in normal models: bias reduction and asymptotic efficiency*, *The Annals of Statistics* **49** (2021), no. 5, 2577–2610. MR4338376
- [21] L. H. Lim, *Tensors in computations*, *Acta Numerica* **30** (2021), 555–764. MR4298222
- [22] K. Lounici, *High-dimensional covariance matrix estimation with missing observations*, *Bernoulli* **20** (2014), no. 3, 1029–1058. MR3217437

- [23] P. McCullagh, *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability*, Chapman and Hall/CRC, 2018. MR0907286
- [24] H. Z. Munthe-Kaas, O. Verdier, and G. Vilmart, *A short proof of Isserlis' theorem*, arXiv preprint arXiv:2503.01588 (2025).
- [25] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Transactions on Signal Processing **65** (2017), no. 13, 3551–3582. MR3666587
- [26] M. Talagrand, *Upper and lower bounds for stochastic processes: Decomposition theorems*, vol. 60, Springer Nature, 2022. MR4381414
- [27] R. Tomioka and T. Suzuki, *Spectral norm of random tensors*, arXiv preprint arXiv:1407.1870 (2014).
- [28] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, arXiv preprint arXiv:1011.3027 (2010). MR2963170
- [29] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, vol. 47, Cambridge University Press, 2018. MR3837109
- [30] R. Vershynin, *Concentration inequalities for random tensors*, Bernoulli **26** (2020), no. 4, 3139–3162. MR4140540
- [31] G. Wick, *The evaluation of the collision matrix*, Physical Review **80** (1950), no. 2, 268. MR0038281
- [32] P. W. Zehar, *Invariance of maximum likelihood estimators*, The Annals of Mathematical Statistics **37** (1966), no. 3, 744. MR0193707
- [33] N. Zhivotovskiy, *Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle*, Electronic Journal of Probability **29** (2024), 1–28. MR4693860
- [34] Z. Zhou and Y. Zhu, *Sparse random tensors: Concentration, regularization and applications*, Electronic Journal of Statistics **15** (2021), no. 1, 2483–2516. MR4255324

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <https://imstat.org/shop/donation/>