Electronic Journal of Probability

Electron. J. Probab. **24** (2019), no. 133, 1–32.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP392

Moment inequalities for matrix-valued U-statistics of order 2*

Stanislav Minsker† Xiaohan Wei‡

Abstract

We present Rosenthal-type moment inequalities for matrix-valued U-statistics of order 2. As a corollary, we obtain new matrix concentration inequalities for U-statistics. One of our main technical tools, a version of the non-commutative Khintchine inequality for the spectral norm of the Rademacher chaos, could be of independent interest.

Keywords: U-statistics; moment inequalities; concentration inequalities; Khintchine inequality.
AMS MSC 2010: 60E15.
Submitted to EJP on August 1, 2018, final version accepted on November 7, 2019.

1 Introduction

Since being introduced by W. Hoeffding [16], U-statistics have become an active topic of research. Many classical results in estimation and testing are related to U-statistics; detailed treatment of the subject can be found in excellent monographs [7, 20, 30, 21]. A large body of research has been devoted to understanding the asymptotic behavior of real-valued U-statistics. Such asymptotic results, as well as moment and concentration inequalities, are discussed in the works [8, 7, 12, 14, 18, 11, 17], among others. The case of vector-valued and matrix-valued U-statistics received less attention; natural examples of matrix-valued U-statistics include various estimators of covariance matrices, such as the usual sample covariance matrix and the estimators based on Kendall's tau [37, 15].

Exponential and moment inequalities for Hilbert space-valued U-statistics have been developed in [2]. The goal of the present work is to obtain moment and concentration inequalities for generalized degenerate U-statistics of order 2 with values in the set of matrices with complex-valued entries equipped with the operator (spectral) norm. The emphasis is made on expressing the upper bounds in terms of *computable* parameters. Our results extend the matrix Rosenthal's inequality for the sums of independent random matrices due to Chen, Gittens and Tropp [5] (see also [19, 25]) to the framework of

*Authors gratefully acknowledge support by the National Science Foundation grants DMS-1712956 and CCF-1908905.

†Department of Mathematics, University of Southern California. E-mail: minsker@usc.edu
‡Department of Electrical Engineering, University of Southern California. E-mail: xiaohanw@usc.edu

U-statistics. As a corollary of our bounds, we deduce a variant of the Matrix Bernstein inequality for U-statistics of order 2.

We also discuss connections of our bounds with general moment inequalities for Banach space-valued U-statistics due to R. Adamczak [1], and leverage Adamczak's inequalities to obtain additional refinements and improvements of the results.

We note that U-statistics with values in the set of self-adjoint matrices have been considered in [6], however, most results in that work deal with the element-wise sup-norm, while we are primarily interested in results about the moments and tail behavior of the spectral norm of U-statistics. Another recent work [26] investigates robust estimators of covariance matrices based on U-statistics, but deals only with the case of non-degenerate U-statititistics that can be reduced to the study of independent sums.

The key technical tool used in our arguments is the extension of the non-commutative Khintchine's inequality (Lemma 3.3) which could be of independent interest.

2 Notation and background material

Given $A \in \mathbb{C}^{d_1\times d_2}$, $A^* \in \mathbb{C}^{d_2\times d_1}$ will denote the Hermitian adjoint of A. $\mathbb{H}^d \subset \mathbb{C}^{d\times d}$ stands for the set of all self-adjoint matrices. If $A = A^*$, we will write $\lambda_{\text{max}}(A)$ and $\lambda_{\text{min}}(A)$ for the largest and smallest eigenvalues of A.

Everywhere below, $\|\cdot\|$ stands for the spectral norm $\|A\| := \sqrt{\lambda_{\text{max}}(A^*A)}$. If $d_1 = d_2 = d$, we denote by $\text{tr}(A)$ the trace of A. The Schatten p-norm of a matrix A is defined as $\|A\|_{S_p} = \left(\text{tr}(A^*A)^{p/2}\right)^{1/p}$. When $p = 1$, the resulting norm is called the nuclear norm and will be denoted by $\|\cdot\|_*$. The Schatten 2-norm is also referred to as the Frobenius norm or the Hilbert-Schmidt norm, and is denoted by $\|\cdot\|_{\text{F}}$; and the associated inner product is $\langle A_1, A_2\rangle = \text{tr}(A_1^*A_2)$.

Given $z \in \mathbb{C}^d$, $\|z\|_2 = \sqrt{z^*z}$ stands for the usual Euclidean norm of z. Let $A,\ B \in \mathbb{H}^d$. We will write $A \succeq B (\text{or} A \succ B)$ iff $A - B$ is nonnegative (or positive) definite. For $a, b \in \mathbb{R}$, we set $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. We use C to denote absolute constants that can take different values in various places.

Finally, we introduce the so-called Hermitian dilation which is a tool that often allows to reduce the problems involving general rectangular matrices to the case of Hermitian matrices.

Definition 2.1. *Given a rectangular matrix $A \in \mathbb{C}^{d_1\times d_2}$, the Hermitian dilation $\mathcal{D}$: $\mathbb{C}^{d_1\times d_2} \mapsto \mathbb{C}^{(d_1+d_2)\times(d_1+d_2)}$ is defined as*

$$\mathcal{D}(A) = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}. \tag{2.1}$$

Since $\mathcal{D}(A)^2 = \begin{pmatrix} AA^* & 0 \\ 0 & A^*A \end{pmatrix}$, it is easy to see that $\|\mathcal{D}(A)\| = \|A\|$.

The rest of the paper is organized as follows. Section 2.1 contains the necessary background on U-statistics. Section 3 contains our main results – bounds on the $\mathbb{H}^d$-valued Rademacher chaos and moment inequalities for $\mathbb{H}^d$-valued U-statistics of order 2. Section 4 provides comparison of our bounds to relevant results in the literature, and discusses further improvements. Finally, Section 5 contains the technical background and proofs of the main results.

2.1 Background on U-statistics

Consider a sequence of i.i.d. random variables $X_1, \ldots, X_n$ ($n \geq 2$) taking values in a measurable space $(\mathcal{S}, \mathcal{B})$, and let P denote the distribution of X_1. Define

$$I_n^m := \{(i_1, \ldots, i_m) :\ 1 \leq i_j \leq n, ,\ i_j \neq i_k \text{ if } j \neq k\},$$

and assume that $H_{i_1,\ldots,i_m} : \mathcal{S}^m \to \mathbb{H}^d$, $(i_1,\ldots,i_m) \in I_n^m$, $2 \le m \le n$, are $\mathcal{S}^m$-measurable, permutation-symmetric kernels, meaning that $H_{i_1,\ldots,i_m}(x_1,\ldots,x_m) = H_{i_{\pi_1},\ldots,i_{\pi_m}}(x_{\pi_1},\ldots,x_{\pi_m})$ for any $(x_1,\ldots,x_m) \in \mathcal{S}^m$ and any permutation π. For example, when $m = 2$, this conditions reads as $H_{i_1,i_2}(x_1,x_2) = H_{i_2,i_1}(x_2,x_1)$ for all $i_1 \ne i_2$ and x_1, x_2. The generalized U-statistic is defined as [7]

$$U_n := \sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m}). \tag{2.2}$$

When $H_{i_1,\ldots,i_m} \equiv H$, we obtain the classical U-statistics. It is often easier to work with the decoupled version of U_n defined as

$$U_n' = \sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}\left(X_{i_1}^{(1)},\ldots,X_{i_m}^{(m)}\right),$$

where $\left\{X_i^{(k)}\right\}_{i=1}^n$, $k = 1,\ldots,m$ are independent copies of the sequence $X_1,\ldots,X_n$. Our ultimate goal is to obtain the moment and deviation bounds for the random variable $\|U_n - \mathbb{E}U_n\|$.

Next, we recall several useful facts about U-statistics. The projection operator $\pi_{m,k}$ $(k \le m)$ is defined as

$$\pi_{m,k}H(\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_k}) := (\delta_{\mathbf{x}_{i_1}} - P)\ldots(\delta_{\mathbf{x}_{i_k}} - P)P^{m-k}H,$$

where

$$\mathcal{Q}^m H := \int \ldots \int H(\mathbf{y}_1,\ldots,\mathbf{y}_m)dQ(\mathbf{y}_1)\ldots dQ(\mathbf{y}_m),$$

for any probability measure Q on $(\mathcal{S},\mathcal{B})$, and δ_x is a Dirac measure concentrated at $x \in \mathcal{S}$. For example, $\pi_{m,1}H(x) = \mathbb{E}\left[H(X_1,\ldots,X_m)|X_1 = x\right] - \mathbb{E}H(X_1,\ldots,X_m)$.

Definition 2.2. *Let $F : \mathcal{S}^m \to \mathbb{H}^d$ be a measurable function. We will say that F is P-degenerate of order r $(1 \le r < m)$ iff*

$$\mathbb{E}F(\mathbf{x}_1,\ldots,\mathbf{x}_r,X_{r+1},\ldots,X_m) = 0 \;\forall \mathbf{x}_1,\ldots,\mathbf{x}_r \in \mathcal{S},$$

and $\mathbb{E}F(\mathbf{x}_1,\ldots,\mathbf{x}_r,\mathbf{x}_{r+1},X_{r+2},\ldots,X_m)$ is not a constant function. Otherwise, F is non-degenerate.

For instance, it is easy to check that $\pi_{m,k}H$ is degenerate of order $k-1$. If F is degenerate of order $m-1$, then it is called *completely degenerate*. From now on, we will only consider generalized U-statistics of order $m = 2$ with completely degenerate (that is, degenerate of order 1) kernels. The case of non-degenerate U-statistics is easily reduced to the degenerate case via the *Hoeffding's decomposition;* see page 137 in [7] for the details.

3 Main results

Rosenthal-type moment inequalities for sums of independent matrices have appeared in a number of previous works, including [5, 25, 31]. For example, the following inequality follows from Theorem A.1 in [5]:

Lemma 3.1 (Matrix Rosenthal inequality)**.** *Suppose that $q \ge 1$ is an integer and fix $r \ge q \vee \log d$. Consider a finite sequence of $\{\mathbf{Y}_i\}$ of independent $\mathbb{H}^d$-valued random* *matrices. Then*

$$\left(\mathbb{E}\left\|\sum_i (\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i)\right\|^{2q}\right)^{1/2q} \le 2\sqrt{er}\left\|\left(\sum_i \mathbb{E}\left(\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i\right)^2\right)^{1/2}\right\| + 4\sqrt{2}er\left(\mathbb{E}\max_i \|\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i\|^{2q}\right)^{1/2q}. \quad (3.1)$$

The bound above improves upon the moment inequality that follows from the matrix Bernstein's inequality (see Theorem 1.6.2 in [31]):

Lemma 3.2 (Matrix Bernstein's inequality)**.** *Consider a finite sequence of $\{\mathbf{Y}_i\}$ of independent $\mathbb{H}^d$-valued random matrices such that $\|\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i\| \le B$ almost surely. Then*

$$\Pr\left(\left\|\sum_i (\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i)\right\| \ge 2\sigma\sqrt{u} + \frac{4}{3}Bu\right) \le 2de^{-u},$$

where $\sigma^2 := \left\|\sum_i \mathbb{E}\left(\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i\right)^2\right\|$.

Indeed, Lemma 5.8 implies, with $a_0 = C\left(\sigma\sqrt{\log(2d)} + B\log(2d)\right)$ for some absolute constant $C > 0$ and after some simple algebra, that

$$\left(\mathbb{E}\left\|\sum_i (\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i)\right\|^{q}\right)^{1/q} \le C_2\left(\sqrt{q + \log(2d)}\,\sigma + (q + \log(2d))B\right),$$

for an absolute constant $C_2 > 0$ and all $q \ge 1$. This bound is weaker than (3.1) as it requires almost sure boundedness of $\|\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_i\|$ for all i. One the the main goals of this work is to obtain operator norm bounds similar to inequality (3.1) for $\mathbb{H}^d$-valued U-statistics of order 2.

3.1 Degenerate U-statistics of order 2

Moment bounds for scalar U-statistics are well-known, see for example the work [12] and references therein. Moreover, in [1], author obtained moment inequalities for general Banach-space valued U-statistics. Here, we aim at improving these bounds for the special case of $\mathbb{H}^d$-valued U-statistics of order 2. We discuss connections and provide comparison of our results with the bounds obtained by R. Adamczak [1] in Section 4.

3.2 Matrix Rademacher chaos

The starting point of our investigation is a moment bound for the matrix Rademacher chaos of order 2. This bound generalizes the spectral norm inequality for the matrix Rademacher series, see [31, 34, 35, 36]. We recall Khintchine's inequality for the matrix Rademacher series for the ease of comparison: let $A_1, \ldots, A_n \in \mathbb{H}^d$ be a sequence of fixed matrices, and $\varepsilon_1, \ldots, \varepsilon_n$ – a sequence of i.i.d. Rademacher random variables. Then

$$\left(\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i A_i\right\|^2\right)^{1/2} \le \sqrt{e(1 + 2\log d)} \cdot \left\|\sum_{i=1}^n A_i^2\right\|^{1/2}. \quad (3.2)$$

Furthermore, Jensen's inequality implies this bound is tight (up to a logarithmic factor). Note that the expected norm of $\sum \varepsilon_i A_i$ is controlled by the single "matrix variance" parameter $\left\|\sum_{i=1}^n A_i^2\right\|$. Next, we state the main result of this section, the analogue of inequality (3.2) for the Rademacher chaos of order 2.

Lemma 3.3. *Let $\{A_{i_1,i_2}\}_{i_1,i_2=1}^n \in \mathbb{H}^d$ be a sequence of fixed matrices. Assume that $\left\{\varepsilon_j^{(i)}\right\}_{j\in\mathbb{N}}$, $i = 1,2$, are two independent sequences of i.i.d. Rademacher random variables, and define*

$$X = \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}.$$

Then for any $q \geq 1$,

$$\max\left\{\|GG^*\|, \left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\|\right\}^{1/2} \leq \left(\mathbb{E}\|X\|^{2q}\right)^{1/(2q)}$$
$$\leq \frac{4}{\sqrt{e}}\cdot r\cdot \max\left\{\|GG^*\|, \left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\|\right\}^{1/2}, \quad (3.3)$$

where $r := q \vee \log d$, and the matrix $G \in \mathbb{H}^{nd}$ is defined via its block structure as

$$G := \begin{pmatrix} 0 & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & 0 & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \dots & 0 \end{pmatrix}. \quad (3.4)$$

Remark 3.4 (Constants in Lemma 3.3)**.** Matrix Rademacher chaos of order 2 has been studied previously in [29], [27] and [28], where Schatten-p norm upper bounds were obtained by iterating Khintchine's inequality for Rademacher series. Specifically, the following bound holds for all $p \geq 1$ (see Lemma 5.4 for the details):

$$\mathbb{E}\,\|X\|_{S_{2p}}^{2p} \leq 2\left(\frac{2\sqrt{2}}{e}p\right)^{2p} \max\left\{\left\|(GG^*)^{1/2}\right\|_{S_{2p}}^{2p}, \left\|\left(\sum_{i_1,i_2=1}^n A_{i_1,i_2}^2\right)^{1/2}\right\|_{S_{2p}}^{2p}\right\}.$$

Using the fact that for any $B \in \mathbb{H}^d$, $\|B\| \leq \|B\|_{S_{2p}} \leq d^{1/2p}\|B\|$ and taking $p = q \vee \log(nd)$, one could obtain a "naïve" extension of the inequality above, namely

$$\left(\mathbb{E}\|X\|^{2q}\right)^{1/(2q)} \leq C\max\left(q, \log(nd)\right)\max\left\{\|GG^*\|, \left\|\sum_{(i_1,i_2)\in I_2^n} A_{i_1,i_2}^2\right\|\right\}^{1/2}$$

that contains an extra $\log(n)$ factor which is removed in Lemma 3.3.

One may wonder if the term $\|GG^*\|$ in Lemma 3.3 is redundant. For instance, in the case when $\{A_{i_1,i_2}\}_{i_1,i_2}$ are scalars, it is easy to see $\left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\| \geq \|GG^*\|$. However, a more careful examination shows that there is no strict dominance among $\|GG^*\|$ and $\left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\|$. The following example presents a situation where $\left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\| < \|GG^*\|$.

Example 3.5. Assume that $d \geq n \geq 2$, let $\{\mathbf{a}_1,\dots,\mathbf{a}_d\}$ be any orthonormal basis in $\mathbb{R}^d$, and $\mathbf{a} := [\mathbf{a}_1^T,\dots,\mathbf{a}_n^T]^T \in \mathbb{R}^{nd}$ be the "vertical concatenation" of $\mathbf{a}_1,\dots,\mathbf{a}_d$. Define

$$A_{i_1,i_2} := \mathbf{a}_{i_1}\mathbf{a}_{i_2}^T + \mathbf{a}_{i_2}\mathbf{a}_{i_1}^T,\quad i_1,i_2 \in \{1,2,\dots,n\},$$

and

$$X := \sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} A_{i_1,i_2}.$$

Then $\|GG^*\| = \left\|GG^T\right\| \geq (n-2)\|\mathbf{a}\|_2^2 = (n-2)n$, and $\left\|\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2}\right\| = 2(n-1)$. Details are outlined in Section 5.4.

It follows from Lemma 5.1 that

$$\|GG^*\| \leq \sum_{i_1}\left\|\sum_{i_2:i_2\neq i_1} A^2_{i_1,i_2}\right\|. \tag{3.5}$$

Often, this inequality yields a "computable" upper bound for the right-hand side of the inequality (3.3), however, in some cases it results in the loss of precision, as the following example demonstrates.

Example 3.6. Assume that n is even, $d \geq n \geq 2$, let $\{\mathbf{a}_1,\ldots,\mathbf{a}_d\}$ be an orthonormal basis in $\mathbb{R}^d$, and let $\mathcal{C}\in\mathbb{R}^{n\times n}$ be an orthogonal matrix with entries $c_{i,j}$ such that $c_{i,i}=0$ for all i. Define

$$A_{i_1,i_2} = c_{i_1,i_2}\left(\mathbf{a}_{i_1}\mathbf{a}_{i_2}^T + \mathbf{a}_{i_2}\mathbf{a}_{i_1}^T\right),\ \ i_1,i_2\in\{1,2,\ldots,n\},$$

and $X := \sum_{(i_1,i_2)\in I_n^2} \varepsilon^{(1)}_{i_1}\varepsilon^{(2)}_{i_2} A_{i_1,i_2}$. Then $\|GG^*\|=1$, $\left\|\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2}\right\| = 2$, but

$$\sum_{i_1}\left\|\sum_{i_2:i_2\neq i_1} A^2_{i_1,i_2}\right\| = n.$$

Details are outlined in Section 5.4.

3.3 Moment inequalities for degenerate U-statistics of order 2

Let $H_{i_1,i_2}:\mathcal{S}\times\mathcal{S}\mapsto\mathbb{H}^d$, $(i_1,i_2)\in I_n^2$, be a sequence of degenerate kernels, for example, $H_{i_1,i_2}(x_1,x_2) = \pi_{2,2}\widehat{H}_{i_1,i_2}(x_1,x_2)$ for some non-degenerate permutation-symmetric $\widehat{H}_{i_1,i_2}$. Recall that U_n, the generalized U-statistic of order 2, has the form

$$U_n := \sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}(X_{i_1},X_{i_2}).$$

Everywhere below, $\mathbb{E}_j[\cdot]$, $j=1,2$, stands for the expectation with respect to $\left\{X_i^{(j)}\right\}_{i=1}^n$ only (that is, conditionally on all other random variables). The following Theorem is our most general result; it can be used as a starting point to derive more refined bounds.

Theorem 3.7. *Let $\left\{X_i^{(j)}\right\}_{i=1}^n$, $j=1,2$, be $\mathcal{S}$-valued i.i.d. random variables, $H_{i,j}$: $\mathcal{S}\times\mathcal{S}\mapsto\mathbb{H}^d$ – permutation-symmetric degenerate kernels. Then for all $q\geq 1$ and $r=\max(q,\log(ed))$,*

$$\begin{aligned}\left(\mathbb{E}\|U_n\|^{2q}\right)^{1/2q} \leq& 4\left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^{2q}\right)^{1/2q}\\ \leq& 128/\sqrt{e}\left[16r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H^2_{i_1,i_2}\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)}\right.\\ &\left.+r\left\|\sum_{(i_1,i_2)\in I_n^2}\mathbb{E}H^2_{i_1,i_2}\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^{1/2} + r\left(\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\|^q\right)^{1/2q}\right],\end{aligned}$$

where the matrix $\widetilde{G} \in \mathbb{H}^{nd}$ *is defined as*

$$\widetilde{G} := \begin{pmatrix} 0 & H_{1,2}\left(X_1^{(1)}, X_2^{(2)}\right) & \dots & H_{1,n}\left(X_1^{(1)}, X_n^{(2)}\right) \\ H_{2,1}\left(X_2^{(1)}, X_1^{(2)}\right) & 0 & \dots & H_{2,n}\left(X_2^{(1)}, X_n^{(2)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ H_{n,1}\left(X_n^{(1)}, X_1^{(2)}\right) & H_{n,2}\left(X_n^{(1)}, X_2^{(2)}\right) & \dots & 0 \end{pmatrix}. \tag{3.6}$$

Proof. See Section 5.2.3. □

The following lower bound (proven in Section 5.2.4) demonstrates that all the terms in the bound of Theorem 3.7 are necessary.

Lemma 3.8. *Under the assumptions of Theorem 3.7,*

$$\left(\mathbb{E}\,\|U_n\|^{2q}\right)^{1/2q} \geq C\left[\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2: i_2\neq i_1} H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)} + \left(\mathbb{E}\left\|\mathbb{E}_2 \widetilde{G}\widetilde{G}^*\right\|^q\right)^{1/2q} + \left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^q\right)^{1/2q}\right]$$

where $C > 0$ *is an absolute constant.*

Example 3.9. Let $\{A_{i_1,i_2}\}_{1\leq i_1<i_2\leq n}$ be fixed elements of $\mathbb{H}^d$ and $X_1,\dots,X_n$ – centered i.i.d. real-valued random variables such that $\mathrm{Var}(X_1) = 1$. Consider $\mathbf{Y} := \sum_{i_1\neq i_2} A_{i_1,i_2} X_{i_1} X_{i_2}$, where $A_{i_2,i_1} = A_{i_1,i_2}$ for $i_2 > i_1$. We will apply Theorem 3.7 to obtain the bounds for $\left(\mathbb{E}\|\mathbf{Y}\|^{2q}\right)^{1/2q}$. In this case, $H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) = A_{i_1,i_2} X_{i_1}^{(1)} X_{i_2}^{(2)}$, and it is easy to see that

$$\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2: i_2\neq i_1} H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)} \leq \max_i \left\|\sum_{j\neq i} A^2_{i,j}\right\|^{1/2} \left(\mathbb{E}\max_{1\leq i\leq n}|X_i|^{2q}\right)^{1/q}$$

and $\left\|\sum_{(i_1,i_2)\in I_n^2} \mathbb{E} H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^{1/2} = \left\|\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2}\right\|^{1/2}$. Moreover,

$$\left(\mathbb{E}_2 \widetilde{G}\widetilde{G}^*\right)_{i,j} = X_i^{(1)} X_j^{(1)} \sum_{k\neq i,j} A_{i,k} A_{j,k},$$

implying that $\mathbb{E}_2 \widetilde{G}\widetilde{G}^* = D\,G\,D$, where G is defined as in (3.4) and $D \in \mathbb{H}^{nd}$ is a diagonal matrix $D = \mathrm{diag}(X_1^{(1)},\dots,X_n^{(1)}) \otimes I_d$, where $\otimes$ denotes the Kronecker product. It yields that $\left\|\mathbb{E}_2 \widetilde{G}\widetilde{G}^*\right\| \leq \max_i \left|X_i^{(1)}\right|^2 \cdot \|GG^*\|$, hence

$$\left(\mathbb{E}\left\|\mathbb{E}_2 \widetilde{G}\widetilde{G}^*\right\|^q\right)^{1/2q} \leq \|GG^*\|^{1/2}\left(\mathbb{E}\max_{1\leq i\leq n}|X_i|^{2q}\right)^{1/2q}.$$

Combining the inequalities above, we deduce from Theorem 3.7 that

$$\left(\mathbb{E}\|\mathbf{Y}\|^{2q}\right)^{1/2q} \leq C\left[r\left(\left\|\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2}\right\|^{1/2} + \left(\mathbb{E}\max_{1\leq i\leq n}|X_i|^{2q}\right)^{1/2q}\|GG^*\|^{1/2}\right) + r^{3/2}\max_{i_1}\left\|\sum_{i_2\neq i_1} A^2_{i_1,i_2}\right\|^{1/2}\left(\mathbb{E}\max_{1\leq i\leq n}|X_i|^{2q}\right)^{1/q}\right], \tag{3.7}$$

where $r = \max(q, \log(ed))$. If for instance $|X_1| \le M$ almost surely for some $M \ge 1$, it follows that

$$\left(\mathbb{E}\|\mathbf{Y}\|^{2q}\right)^{1/2q} \le C\left[r\left(M\ \|GG^*\|^{1/2} + \left\| \sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2} \right\|^{1/2} \right) + r^{3/2} M^2 \max_{i_1} \left\| \sum_{i_2 \ne i_1} A^2_{i_1,i_2} \right\|^{1/2} \right].$$

On the other hand, if X_1 is not bounded but is sub-Gaussian, meaning that $\left(\mathbb{E}|X_1|^q\right)^{1/q} \le C\sigma\sqrt{q}$ for all $q \in \mathbb{N}$ and some $\sigma > 0$, then it is easy to check that

$$\left(\mathbb{E} \max_{1\le i\le n} |X_i|^{2q}\right)^{1/2q} \le C_1\sqrt{\log(n)}\sigma\sqrt{2q},$$

and the estimate for $\left(\mathbb{E}\,\|\mathbf{Y}\|^{2q}\right)^{1/2q}$ follows from (3.7).

Our next goal is to obtain more "user-friendly" versions of the upper bound, and we first focus on the term $\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\|^q$ appearing in Theorem 3.7 that might be difficult to deal with directly. It is easy to see that the (i,j)-th block of the matrix $\mathbb{E}_2\widetilde{G}\widetilde{G}^*$ is

$$\left(\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right)_{i,j} = \sum_{k\ne i,j} \mathbb{E}_2\left[H_{i,k}(X_i^{(1)}, X_k^{(2)}) H_{j,k}(X_j^{(1)}, X_k^{(2)}) \right].$$

It follows from Lemma 5.1 that

$$\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\| \le \sum_i \left\| \left(\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right)_{i,i} \right\| = \sum_{i_1} \left\| \sum_{i_2 : i_2\ne i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\|, \tag{3.8}$$

hence

$$\begin{aligned} \left(\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\|^q\right)^{1/2q} &\le \left(\mathbb{E}\left(\sum_{i_1} \left\| \sum_{i_2:i_2\ne i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\| \right)^q \right)^{1/2q} \\ &\le \left(\sum_{i_1} \mathbb{E} \left\| \sum_{i_2:i_2\ne i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\| \right)^{1/2} \\ &\qquad + 2\sqrt{2eq}\left(\mathbb{E}\max_{i_1} \left\| \sum_{i_2:i_2\ne i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\|^q \right)^{1/2q}, \end{aligned}$$

where we used Rosenthal's inequality (Lemma 5.5 applied with $d = 1$) in the last step. Together with the fact that $\left\|\mathbb{E}H^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\| \le \mathbb{E}\left\|\mathbb{E}_2 H^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|$ for all i_1, i_2, and the inequality

$$\begin{aligned} &\left(\mathbb{E}\max_{i_1} \left\| \sum_{i_2:i_2\ne i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\|^q \right)^{1/2q} \\ &\le \left(\mathbb{E}\max_{i_1} \left\| \sum_{i_2:i_2\ne i_1} H^2_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right) \right\|^q \right)^{1/(2q)}, \end{aligned}$$

we obtain the following result.

Corollary 3.10. *Under the assumptions of Theorem 3.7,*

$$\left(\mathbb{E}\left\|U_n\right\|^{2q}\right)^{1/2q} \le 256/\sqrt{e}\Bigg[r\left(\sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2} + 11\,r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/(2q)}\Bigg].$$

Remark 3.11. Assume that $H_{i,j} = H$ is independent of i, j and is such that $\|H(x_1, x_2)\| \le M$ for all $x_1, x_2 \in S$. Then

$$\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q \le (n-1)^q M^{2q},$$

and it immediately follows from Lemma 5.7 and Corollary 3.10 that for all $t \ge 1$ and an absolute constant $C > 0$,

$$\Pr\left(\|U_n\| \ge C\left(\sqrt{\mathbb{E}\left\|\mathbb{E}_2 H^2(X^{(1)}_1,X^{(2)}_2)\right\|}\,(t+\log d)\cdot n + M\sqrt{n}\,(t+\log d)^{3/2}\right)\right) \le e^{-t}. \tag{3.9}$$

Next, we obtain further refinements of the result that follow from estimating the term

$$r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/(2q)}.$$

Lemma 3.12. *Under the assumptions of Theorem 3.7,*

$$\begin{aligned}
&r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/(2q)}\\
&\le 4e\sqrt{2}\sqrt{1+\frac{\log d}{q}}\Bigg[r\left(\sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}\\
&\quad + r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/2q}\\
&\quad + r^2\left(\sum_{i_1}\mathbb{E}\max_{i_2:i_2\neq i_1}\left\|H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/2q}\Bigg].
\end{aligned}$$

Proof. See Section 5.2.5. ☐

One of the key features of the bounds established above is the fact that they yield estimates for $\mathbb{E}\|U_n\|$: for example, Theorem 3.7 implies that

$$\mathbb{E}\|U_n\| \le C\log d\Bigg(\left(\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\|\right)^{1/2} + \left\|\sum_{(i_1,i_2)\in I_n^2}\mathbb{E}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{1/2} + \sqrt{\log d}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}\Bigg) \tag{3.10}$$

for some absolute constant C. On the other hand, direct application of the non-commutative Khintchine's inequality (3.2) followed by Rosenthal's inequality (Lemma 5.5) only gives that

$$\begin{aligned}\mathbb{E}\|U_n\| \leq & C\log d\left(\sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1} H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2} \\ \leq & C\log d\Bigg(\left(\sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2} \\ & + \sqrt{\log d}\left(\sum_{i_1}\mathbb{E}\max_{i_2:i_2\neq i_1}\left\|H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}\Bigg),\end{aligned} \tag{3.11}$$

and it is easy to see that the right-hand side of (3.10) is never worse than the bound (3.11). To verify that it can be strictly better, consider the framework of Example 3.6, where it is easy to check (following the same calculations as those given in Section 5.4) that

$$\left(\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\|\right)^{1/2}=1,\quad \left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}=1,$$
$$\left\|\sum_{(i_1,i_2)\in I^2_n}\mathbb{E}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{1/2}=2,$$

while $\left(\sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}=\sqrt{n}$.

Remark 3.13 (Extensions to rectangular matrices)**.** All results in this section can be extended to the general case of $\mathbb{C}^{d_1\times d_2}$-valued kernels by considering the Hermitian dilation $\mathcal{D}(U_n)$ of U_n as defined in (2.1), namely

$$\mathcal{D}(U_n)=\sum_{(i_1,i_2)\in I^2_n}\mathcal{D}\left(H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right)\in\mathbb{H}^{d_1+d_2},$$

and observing that $\|U_n\|=\|\mathcal{D}(U_n)\|$.

4 Adamczak's moment inequality for U-statistics

The paper [1] by R. Adamczak developed moment inequalities for general Banach space-valued completely degenerate U-statistics of arbitrary order. More specifically, application of Theorem 1 in [1] to our scenario $\mathbb{B}=\left(\mathbb{H}^d,\ \|\cdot\|\right)$ and $m=2$ yields the following bounds for all $q\geq 1$ and $t\geq 2$:

$$\begin{aligned}\left(\mathbb{E}\|U_n\|^{2q}\right)^{1/(2q)} &\leq C\left(\mathbb{E}\|U_n\|+\sqrt{q}\cdot A+q\cdot B+q^{3/2}\cdot\Gamma+q^2\cdot D\right), \\ \Pr\left(\|U_n\|\geq C\left(\mathbb{E}\|U_n\|+\sqrt{t}\cdot A+t\cdot B+t^{3/2}\cdot\Gamma+t^2\cdot D\right)\right) &\leq e^{-t},\end{aligned} \tag{4.1}$$

where C is an absolute constant, and the quantities A,B,Γ,D will be specified below (see Section 5.3 for the complete statement of Adamczak's result). Notice that inequality (4.1) contains the "sub-Gaussian" term corresponding to $\sqrt{q}$ that did not appear in the previously established bounds.

We should mention another important distinction between (4.1) and the results of Theorem 3.7 and its corollaries, such as inequality (3.9): while (4.1) describes the deviations of $\|U_n\|$ from its expectation, (3.9) states that U_n is close to its expectation *as a random matrix*; similar connections exist between the Matrix Bernstein inequality [33] and Talagrand's concentration inequality [3]. It particular, (4.1) can be combined with a bound (3.10) for $\mathbb{E}\|U_n\|$ to obtain a moment inequality that is superior (in a certain range of q) to the results derived from Theorem 3.7.

Theorem 4.1. *Inequalities (4.1) hold with the following choice of A, B, Γ and D:*

$$A = \sqrt{\log(de)} \left(\mathbb{E} \left\| \mathbb{E}_2 \widetilde{G}\widetilde{G}^* \right\| + \left\| \sum_{(i_1,i_2)\in I_n^2} \mathbb{E} H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\| \right)^{1/2}$$
$$+ \log(de) \left(\mathbb{E} \max_{i_1} \left\| \sum_{i_2: i_2\neq i_1} H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\| \right)^{1/2},$$
$$B = \left(\sup_{z\in\mathbb{C}^d: \|z\|_2\le 1} \sum_{(i_1,i_2)\in I_n^2} \mathbb{E}\left(z^* H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) z \right)^2 \right)^{1/2}$$
$$\le \left(\left\| \sum_{(i_1,i_2)\in I_n^2} \mathbb{E} H^2_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \right\| \right)^{1/2},$$
$$\Gamma = \sqrt{1 + \frac{\log d}{q}} \left(\sum_{i_1} \mathbb{E}_1 \left\| \sum_{i_2: i_2\neq i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/2q},$$
$$D = \left(\sum_{(i_1,i_2)\in I_n^2} \mathbb{E} \left\| H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/(2q)}$$
$$+ \left(1 + \frac{\log d}{q}\right) \left(\sum_{i_1} \mathbb{E} \max_{i_2: i_2\neq i_1} \left\| H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/2q},$$

where $\widetilde{G}_i$ were defined in (3.6).

Proof. See Section 5.3. □

It is possible to further simplify the bounds for A (via Lemma 3.12) and D to deduce that one can choose

$$A = \log(de) \left(\mathbb{E} \left\| \mathbb{E}_2 \widetilde{G}\widetilde{G}^* \right\| + \left\| \sum_{(i_1,i_2)\in I_n^2} \mathbb{E} H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\| \right)^{1/2},$$
$$B = \left(\sup_{z\in\mathbb{C}^d: \|z\|_2\le 1} \sum_{(i_1,i_2)\in I_n^2} \mathbb{E}\left(z^* H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) z \right)^2 \right)^{1/2},$$
$$\Gamma = (\log(de))^{3/2} \left(\sum_{i_1} \mathbb{E}_1 \left\| \sum_{i_2: i_2\neq i_1} \mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/2q},$$
$$D = \log(de) \left(\sum_{(i_1,i_2)\in I_n^2} \mathbb{E} \left\| H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/(2q)}. \tag{4.2}$$

The upper bound for A can be modified even further as in (3.8), using the fact that

$$\mathbb{E}\left\|\mathbb{E}_2\widetilde{G}\widetilde{G}^*\right\| \le \sum_{i_1}\mathbb{E}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|.$$

5 Proofs

5.1 Tools from probability theory and linear algebra

This section summarizes several facts that will be used in our proofs. The first inequality is a bound connecting the norm of a matrix to the norms of its blocks.

Lemma 5.1. *Let $M \in \mathbb{H}^{d_1+d_2}$ be nonnegative definite and such that $M = \begin{pmatrix} A & X \\ X^* & B\end{pmatrix}$, where $A \in \mathbb{H}^{d_1}$ and $B \in \mathbb{H}^{d_2}$. Then*

$$|||M||| \le |||A||| + |||B|||$$

for any unitarily invariant norm $|||\cdot|||$.

Proof. It follows from the result in [4] that under the assumptions of the lemma, there exist unitary operators U, V such that

$$\begin{pmatrix} A & X \\ X^* & B\end{pmatrix} = U\begin{pmatrix} A & 0 \\ 0 & 0\end{pmatrix}U^* + V\begin{pmatrix} 0 & 0 \\ 0 & B\end{pmatrix}V^*,$$

hence the result is a consequence of the triangle inequality. □

The second result is the well-known decoupling inequality for U-statistics due to de la Pena and Montgomery-Smith [8].

Lemma 5.2. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent random variables with values in a measurable space $(S,\mathcal{S})$, and let $\{X_i^{(k)}\}_{i=1}^n$, $k=1,2,\ldots,m$ be m independent copies of this sequence. Let B be a separable Banach space and, for each $(i_1,\ldots,i_m)\in I_n^m$, let $H_{i_1,\ldots,i_m}: S^m \to B$ be a measurable function. Moreover, let $\Phi:[0,\infty)\to[0,\infty)$ be a convex nondecreasing function such that*

$$\mathbb{E}\Phi(\|H_{i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m})\|) < \infty$$

for all $(i_1,\ldots,i_m)\in I_n^m$. Then

$$\mathbb{E}\Phi\left(\left\|\sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m})\right\|\right) \le \\ \mathbb{E}\Phi\left(C_m\left\|\sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}\left(X^{(1)}_{i_1},\ldots,X^{(m)}_{i_m}\right)\right\|\right),$$

where $C_m := 2^m(m^m-1)\cdot((m-1)^{m-1}-1)\cdot\ldots\cdot 3$. Moreover, if $H_{i_1,\ldots,i_m}$ is P-canonical, then the constant C_m can be taken to be m^m. Finally, there exists a constant $D_m>0$ such that for all $t>0$,

$$\Pr\left(\left\|\sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}(X_{i_1},\ldots,X_{i_m})\right\| \ge t\right) \\ \le D_m\Pr\left(D_m\left\|\sum_{(i_1,\ldots,i_m)\in I_n^m} H_{i_1,\ldots,i_m}\left(X^{(1)}_{i_1},\ldots,X^{(m)}_{i_m}\right)\right\| \ge t\right).$$

Furthermore, if $H_{i_1,\dots,i_m}$ is permutation-symmetric, then, both of the above inequalities can be reversed (with different constants C_m and D_m).

The following results are the variants of the non-commutative Khintchine's inequalities (that first appeared in the works by Lust-Piquard and Pisier) for the Rademacher sums and the Rademacher chaos with explicit constants, see [24, 23], page 111 in [27], Theorems 6.14, 6.22 in [28] and Corollary 20 in [32].

Lemma 5.3. *Let $B_j \in \mathbb{C}^{r\times t}$, $j = 1,\dots,n$ be the matrices of the same dimension, and let $\{\varepsilon_j\}_{j\in\mathbb{N}}$ be a sequence of i.i.d. Rademacher random variables. Then for any $p \geq 1$,*

$$\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j B_j\right\|_{S_{2p}}^{2p} \leq \left(\frac{2\sqrt{2}}{e}p\right)^p \cdot \max\left\{\left\|\left(\sum_{j=1}^n B_j B_j^*\right)^{1/2}\right\|_{S_{2p}}^{2p}, \left\|\left(\sum_{j=1}^n B_j^* B_j\right)^{1/2}\right\|_{S_{2p}}^{2p}\right\}.$$

Lemma 5.4. *Let $\{A_{i_1,i_2}\}_{i_1,i_2=1}^n$ be a sequence of Hermitian matrices of the same dimension, and let $\left\{\varepsilon_i^{(k)}\right\}_{i=1}^n$, $k = 1, 2$, be i.i.d. Rademacher random variables. Then for any $p \geq 1$,*

$$\mathbb{E}\left\|\sum_{i_1=1}^n\sum_{i_2=1}^n A_{i_1,i_2}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}\right\|_{S_{2p}}^{2p} \leq$$

$$2\left(\frac{2\sqrt{2}}{e}p\right)^{2p} \max\left\{\left\|(GG^*)^{1/2}\right\|_{S_{2p}}^{2p}, \left\|\left(\sum_{i_1,i_2=1}^n A_{i_1,i_2}^2\right)^{1/2}\right\|_{S_{2p}}^{2p}\right\},$$

where the matrix $G \in \mathbb{H}^{nd}$ is defined as

$$G := \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}.$$

The following result (Theorem A.1 in [5]) is a variant of matrix Rosenthal's inequality for nonnegative definite matrices.

Lemma 5.5. *Let $Y_1,\dots,Y_n \in \mathbb{H}^d$ be a sequence of independent nonnegative definite random matrices. Then for all $q \geq 1$ and $r = \max(q, \log(d))$,*

$$\left(\mathbb{E}\left\|\sum_j Y_j\right\|^q\right)^{1/2q} \leq \left\|\sum_j \mathbb{E}Y_j\right\|^{1/2} + 2\sqrt{2er}\left(\mathbb{E}\max_j \|Y_j\|^q\right)^{1/2q}.$$

The next inequality (see equation (2.6) in [12]) allows to replace the sum of moments of nonnegative random variables with maxima.

Lemma 5.6. *Let $\xi_1,\dots,\xi_n$ be independent random variables. Then for all $q > 1$ and $\alpha \geq 0$,*

$$q^{\alpha q}\sum_{i=1}^{n}|\xi_i|^q \le 2(1+q^\alpha)\max\left(q^{\alpha q}\mathbb{E}\max_i|\xi_i|^q, \left(\sum_{i=1}^{n}\mathbb{E}|\xi_i|\right)^q\right).$$

Finally, the following inequalities allow transitioning between moment and tail bounds.

Lemma 5.7. *Let X be a random variable satisfying $(\mathbb{E}|X|^p)^{1/p} \le a_4p^2 + a_3p^{3/2} + a_2p + a_1\sqrt{p} + a_0$ for all $p \ge 2$ and some positive real numbers $a_j,\ j = 0,\dots,3$. Then for any $u \ge 2$,*

$$\Pr\left(|X| \ge e(a_4u^2 + a_3u^{3/2} + a_2u + a_1\sqrt{u} + a_0)\right) \le \exp(-u).$$

See Proposition 7.11 and 7.15 in [10] for the proofs of closely related bounds.

Lemma 5.8. *Let X be a random variable such that $\Pr(|X| \ge a_0 + a_1\sqrt{u} + a_2u) \le e^{-u}$ for all $u \ge 1$ and some $0 \le a_0, a_1, a_2 < \infty$. Then*

$$(\mathbb{E}|X|^p)^{1/p} \le C(a_0 + a_1\sqrt{p} + a_2p)$$

for an absolute constant $C > 0$ and all $p \ge 1$.

The proof follows from the formula $\mathbb{E}|X|^p = p\int_0^\infty \Pr(|X| \ge t)\, t^{p-1}dt$, see Lemma A.2 in [9] and Proposition 7.14 in [10] for the derivation of similar inequalities. Next, we will use Lemma 5.2 combined with a well-known argument to obtain the symmetrization inequality for degenerate U-statistics.

Lemma 5.9. *Let $H_{i_1,i_2} : S \times S \mapsto \mathbb{H}^d$ be degenerate kernels, $X_1,\dots,X_n$ – i.i.d. S-valued random variables, and assume that $\{X_i^{(k)}\}_{i=1}^n$, $k = 1,2$, are independent copies of this sequence. Moreover, let $\left\{\varepsilon_i^{(k)}\right\}_{i=1}^n$, $k = 1,2$, be i.i.d. Rademacher random variables. Define*

$$U_n' := \sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right). \tag{5.1}$$

Then for any $p \ge 1$,

$$\left(\mathbb{E}\,\|U_n\|^p\right)^{1/p} \le 16\left(\mathbb{E}\,\|U_n'\|^p\right)^{1/p}.$$

Proof. Note that

$$\begin{aligned}\mathbb{E}\|U_n\|^p =& \mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}(X_{i_1}, X_{i_2})\right\|^p \\ \le& \mathbb{E}\left\|2^2\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p,\end{aligned}$$

where the inequality follows from the fact that H_{i_1,i_2} is $\mathcal{P}$-canonical, hence Lemma 5.2 applies with constant equal to $C_2 = 4$.

Next, for $i = 1,2$, let $\mathbb{E}_i[\cdot]$ stand for the expectation with respect to $\left\{X_j^{(i)}, \varepsilon_j^{(i)}\right\}_{j\ge1}$ only (that is, conditionally on $\left\{X_j^{(k)}, \varepsilon_j^{(k)}\right\}_{j\ge1}$, $k \ne i$). Using iterative expectations and the symmetrization inequality for the Rademacher sums twice (see Lemma 6.3 in [22]), we deduce that

$$
\begin{aligned}
\mathbb{E}\|U_n\|^p \leq & 4^p\, \mathbb{E}\left\| \sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p \\
= & 4^p\, \mathbb{E}\left[\mathbb{E}_1 \left\| \sum_{i_1=1}^{n} \sum_{i_2\neq i_1} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p\right] \\
\leq & 4^p\, \mathbb{E}\left[\mathbb{E}_1 \left\| 2\sum_{i_1=1}^{n} \varepsilon_{i_1}^{(1)} \sum_{i_2\neq i_1} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p\right] \\
= & 4^p\, \mathbb{E}\left[\mathbb{E}_2 \left\| 2\sum_{i_2=1}^{n} \sum_{i_1\neq i_2} \varepsilon_{i_1}^{(1)} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p\right] \\
\leq & 4^p\, \mathbb{E}\left[\left\| 4\sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^p\right]. \qquad \square
\end{aligned}
$$

5.2 Proofs of results in Section 3

5.2.1 Proof of Lemma 3.3

Recall that

$$
X = \sum_{i_1=1}^{n} \sum_{i_2\neq i_1} A_{i_1,i_2}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)},
$$

where $A_{i_1,i_2} \in \mathbb{H}^d$ for all i_1, i_2, and let $C_p := 2\left(\frac{2\sqrt{2}}{e}p\right)^{2p}$. We will first establish the upper bound. Application of Lemma 5.4 (Khintchine's inequality) to the sequence of matrices $\{A_{i_1,i_2}\}_{i_1,i_2=1}^n$ such that $A_{j,j} = 0$ for $j = 1, \ldots, n$ yields

$$
\left(\mathbb{E}\|X\|_{S_{2p}}^{2p}\right)^{1/2p} \leq 2^{1/2p}\frac{2\sqrt{2}}{e}\cdot p \cdot \max\left\{\left\|(GG^*)^{1/2}\right\|_{S_{2p}}, \left\|\left(\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right)^{1/2}\right\|_{S_{2p}}\right\}, \tag{5.2}
$$

where

$$
G := \begin{pmatrix} 0 & A_{12} & \ldots & A_{1n} \\ A_{21} & 0 & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{nd\times nd}.
$$

Our goal is to obtain a version of inequality (5.2) for $p = \infty$. To this end, we need to find an upper bound for

$$
\inf_{p\geq q}\left[\frac{p\cdot \max\left\{\left\|(GG^*)^{1/2}\right\|_{S_{2p}}, \left\|\left(\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right)^{1/2}\right\|_{S_{2p}}\right\}}{\max\left\{\left\|(GG^*)^{1/2}\right\|, \left\|\left(\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right)^{1/2}\right\|\right\}}\right].
$$

Since G is a $nd \times nd$ matrix, a naive upper bound is of order $\log(nd)$. We will show that it can be improved to $\log d$. To this end, we need to distinguish between the cases when the maximum in (5.2) is attained by the first or second term. Define

$$
\widehat{B}_{i_1,i_2} = [0 \mid 0 \mid \ldots \mid A_{i_1,i_2} \mid \ldots \mid 0 \mid 0] \in \mathbb{C}^{d\times nd},
$$

where $A_{i_1 i_2}$ sits on the i_1-th position of the above block matrix. Moreover, let

$$B_{i_2} = \sum_{i_1 : i_1 \neq i_2} \widehat{B}_{i_1,i_2}. \tag{5.3}$$

Then it is easy to see that

$$GG^* = \sum_{i_2} B^*_{i_2} B_{i_2},$$
$$\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2} = \sum_{i_2} B_{i_2} B^*_{i_2}.$$

The following bound gives a key estimate.

Lemma 5.10. *Let $M_1, \dots, M_N$ be a sequence of $\mathbb{C}^{d\times nd}$-valued matrices. Let $\lambda_1, \dots, \lambda_{nd}$ be eigenvalues of $\sum_j M_j^* M_j$ and let $\nu_1, \dots, \nu_d$ be eigenvalues of $\sum_j M_j M_j^*$. Then $\sum_{i=1}^{nd} \lambda_i = \sum_{j=1}^{d} \nu_j$. Furthermore, if $\max_i \lambda_i \le \frac{1}{d}\sum_{j=1}^{d} \nu_j$, then*

$$\left\| \left(\sum_j M_j M_j^* \right)^{1/2} \right\|^{2p}_{S_{2p}} \ge \left\| \left(\sum_j M_j^* M_j \right)^{1/2} \right\|^{2p}_{S_{2p}},$$

for any integer $p \ge 2$.

The proof of the Lemma is given in Section 5.2.2. We will apply this fact with $M_j = B_j$, $j = 1, \dots, n$. Assuming that $\max_i \lambda_i \le \frac{1}{d}\sum_{j=1}^{nd} \lambda_j$, it is easy to see that the second term in the maximum in (5.2) dominates, hence

$$\begin{aligned} \mathbb{E}\|X\|^{2p}_{S_{2p}} \le C_p \left\| \left(\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2} \right)^{1/2} \right\|^{2p}_{S_{2p}} &= C_p \operatorname{tr} \left(\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1 i_2} \right)^p \\ &\le C_p \cdot d \cdot \left\| \left(\sum_{(i_1,i_2)\in I_n^2} A^2_{i_1 i_2} \right)^p \right\| = C_p \cdot d \cdot \left\| \sum_{(i_1,i_2)\in I_n^2} A^2_{i_1,i_2} \right\|^p, \end{aligned} \tag{5.4}$$

where the last equality follows from the fact that for any positive semidefinite matrix H, $\|H^p\| = \|H\|^p$. On the other hand, when $\max_i \lambda_i > \frac{1}{d}\sum_{j=1}^{nd} \lambda_j$, it is easy to see that for all $p \ge 1$,

$$d > \sum_{j=1}^{nd} \frac{\lambda_j}{\max_i \lambda_i} \ge \sum_{j=1}^{nd} \left(\frac{\lambda_j}{\max_i \lambda_i} \right)^p,$$

which in turn implies that

$$d \left(\max_i \lambda_i \right)^p \ge \sum_{j=1}^{nd} \lambda_j^p. \tag{5.5}$$

Moreover,

$$\left\| \left(\sum_{i_2} B^*_{i_2} B_{i_2} \right)^{1/2} \right\|^{2p}_{S_{2p}} = \operatorname{tr} \left(\left(\sum_{i_2} B^*_{i_2} B_{i_2} \right)^p \right) = \sum_{i=1}^{nd} \lambda_i^p. \tag{5.6}$$

Combining (5.5), (5.6), we deduce that

$$\left\| \left(\sum_{i_2} B_{i_2}^* B_{i_2} \right)^{1/2} \right\|_{S_{2p}}^{2p} \leq d \left\| \left(\sum_{i_2} B_{i_2}^* B_{i_2} \right)^p \right\| = d \left\| (GG^*)^p \right\| = d \left\| GG^* \right\|^p ,$$

where the second from the last equality follows again from the fact that for any positive semi-definite matrix H, $\|H^p\| = \|H\|^p$. Thus, combining the bound above with (5.2) and (5.4), we obtain

$$\mathbb{E}\|X\|_{S_{2p}}^{2p} \leq d \cdot C_p \max \left\{ \|GG^*\|^p , \left\| \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 \right\|^p \right\}.$$

Finally, set $p = \max(q, \log(d))$ and note that $d^{1/2p} \leq \sqrt{e}$, hence

$$\left(\mathbb{E}\|X\|^{2q}\right)^{1/2q} \leq \left(\mathbb{E}\|X\|^{2p}\right)^{1/2p} \leq \frac{4}{\sqrt{e}} \max\{\log d, q\} \cdot \max \left\{ \|GG^*\| , \left\| \sum_{(i_1,i_2)\in I_2^n} A_{i_1,i_2}^2 \right\| \right\}^{1/2} .$$

This finishes the proof of upper bound.

Now, we turn to the lower bound. Let $\mathbb{E}_1[\cdot]$ stand for the expectation with respect to $\left\{\varepsilon_j^{(1)}\right\}_{j\geq 1}$ only. Then

$$\begin{aligned}
\left(\mathbb{E}\|X\|^{2p}\right)^{1/(2p)} \geq \left(\mathbb{E}\|X\|^{2}\right)^{1/2} &= \left(\mathbb{E}\mathbb{E}_1 \left\| \left(\sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} A_{i_1,i_2} \right)^2 \right\| \right)^{1/2} \\
&\geq \left(\mathbb{E} \left\| \mathbb{E}_1 \left(\sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} A_{i_1,i_2} \right)^2 \right\| \right)^{1/2} \\
&= \left(\mathbb{E} \left\| \sum_{i_1} \left(\sum_{i_2 : i_2 \neq i_1} \varepsilon_{i_2}^{(2)} A_{i_1,i_2} \right)^2 \right\| \right)^{1/2} .
\end{aligned}$$

It is easy to check that

$$\sum_{i_1=1}^{n} \left(\sum_{i_2 : i_2 \neq i_1} \varepsilon_{i_2}^{(2)} A_{i_1,i_2} \right)^2 = \left(\sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right) \left(\sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right)^* ,$$

where B_i were defined in (5.26). Hence

$$\left(\mathbb{E}\|X\|^{2p}\right)^{1/(2p)} \geq \left(\mathbb{E} \left\| \left(\sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right) \left(\sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right)^* \right\| \right)^{1/2} .$$

Next, for any matrix $A \in \mathbb{C}^{d_1 \times d_2}$,

$$\left\| \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}^2 \right\| = \left\| \begin{pmatrix} A^*A & 0 \\ 0 & AA^* \end{pmatrix} \right\| = \max\{\|A^*A\|, \|AA^*\|\} = \|AA^*\|,$$

where the last equality follows from the fact that $\|AA^*\| = \|A^*A\|$. Taking $A = \sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2}$ yields that

$$\begin{aligned}
\left(\mathbb{E}\|X\|^{2p}\right)^{1/(2p)} \geq (\mathbb{E}\,\|BB^*\|)^{1/2} &= \left(\mathbb{E}\left\|\left(\sum_{i_2} \varepsilon_{i_2}^{(2)} \begin{pmatrix} 0 & B_{i_2}^* \\ B_{i_2} & 0 \end{pmatrix}\right)^2\right\|\right)^{1/2} \\
&\geq \left\|\mathbb{E}\left(\sum_{i_2} \varepsilon_{i_2}^{(2)} \begin{pmatrix} 0 & B_{i_2}^* \\ B_{i_2} & 0 \end{pmatrix}\right)^2\right\|^{1/2} = \left\|\sum_{i_2} \begin{pmatrix} B_{i_2}^* B_{i_2} & 0 \\ 0 & B_{i_2} B_{i_2}^* \end{pmatrix}\right\|^{1/2} \\
&= \max\left\{\left\|\sum_{i_2} B_{i_2}^* B_{i_2}\right\|, \left\|\sum_{i_2} B_{i_2} B_{i_2}^*\right\|\right\}^{1/2} \\
&= \max\left\{\|GG^*\|, \left\|\sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2\right\|\right\}^{1/2}.
\end{aligned}$$

5.2.2 Proof of Lemma 5.10

The equality of traces is obvious since

$$\operatorname{tr}\left(\sum_{j=1}^N M_j M_j^*\right) = \sum_{j=1}^N \operatorname{tr}\left(M_j M_j^*\right) = \sum_{j=1}^N \operatorname{tr}\left(M_j^* M_j\right) = \operatorname{tr}\left(\sum_{j=1}^N M_j^* M_j\right).$$

Set

$$S := \sum_{i=1}^{nd} \lambda_i = \sum_{i=1}^{d} \nu_i.$$

Note that

$$\begin{aligned}
\left\|\left(\sum_{j=1}^N M_j^* M_j\right)^{1/2}\right\|_{S_{2p}}^{2p} &= \operatorname{tr}\left(\left(\sum_{j=1}^N M_j^* M_j\right)^p\right) = \sum_{i=1}^{nd} \lambda_i^p, \\
\left\|\left(\sum_{j=1}^N M_j M_j^*\right)^{1/2}\right\|_{S_{2p}}^{2p} &= \operatorname{tr}\left(\left(\sum_{j=1}^N M_j M_j^*\right)^p\right) = \sum_{i=1}^{d} \nu_i^p.
\end{aligned}$$

Moreover, $\lambda_i \geq 0$, $\nu_j \geq 0$ for all i, j, and $\max_i \lambda_i \leq \frac{1}{d}\sum_{j=1}^d \nu_j = \frac{S}{d}$ by assumption. It is clear that

$$\begin{aligned}
\left\|\left(\sum_{j=1}^N M_j^* M_j\right)^{1/2}\right\|_{S_{2p}}^{2p} &\leq \max_{0\leq\lambda_i\leq\frac{S}{2d},\ \sum_{i=1}^{nd}\lambda_i=S} \sum_{i=1}^{nd} \lambda_i^p, \\
\left\|\left(\sum_{j=1}^N M_j M_j^*\right)^{1/2}\right\|_{S_{2p}}^{2p} &\geq \min_{\nu_i\geq 0\ ,\sum_{i=1}^{d}\nu_i=S} \sum_{i=1}^{d} \nu_i^p.
\end{aligned}$$

Hence, it is enough to show that

$$\max_{0\leq\lambda_i\leq\frac{S}{d},\ \sum_{i=1}^{nd}\lambda_i=S} \sum_{i=1}^{nd} \lambda_i^p \leq \min_{\nu_i\geq 0,\ \sum_{i=1}^{d}\nu_i=S} \sum_{i=1}^{d} \nu_i^p. \tag{5.7}$$

The right hand side of the inequality (5.7) can be estimated via Jensen's inequality as

$$\begin{aligned}\min_{\nu_i\ge 0,\ \sum_{i=1}^d \nu_i = S} \sum_{i=1}^d \nu_i^p &= d\cdot \min_{\nu_i\ge 0,\ \sum_{i=1}^d \nu_i = S} \frac{1}{d}\sum_{i=1}^d \nu_i^p \\ &\ge d\cdot \min_{\nu_i\ge 0,\ \sum_{i=1}^d \nu_i = S} \left(\frac{1}{d}\sum_{i=1}^d \nu_i\right)^p = d\cdot\left(\frac{S}{d}\right)^p. \end{aligned} \tag{5.8}$$

It remains to show that $\sum_{i=1}^{nd}\lambda_i^p \le d\cdot\left(\frac{S}{d}\right)^p$. For a sequence $\{a_j\}_{j=1}^N \subset \mathbb{R}$, let $a_{(j)}$ be the j-th smallest element of the sequence, where the ties are broken arbitrary. A sequence $\{a_j\}_{j=1}^N$ majorizes a sequence $\{b_j\}_{j=1}^N$ whenever $\sum_{j=0}^k a_{(N-j)} \ge \sum_{j=0}^k b_{(N-j)}$ for all $0\le k\le N-2$, and $\sum_j a_j = \sum_j b_j$. A function $g:\mathbb{R}^N\mapsto\mathbb{R}$ is called Schur-convex if $g(a_1,\dots,a_N)\ge g(b_1,\dots,b_N)$ whenever $\{a_j\}_{j=1}^N$ majorizes $\{b_j\}_{j=1}^N$. It is well known that if $f:\mathbb{R}\mapsto\mathbb{R}$ is convex, then $g(a_1,\dots,a_N)=\sum_{j=1}^N f(a_j)$ is Schur convex. In particular, $g(a_1,\dots,a_N)=\sum_{j=1}^N a_j^p$, where $a_1,\dots,a_N\ge 0$, is Schur convex for $p\ge 1$. Consider the sequence $a_1=\ldots=a_d=\frac{S}{d}$, $a_{d+1}=\ldots=a_{nd}=0$ and $b_1=\lambda_1,\dots,b_{nd}=\lambda_{nd}$. Since $\max_i \lambda_i \le \frac{S}{d}$ by assumption, the sequence $\{a_j\}$ majorizes $\{b_j\}$, hence Schur convexity yields that $\sum_{i=1}^{nd}\lambda_i^p \le \sum_{i=1}^d \left(\frac{S}{d}\right)^p = d\cdot\left(\frac{S}{d}\right)^p$, implying the result.[1]

5.2.3 Proof of Theorem 3.7

The first inequality in the statement of the theorem follows immediately from Lemma 5.2. Next, it is easy to deduce from the proof of Lemma 5.9 that

$$\left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^{2q}\right)^{1/2q} \le 4\left(\mathbb{E}\left\|U_n'\right\|^{2q}\right)^{1/2q}, \tag{5.9}$$

where U_n' was defined in (5.1). Applying Lemma 3.3 conditionally on $\{X_i^{(j)}\}_{i=1}^n$, $j=1,2$, we get

$$\begin{aligned}\left(\mathbb{E}\left\|U_n'\right\|^{2q}\right)^{1/2q} &= \left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} H_{i_1,i_2}(X_{i_1}^{(1)},X_{i_2}^{(2)})\right\|^{2q}\right)^{1/(2q)} \\ &\le 4e^{-1/2}\max(q,\log d)\left(\mathbb{E}\max\left\{\|\widetilde{G}\widetilde{G}^*\|, \left\|\sum_{(i_1,i_2)\in I_n^2} H^2_{i_1,i_2}\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|\right\}^q\right)^{1/2q},\end{aligned} \tag{5.10}$$

where $\widetilde{G}$ was defined in (3.6). Let $\widetilde{G}_i$ be the i-th column of $\widetilde{G}$, then

$$\widetilde{G}\widetilde{G}^* = \sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*, \qquad \sum_{(i_1,i_2)\in I_n^2} H^2_{i_1,i_2}(X_{i_1}^{(1)},X_{i_2}^{(2)}) = \sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i.$$

Let $Q_i\in\mathbb{H}^{(n+1)d\times(n+1)d}$ be defined as

$$Q_i = \begin{pmatrix} 0 & \widetilde{G}_i^* \\ \widetilde{G}_i & 0\end{pmatrix},$$

[1] We are thankful to the anonymous Referee for suggesting an argument based on Schur convexity, instead of the original proof that was longer and not as elegant.

so that

$$Q_i^2 = \begin{pmatrix} \widetilde{G}_i^* \widetilde{G}_i & 0 \\ 0 & \widetilde{G}_i \widetilde{G}_i^* \end{pmatrix}.$$

Inequality (5.10) implies that

$$\begin{aligned} &\left(\mathbb{E} \left\| \sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1,i_2}\left(X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{2q} \right)^{1/(2q)} \\ &\qquad \le 4e^{-1/2} \max(q, \log d) \left(\mathbb{E} \left\| \sum_{i=1}^n Q_i^2 \right\|^q \right)^{1/(2q)}. \end{aligned} \tag{5.11}$$

Let $\mathbb{E}_2[\cdot]$ stand for the expectation with respect to $\left\{ X_i^{(2)} \right\}_{i=1}^n$ only (that is, conditionally on $\left\{ X_i^{(1)} \right\}_{i=1}^n$). Then Minkowski inequality followed by the symmetrization inequality imply that

$$\begin{aligned} \left(\mathbb{E} \left\| \sum_{i=1}^n Q_i^2 \right\|^q \right)^{1/(2q)} &\le \left(\mathbb{E} \left\| \sum_{i=1}^n \left(Q_i^2 - \mathbb{E}_2 Q_i^2 \right) \right\|^q \right)^{1/(2q)} + \left(\mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 Q_i^2 \right\|^q \right)^{1/2q} \\ &= \left(\mathbb{E}\, \mathbb{E}_2 \left\| \sum_{i=1}^n Q_i^2 - \mathbb{E}_2 Q_i^2 \right\|^q \right)^{1/(2q)} + \left(\mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 Q_i^2 \right\|^q \right)^{1/2q} \\ &\le \sqrt{2} \left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i Q_i^2 \right\|^q \right)^{1/(2q)} + \left(\mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 Q_i^2 \right\|^q \right)^{1/2q}. \end{aligned} \tag{5.12}$$

Next, we obtain an upper bound for $\left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i Q_i^2 \right\|^q \right)^{1/(2q)}$. To this end, we apply Khintchine's inequality (Lemma 5.3). Denote $C_r := \left(\frac{2\sqrt{2}}{e} r \right)^{2r}$, and let $\mathbb{E}_\varepsilon[\cdot]$ be the expectation with respect to $\{\varepsilon_i\}_{i=1}^n$ only. Then for $r > q$ we deduce that

$$\begin{aligned} \mathbb{E}_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i Q_i^2 \right\|_{S_{2r}}^{2r} &\le C_r^{1/2} \left\| \left(\sum_{i=1}^n Q_i^4 \right)^{1/2} \right\|_{S_{2r}}^{2r} \\ &= C_r^{1/2} \mathrm{tr} \left(\sum_{i=1}^n Q_i^4 \right)^r \le C_r^{1/2} \mathrm{tr} \left(\sum_{i=1}^n Q_i^2 \cdot \|Q_i^2\| \right)^r \\ &\le C_r^{1/2} \max_i \|Q_i^2\|^r \cdot \left\| \left(\sum_{i=1}^n Q_i^2 \right)^{1/2} \right\|_{S_{2r}}^{2r}, \end{aligned}$$

where we used the fact that $Q_i^4 \preceq \|Q_i^2\| Q_i^2$ for all i, and the fact that $A \preceq B$ implies that $\mathrm{tr}\, g(A) \le \mathrm{tr}\, g(B)$ for any non-decreasing $g : \mathbb{R} \mapsto \mathbb{R}$. Next, we will focus on the term

$$\left\| \left(\sum_{i=1}^n Q_i^2 \right)^{1/2} \right\|_{S_{2r}}^{2r} = \mathrm{tr} \left(\left(\sum_{i=1}^n \widetilde{G}_i \widetilde{G}_i^* \right)^r \right) + \mathrm{tr} \left(\left(\sum_{i=1}^n \widetilde{G}_i^* \widetilde{G}_i \right)^r \right).$$

Applying Lemma 5.10 with $M_j = \widetilde{G}_j^*,\ j = 1, \ldots, n$, we deduce that

- if $\left\|\sum_{i=1}^n \widetilde{G}_i\,\widetilde{G}_i^*\right\| \leq \frac{1}{d}\,\mathrm{tr}\left(\sum_{i=1}^n \widetilde{G}_i^*\,\widetilde{G}_i\right)$, then $\left\|\left(\sum_{i=1}^n \widetilde{G}_i\,\widetilde{G}_i^*\right)^{1/2}\right\|_{S_{2r}}^{2r} \leq \left\|\left(\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right)^{1/2}\right\|_{S_{2r}}^{2r}$, which implies that $\mathrm{tr}\left(\left(\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right)^r\right) \leq \mathrm{tr}\left(\left(\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right)^r\right)$, and
$$\left\|\left(\sum_{i=1}^n Q_i^2\right)^{1/2}\right\|_{S_{2r}}^{2r} \leq 2d\cdot\left\|\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right\|^r.$$
- if $\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\| > \frac{1}{d}\mathrm{tr}\left(\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right)$, let λ_j be the j-th eigenvalue of $\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*$, and note that
$$d > \frac{\mathrm{tr}\left(\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right)}{\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\|} = \frac{\mathrm{tr}\left(\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right)}{\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\|} = \sum_{i=1}^{nd}\frac{\lambda_i}{\max_j \lambda_j} \geq \sum_{i=1}^{nd}\left(\frac{\lambda_i}{\max_j \lambda_j}\right)^r,$$
where $r \geq 1$. In turn, it implies that
$$\mathrm{tr}\left(\left(\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right)^r\right) < d\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\|^r.$$
Thus
$$\begin{aligned}\left\|\left(\sum_{i=1}^n Q_i^2\right)^{1/2}\right\|_{S_{2r}}^{2r} &= \mathrm{tr}\left(\left(\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right)^r\right) + \mathrm{tr}\left(\left(\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right)^r\right)\\ &\leq d\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\|^r + d\left\|\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right\|^r.\end{aligned}$$

Putting the bounds together, we obtain that
$$\begin{aligned}\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n \varepsilon_i Q_i^2\right\|_{S_{2r}}^{2r} &\leq 2dC_r^{1/2}\max_i \left\|Q_i^2\right\|^r \max\left\{\left\|\sum_{i=1}^n \widetilde{G}_i\widetilde{G}_i^*\right\|^r, \left\|\sum_{i=1}^n \widetilde{G}_i^*\widetilde{G}_i\right\|^r\right\} \qquad (5.13)\\ &\leq 2dC_r^{1/2}\max_i \left\|Q_i^2\right\|^r\cdot\left\|\sum_{i=1}^n Q_i^2\right\|^r.\end{aligned}$$

Next, observe that for r such that $2r \geq q$, $\mathbb{E}_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^q \leq \left(\mathbb{E}_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^{2r}\right)^{q/2r}$ by Hölder's inequality, hence
$$\begin{aligned}\left(\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^q\right)^{1/2q} &= \left(\mathbb{E}\mathbb{E}_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^q\right)^{1/2q}\\ &\leq \left(\mathbb{E}\left(\mathbb{E}_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^{2r}\right)^{q/2r}\right)^{1/2q}\\ &\leq \left(\mathbb{E}\left(\mathbb{E}_\varepsilon\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|_{S_{2r}}^{2r}\right)^{q/2r}\right)^{1/2q}\\ &\leq \left(2dC_r^{1/2}\right)^{1/4r}\left(\mathbb{E}\left[\max_i \left\|Q_i^2\right\|^{q/2}\cdot\left\|\sum_{i=1}^n Q_i^2\right\|^{q/2}\right]\right)^{1/2q}.\end{aligned}$$

Set $r = q \vee \log d$ and apply Cauchy-Schwarz inequality to deduce that

$$\left(\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j Q_j^2\right\|^q\right)^{1/2q} \leq (8r)^{1/4}\left(\mathbb{E}\max_i \left\|Q_i^2\right\|^q\right)^{1/(4q)}\left(\mathbb{E}\left\|\sum_{i=1}^n Q_i^2\right\|^q\right)^{1/(4q)}. \tag{5.14}$$

Substituting bound (5.14) into (5.12) and letting

$$R_q := \left(\mathbb{E}\left\|\sum_{i=1}^n Q_i^2\right\|^q\right)^{1/(2q)},$$

we obtain

$$R_q \leq (8r)^{1/4}\sqrt{2R_q}\left(\mathbb{E}\max_i \left\|Q_i^2\right\|^q\right)^{1/(4q)} + \left(\mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 Q_i^2\right\|^q\right)^{1/2q}.$$

If $x, a, b > 0$ are such that $x \leq a\sqrt{x} + b$, then $x \leq 4a^2 \vee 2b$, hence

$$R_q \leq 16\sqrt{2r}\left(\mathbb{E}\max_i \left\|Q_i^2\right\|^q\right)^{1/(2q)} + 2\left(\mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 Q_i^2\right\|^q\right)^{1/2q}.$$

Finally, it follows from (5.11) that

$$\begin{aligned}
&\left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} H(X_{i_1}^{(1)}, X_{i_2}^{(2)})\right\|^{2q}\right)^{1/(2q)} \\
&\leq 64\sqrt{\frac{2}{e}}\, r^{3/2}\left(\mathbb{E}\max_i \left\|Q_i^2\right\|^q\right)^{1/(2q)} + \frac{8}{\sqrt{e}} r\left(\mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 Q_i^2\right\|^q\right)^{1/2q} \\
&\leq 64\sqrt{\frac{2}{e}}\, r^{3/2}\left(\mathbb{E}\max_i \left\|\widetilde{G}_i^*\widetilde{G}_i\right\|^q\right)^{1/(2q)} + \frac{8}{\sqrt{e}} r\left(\mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 \widetilde{G}_i\widetilde{G}_i^*\right\|^q + \mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 \widetilde{G}_i^*\widetilde{G}_i\right\|^q\right)^{1/2q} \\
&= 64\sqrt{\frac{2}{e}} r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2: i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)} \\
&\quad + \frac{8}{\sqrt{e}} r\left(\mathbb{E}\left\|\sum_{i=1}^n \mathbb{E}_2 \widetilde{G}_i\widetilde{G}_i^*\right\|^q + \mathbb{E}\left\|\sum_{i_1=1}^n\left(\sum_{i_2: i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right)\right\|^q\right)^{1/2q},
\end{aligned} \tag{5.15}$$

where the last equality follows from the definition of $\widetilde{G}_i$. To bring the bound to its final form, we will apply Rosenthal's inequality (Lemma 5.5) to the last term in (5.15) to get that

$$\begin{aligned}
\left(\mathbb{E}\left\|\sum_{i_1=1}^n\left(\sum_{i_2: i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right)\right\|^q\right)^{1/2q} &\leq \left\|\sum_{(i_1,i_2)\in I_n^2}\mathbb{E}H_{i_1,i_2}^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^{1/2} \\
&\quad + 2\sqrt{2er}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2: i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)}, X_{i_2}^{(2)}\right)\right\|^q\right)^{1/2q}.
\end{aligned}$$

Moreover, Jensen's inequality implies that

$$\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q \le \mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q,$$

hence this term can be combined with one of the terms in (5.15).

5.2.4 Proof of Lemma 3.8

Let $\mathbb{E}_i[\cdot]$ stand for the expectation with respect to the variables with the upper index i only. Since $H_{i_1,i_2}(\cdot,\cdot)$ are permutation-symmetric, we can apply the second part of Lemma 5.2 and (twice) the desymmetrization inequality (see Theorem 3.1.21 in [13]) to get that for some absolute constant $C_0>0$

$$\begin{aligned}\left(\mathbb{E}\left\|U_n\right\|^{2q}\right)^{1/(2q)} \ge& \frac{1}{C_0}\left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{2q}\right)^{1/(2q)}\\ =&\frac{1}{C_0}\left(\mathbb{E}_2\,\mathbb{E}_1\left\|\sum_{i_1}\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{2q}\right)^{1/(2q)}\\ \ge&\frac{1}{2C_0}\left(\mathbb{E}\left\|\sum_{i_1}\varepsilon^{(1)}_{i_1}\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{2q}\right)^{1/(2q)}\\ =&\frac{1}{2C_0}\left(\mathbb{E}\mathbb{E}_2\left\|\sum_{i_2}\sum_{i_1:i_1\neq i_2}\varepsilon^{(1)}_{i_1} H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{2q}\right)^{1/(2q)}\\ \ge&\frac{1}{4C_0}\left(\mathbb{E}\left\|\sum_{(i_1,i_2)\in I_n^2}\varepsilon^{(1)}_{i_1}\varepsilon^{(2)}_{i_2} H_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^{2q}\right)^{1/(2q)}.\end{aligned}$$

Applying the lower bound of Lemma 3.3 conditionally on $\left\{X^{(1)}_i\right\}_{i=1}^n$ and $\left\{X^{(2)}_i\right\}_{i=1}^n$, we obtain

$$\begin{aligned}\left(\mathbb{E}\left\|U_n\right\|^{2q}\right)^{1/(2q)} \ge& c\left(\mathbb{E}\max\left\{\left\|\sum_i \widetilde{G}_i^*\widetilde{G}_i\right\|,\left\|\sum_i \widetilde{G}_i\widetilde{G}_i^*\right\|\right\}^q\right)^{1/2q} \qquad (5.16)\\ \ge&\frac{1}{4\sqrt{2}C_0}\left(\left(\mathbb{E}\left\|\sum_i\widetilde{G}_i\widetilde{G}_i^*\right\|^q\right)^{1/2q}+\left(\mathbb{E}\left\|\sum_i\widetilde{G}_i^*\widetilde{G}_i\right\|^q\right)^{1/2q}\right)\\ \ge&\frac{1}{4\sqrt{2}C_0}\left(\left(\mathbb{E}\left\|\sum_i\mathbb{E}_2\widetilde{G}_i\widetilde{G}_i^*\right\|^q\right)^{1/2q}+\left(\mathbb{E}\left\|\sum_i\mathbb{E}_2\widetilde{G}_i^*\widetilde{G}_i\right\|^q\right)^{1/2q}\right),\end{aligned}$$

where $\widetilde{G}_i$ is the i-th column if the matrix $\widetilde{G}$ defined in (3.6); we also used the identities $\widetilde{G}\widetilde{G}^*=\sum_{i=1}^n\widetilde{G}_i\widetilde{G}_i^*$, $\sum_{(i_1,i_2)\in I_n^2}H^2_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})=\sum_{i=1}^n\widetilde{G}_i^*\widetilde{G}_i$. The inequality above takes care of the second and third terms in the lower bound of the lemma. To show that the first term is necessary, let

$$Q_i = \begin{pmatrix} 0 & \widetilde{G}_i^* \\ \widetilde{G}_i & 0 \end{pmatrix}.$$

It follows from the first line of (5.16) that

$$\left(\mathbb{E}\left\|U_n\right\|^{2q}\right)^{1/(2q)} \geq \frac{1}{4C_0}\left(\mathbb{E}\left\|\sum_{i=1}^n Q_i^2\right\|^q\right)^{1/(2q)}.$$

Let i_* be the smallest value of $i \leq n$ where $\max_i \left\|Q_i^2\right\|$ is achieved. Then $\sum_{i=1}^n Q_i^2 \succeq Q_{i_*}^2$, hence $\left\|Q_{i_*}^2\right\| \leq \left\|\sum_{i=1}^n Q_i^2\right\|$. Jensen's inequality implies that

$$\begin{aligned}\left(\mathbb{E}\left\|U_n\right\|^{2q}\right)^{1/(2q)} &\geq \frac{1}{4C_0}\left(\mathbb{E}\left\|\sum_{i=1}^n Q_i^2\right\|^q\right)^{1/(2q)} \\ &\geq \frac{1}{4C_0}\left(\mathbb{E}\max_i\left\|Q_i^2\right\|^q\right)^{1/(2q)} \geq \frac{1}{4C_0}\left(\mathbb{E}\max_i\left\|\widetilde{G}_i^*\widetilde{G}_i\right\|^q\right)^{1/2q},\end{aligned}$$

where the last equality holds since $\left\|\widetilde{G}_i^*\widetilde{G}_i\right\| = \left\|\widetilde{G}_i\widetilde{G}_i^*\right\|$. The claim follows.

5.2.5 Proof of Lemma 3.12

Note that

$$\begin{aligned} &r^{3/2}\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)} \\ &\qquad\leq r^{3/2}\left(\mathbb{E}\sum_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)} \\ &\qquad= r^{3/2}\left(\mathbb{E}_1\sum_{i_1}\mathbb{E}_2\left\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\right)^{1/(2q)}. \end{aligned} \tag{5.17}$$

Next, Lemma 5.5 implies that, for $r = \max(q, \log(d))$,

$$\begin{aligned}\mathbb{E}_2\left\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q \leq 2^{2q-1}\Bigg[&\left\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q \\ &+ (2\sqrt{2e})^{2q} r^q\,\mathbb{E}_2\max_{i_2:i_2\neq i_1}\left\|H^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\Bigg].\end{aligned} \tag{5.18}$$

We will now apply Lemma 5.6 with $\alpha = 1$ and $\xi_{i_1} := \left\|\sum_{i_2\neq i_1}\mathbb{E}_2 H^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|$ to get that

$$\sum_{i_1} \mathbb{E}\xi_{i_1}^q \le 2(1+q)\Bigg(\mathbb{E}\max_{i_1}\Bigg\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\Bigg\|^q \\ + q^{-q}\left(\sum_{i_1}\mathbb{E}\Bigg\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\Bigg\|\right)^q\Bigg). \quad (5.19)$$

Combining (5.17) with (5.18) and (5.19), we obtain (using the inequality $1+q\le e^q$) that

$$r^{3/2}\left(\mathbb{E}\max_{i_1}\Bigg\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\Bigg\|^q\right)^{1/(2q)} \\ \le 4e\sqrt{2}\Bigg[r^{3/2}\left(\mathbb{E}\max_{i_1}\Bigg\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{2}^{(2)}\right)\Bigg\|^q\right)^{1/2q} \\ + r\sqrt{1+\frac{\log d}{q}}\left(\sum_{i_1}\mathbb{E}\Bigg\|\sum_{i_2:i_2\neq i_1}\mathbb{E}_2 H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\Bigg\|\right)^{1/2} \\ + r^2\left(\sum_{i_1}\mathbb{E}\max_{i_2:i_2\neq i_1}\left\|H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|^q\right)^{1/2q}\Bigg], \quad (5.20)$$

which yields the result.

5.3 Proof of Theorem 4.1

Let $J\subseteq I\subseteq\{1,2\}$. We will write $\mathbf{i}$ to denote the multi-index $(i_1,i_2)\in\{1,\dots,n\}^2$. We will also let $\mathbf{i}_I$ be the restriction of $\mathbf{i}$ onto its coordinates indexed by I, and, for a fixed value of $\mathbf{i}_{I^c}$, let $(H_{\mathbf{i}})_{\mathbf{i}_I}$ be the array $\left\{H_{\mathbf{i}},\ \mathbf{i}_I\in\{1,\dots,n\}^{|I|}\right\}$, where $H_{\mathbf{i}} := H_{i_1,i_2}(X_{i_1}^{(1)},X_{i_2}^{(2)})$. Finally, we let $\mathbb{E}_I$ stand for the expectation with respect to the variables with upper indices contained in I only. Following section 2 in [1], we define

$$\left\|(H_{\mathbf{i}})_{\mathbf{i}_I}\right\|_{I,J} = \mathbb{E}_{I\setminus J}\sup\Bigg\{\mathbb{E}_J\sum_{\mathbf{i}_I}\langle\Phi,H_{\mathbf{i}}\rangle\prod_{j\in J} f_{i_j}^{(j)}(X_{i_j}^{(j)}):\ \|\Phi\|_*\le 1, \\ f_i^{(j)}:S\mapsto\mathbb{R}\text{ for all } i,j,\text{ and }\sum_i\mathbb{E}\left|f_i^{(j)}(X_i^{(j)})\right|^2\le 1,\ j\in J\Bigg\} \quad (5.21)$$

and $\left\|(H_{\mathbf{i}})_{\mathbf{i}_\emptyset}\right\|_{\emptyset,\emptyset} := \|H_{\mathbf{i}}\|$, where $\langle A_1,A_2\rangle := \mathrm{tr}(A_1\,A_2^*)$ for $A_1,A_2\in\mathbb{H}^d$ and $\|\cdot\|_*$ denotes the nuclear norm. Theorem 1 in [1] states that for all $q\ge 1$,

$$\left(\mathbb{E}\|U_n\|^{2q}\right)^{1/2q}\le C\left[\sum_{I\subseteq\{1,2\}}\sum_{J\subseteq I} q^{|J|/2+|I^c|}\left(\sum_{\mathbf{i}_{I^c}}\mathbb{E}_{I^c}\left\|(H_{\mathbf{i}})_{\mathbf{i}_I}\right\|_{I,J}^{2q}\right)^{1/2q}\right],$$

where C is an absolute constant. Obtaining upper bounds for each term in the sum above, we get that

$$\left(\mathbb{E}\|U_n\|^{2q}\right)^{1/2q}\le C\left[\mathbb{E}\|U_n\|+\sqrt{q}\cdot A+q\cdot B+q^{3/2}\cdot\Gamma+q^2\cdot D\right],$$

where

$$
\begin{aligned}
A \le & 2\,\mathbb{E}_1 \left(\sup_{\Phi:\|\Phi\|_*\le 1} \sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 \right)^{1/2}, \\
B \le & \left(\sup_{\Phi:\|\Phi\|_*\le 1} \sum_{(i_1,i_2)\in I_n^2} \mathbb{E} \left\langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 \right)^{1/2} \\
& + 2 \left(\sum_{i_2} \mathbb{E}_2 \left(\mathbb{E}_1 \sup_{\Phi:\|\Phi\|_*\le 1} \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle \right)^{2q} \right)^{1/2q}, \\
\Gamma \le & 2 \left(\sum_{i_2} \mathbb{E}_2 \left(\sup_{\Phi:\|\Phi\|_*\le 1} \sum_{i_1} \mathbb{E}_1 \left\langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 \right)^q \right)^{1/2q}, \\
D \le & \left(\sum_{(i_1,i_2)\in I_n^2} \mathbb{E} \left\| H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \right\|^{2q} \right)^{1/2q}.
\end{aligned}
$$

The bounds for A, B, Γ, D above are obtained from (5.21) via the Cauchy-Schwarz inequality. For instance, to get a bound for A, note that it corresponds to the choice $I = \{1,2\}$ and $J = \{1\}$ or $J = \{2\}$. Due to symmetry of the kernels, it suffices to consider the case $J = \{2\}$, and multiply the upper bound by a factor of 2. When $J = \{2\}$,

$$
\begin{aligned}
& \left(\sum_{\mathbf{i}_{I^c}} \mathbb{E}_{I^c} \left\| (H_{\mathbf{i}})_{\mathbf{i}_I} \right\|_{I,J}^{2q} \right)^{1/2q} = \left\| (H_{\mathbf{i}})_{\mathbf{i}_{\{1,2\}}} \right\|_{\{1,2\},\{2\}} \\
&= \mathbb{E}_1 \sup \left\{ \mathbb{E}_2 \sum_{(i_1,i_2)\in I_n^2} \langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \rangle \cdot f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) : \|\Phi\|_* \le 1,\ \sum_{i_2} \mathbb{E} \left| f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) \right|^2 \le 1 \right\} \\
&= \mathbb{E}_1 \sup \left\{ \mathbb{E}_2 \sum_{i_2} \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle \cdot f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) : \|\Phi\|_* \le 1,\ \sum_{i_2} \mathbb{E} \left| f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) \right|^2 \le 1 \right\} \\
&\le \mathbb{E}_1 \sup_{\Phi, f^{(2)}_1, \ldots, f^{(2)}_n} \left\{ \sum_{i_2} \sqrt{ \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 } \sqrt{ \mathbb{E} \left| f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) \right|^2 } \right\} \\
&\le \mathbb{E}_1 \sup \left\{ \left(\sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 \right)^{1/2} : \|\Phi\|_* \le 1 \right\} \\
&= \mathbb{E}_1 \left(\sup_{\Phi:\|\Phi\|_*\le 1} \sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 \right)^{1/2}.
\end{aligned}
$$

It is not hard to see that the inequality above is in fact an equality, and it is attained by setting, for every fixed Φ,

$$
f^{(2)}_{i_2}\left(X^{(2)}_{i_2}\right) = \alpha_{i_2} \frac{\left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle}{\sqrt{ \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2 }},
$$

where $\alpha_{i_2} = \frac{\sqrt{\mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2}}{\sqrt{\sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2}}$ are such that $\sum_{i_2} \alpha_{i_2}^2 = 1$. The bounds for other terms are obtained quite similarly. Next, we will further simplify the upper bounds for A, B, Γ, D by analyzing the supremum over Φ with nuclear norm not exceeding 1. To this end, note that

$$\Phi \mapsto \mathbb{E}\left\langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi\right\rangle^2$$

is a convex function, hence its maximum over the convex set $\{\Phi \in \mathbb{H}^d : \ \|\Phi\|_* \le 1\}$ is attained at an extreme point that in the case of a unit ball for the nuclear norm must be a rank-1 matrix of the form $\phi\phi^*$ for some $\phi \in \mathbb{C}^d$. It implies that

$$\begin{aligned}\sup_{\Phi:\|\Phi\|_*\le 1} \mathbb{E}\left\langle H_{i_1,i_2}(X^{(1)}_1, X^{(2)}_2), \Phi\right\rangle^2 &\le \sup_{\phi:\|\phi\|_2\le 1} \mathbb{E}\left\langle H_{i_1,i_2}(X^{(1)}_1, X^{(2)}_2), \phi\phi^*\right\rangle^2 \\ &\le \left\| \mathbb{E} H^2_{i_1,i_2}(X^{(1)}_1, X^{(2)}_2)\right\|. \end{aligned} \tag{5.22}$$

Moreover,

$$\begin{aligned} &\sum_{i_2} \mathbb{E}_2 \left(\mathbb{E}_1 \sup_{\Phi:\|\Phi\|_*\le 1} \left\langle \sum_{i_1} H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle \right)^{2q} \\ &\quad = \sum_{i_2} \mathbb{E}_2 \left(\mathbb{E}_1 \left\| \sum_{i_1} H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\| \right)^{2q} \\ &\quad \le \sum_{i_2} \mathbb{E}_2\, \mathbb{E}_1 \left\| \sum_{i_1} H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^{2q} \\ &\quad \le \sum_{i_2} \mathbb{E}_2 \Bigg(2\sqrt{er} \left\| \sum_{i_1} \mathbb{E}_1 H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^{1/2} \\ &\qquad + 4\sqrt{2}er \left(\mathbb{E}_1 \max_{i_1} \left\| H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^{2q} \right)^{1/2q} \Bigg)^{2q}, \end{aligned} \tag{5.23}$$

where we have used Lemma 3.1 in the last step, and $r = q \vee \log d$. Combining (5.22), (5.23), we get that

$$\begin{aligned} B \le &\left(\left\| \sum_{(i_1,i_2)\in I^2_n} \mathbb{E} H^2_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \right\| \right)^{1/2} \\ &+ 4\sqrt{er} \left(\sum_{i_2} \mathbb{E}_2 \left\| \sum_{i_1} \mathbb{E}_1 H^2_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^q \right)^{1/2q} \\ &+ 8\sqrt{2}er \left(\sum_{i_2} \mathbb{E} \max_{i_1} \left\| H_{i_1,i_2}\left(X^{(1)}_{i_1}, X^{(2)}_{i_2}\right) \right\|^{2q} \right)^{1/2q}. \end{aligned} \tag{5.24}$$

It is also easy to get the bound for Γ: first, recall that

$$\Phi \mapsto \sum_{i_1} \mathbb{E}_1 \left\langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \Phi \right\rangle^2$$

is a convex function, hence its maximum over the convex set $\{\Phi \in \mathbb{H}^d : \ \|\Phi\|_* \le 1\}$ is attained at an extreme point of the form $\phi\phi^*$ for some unit vector ϕ. Moreover,

$$\begin{aligned} \left\langle H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}), \phi\phi^* \right\rangle^2 &= \phi^* H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \phi\phi^* H_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \phi \\ &\le \phi^* H^2_{i_1,i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \phi \end{aligned}$$

due to the fact that $\phi\phi^* \preceq I$. Hence

$$\sup_{\Phi:\|\Phi\|_*\le 1}\sum_{i_1}\mathbb{E}_1\left\langle H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2}),\Phi\right\rangle^2 \le \sup_{\phi:\ \|\phi\|_2=1}\sum_{i_1}\mathbb{E}_1\left(\phi^* H^2_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})\phi\right)$$
$$= \sup_{\phi:\ \|\phi\|_2=1}\phi^*\left(\mathbb{E}_1\sum_{i_1}H^2_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})\right)\phi = \left\|\mathbb{E}_1\sum_{i_1}H^2_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})\right\|,$$

and we conclude that

$$\Gamma \le 2\left(\sum_{i_2}\mathbb{E}_2\left(\sup_{\Phi:\|\Phi\|_*\le 1}\sum_{i_1}\mathbb{E}_1\left\langle H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2}),\Phi\right\rangle^2\right)^q\right)^{1/2q}$$
$$\le 2\left(\sum_{i_2}\mathbb{E}_2\left\|\sum_{i_1}\mathbb{E}_1 H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|^q\right)^{1/2q}. \tag{5.25}$$

The bound for A requires a bit more work. The following inequality holds:

Lemma 5.11. *The following inequality holds:*

$$A \le 2\mathbb{E}_1\left(\sup_{\Phi:\|\Phi\|_*\le 1}\sum_{i_2}\mathbb{E}_2\left\langle\sum_{i_1}H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2}),\Phi\right\rangle^2\right)^{1/2} \le$$
$$64\sqrt{e}\log(de)\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\ne i_1}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2}$$
$$+8\sqrt{2e\log(de)}\left(\mathbb{E}\left\|\sum_i\mathbb{E}_2\widetilde{G}_i\widetilde{G}_i^*\right\|+\left\|\sum_{(i_1,i_2)\in I_n^2}\mathbb{E}H^2_{i_1,i_2}\left(X^{(1)}_{i_1},X^{(2)}_{i_2}\right)\right\|\right)^{1/2},$$

where $\widetilde{G}$ was defined in (3.6).

Combining the bounds (5.24), (5.25) and Lemma 5.11, and grouping the terms with the same power of q, we get the result of Theorem 4.1.

It remains to prove Lemma 5.11. To this end, note that Jensen's inequality and an argument similar to (5.22) imply that

$$\mathbb{E}_1\left(\sup_{\Phi:\|\Phi\|_*\le 1}\sum_{i_2}\mathbb{E}_2\left\langle\sum_{i_1}H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2}),\Phi\right\rangle^2\right)^{1/2}$$
$$\le\left(\mathbb{E}\sup_{\Phi:\|\Phi\|_*\le 1}\sum_{i_2}\left\langle\sum_{i_1}H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2}),\Phi\right\rangle^2\right)^{1/2}$$
$$\le\left(\mathbb{E}\left\|\sum_{i_2}\left(\sum_{i_1}H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})\right)^2\right\|\right)^{1/2}.$$

Next, arguing as in the proof of Lemma 3.3, we define

$$\widehat{B}_{i_1,i_2} = [0\mid 0\mid\ldots\mid H_{i_1,i_2}(X^{(1)}_{i_1},X^{(2)}_{i_2})\mid\ldots\mid 0\mid 0]\in\mathbb{R}^{d\times nd},$$

where H_{i_1,i_2} sits on the i_1-th position of the block matrix above. Moreover, let

$$B_{i_2} = \sum_{i_1:i_1\ne i_2}\widehat{B}_{i_1,i_2}. \tag{5.26}$$

Using the representation (5.26), we have

$$\left(\mathbb{E}\left\|\sum_{i_2}\left(\sum_{i_1} H_{i_1,i_2}(X_{i_1}^{(1)},X_{i_2}^{(2)})\right)^2\right\|\right)^{1/2} = \left(\mathbb{E}\left\|\left(\sum_{i_2} B_{i_2}\right)\left(\sum_{i_2} B_{i_2}\right)^*\right\|\right)^{1/2}$$
$$= \left(\mathbb{E}\left\|\left(\sum_{i_2}\begin{pmatrix} 0 & B_{i_2}^* \\ B_{i_2} & 0\end{pmatrix}\right)^2\right\|\right)^{1/2} \leq 2\left(\mathbb{E}\left\|\left(\sum_{i_2}\varepsilon_{i_2}\begin{pmatrix} 0 & B_{i_2}^* \\ B_{i_2} & 0\end{pmatrix}\right)\right\|^2\right)^{1/2},$$

where $\{\varepsilon_{i_2}\}_{i_2=1}^n$ is sequence of i.i.d. Rademacher random variables, and the last step follows from the symmetrization inequality. Next, Khintchine's inequality (3.2) yields that

$$\begin{aligned}
A \leq & 4\sqrt{e(1+2\log d)}\left(\mathbb{E}\left\|\sum_{i_2}\begin{pmatrix} B_{i_2}^*B_{i_2} & 0 \\ 0 & B_{i_2}B_{i_2}^*\end{pmatrix}\right\|\right)^{1/2} \\
= & 4\sqrt{e(1+2\log d)}\left(\mathbb{E}\max\left\{\left\|\sum_{i_2} B_{i_2}^*B_{i_2}\right\|, \left\|\sum_{i_2} B_{i_2}B_{i_2}^*\right\|\right\}\right)^{1/2} \\
= & 4\sqrt{e(1+2\log d)}\left(\mathbb{E}\max\left\{\left\|\widetilde{G}\widetilde{G}^*\right\|, \left\|\sum_{(i_1,i_2)\in I_n^2} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|\right\}\right)^{1/2}.
\end{aligned}$$

Note that the last expression is of the same form as equation (5.10) in the proof of Theorem 3.7 with $q=1$. Repeating the same argument, one can show that

$$\begin{aligned}
A \leq & 64\sqrt{e}\log(de)\left(\mathbb{E}\max_{i_1}\left\|\sum_{i_2:i_2\neq i_1} H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|\right)^{1/2} \\
& + 8\sqrt{2e\log(de)}\left(\mathbb{E}\left\|\sum_i \mathbb{E}_2\widetilde{G}_i\widetilde{G}_i^*\right\| + \left\|\sum_{(i_1,i_2)\in I_n^2}\mathbb{E}H_{i_1,i_2}^2\left(X_{i_1}^{(1)},X_{i_2}^{(2)}\right)\right\|\right)^{1/2},
\end{aligned}$$

which is an analogue of (5.15).

5.4 Calculations related to Examples 3.5 and 3.6

We will first estimate $\|GG^*\|$. Note that the (i,i)-th block of the matrix GG^* is

$$(GG^*)_{ii} = \sum_{j:j\neq i} A_{i,j}^2 = \sum_{j:j\neq i}\left(\mathbf{a}_i\mathbf{a}_j^T + \mathbf{a}_j\mathbf{a}_i^T\right)^2 = (n-1)\mathbf{a}_i\mathbf{a}_i^T + \sum_{j\neq i}\mathbf{a}_j\mathbf{a}_j^T.$$

The (i,j)-block for $j\neq i$ is

$$(GG^*)_{ij} = \sum_{k\neq i,j} A_{i,k}A_{j,k} = \sum_{k\neq i,j}\left(\mathbf{a}_i\mathbf{a}_k^T + \mathbf{a}_k\mathbf{a}_i^T\right)\left(\mathbf{a}_j\mathbf{a}_k^T + \mathbf{a}_k\mathbf{a}_j^T\right) = (n-2)\mathbf{a}_i\mathbf{a}_j^T.$$

We thus obtain that

$$GG^* = (n-2)\mathbf{a}\mathbf{a}^T + \mathrm{Diag}\left(\underbrace{\sum_{j=1}^n \mathbf{a}_j\mathbf{a}_j^T,\ldots,\sum_{j=1}^n \mathbf{a}_j\mathbf{a}_j^T}_{\text{n terms}}\right),$$

where $\mathrm{Diag}(\cdot)$ denotes the block-diagonal matrix with diagonal blocks in the brackets. Since

$$\mathrm{Diag}\left(\sum_{j=1}^{n}\mathbf{a}_j\mathbf{a}_j^T,\ldots,\sum_{j=1}^{n}\mathbf{a}_j\mathbf{a}_j^T\right)\succeq 0,$$

it follows that

$$\|GG^*\|\geq (n-2)\|\mathbf{a}\|_2^2=(n-2)n.$$

On the other hand,

$$\begin{aligned}\left\|\sum_{(i_1,i_2)\in I_n^2}A_{i_1,i_2}^2\right\| &= \left\|\sum_{(i_1,i_2)\in I_n^2}\left(\mathbf{a}_{i_1}\mathbf{a}_{i_2}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_1}^T\right)^2\right\| = \left\|\sum_{(i_1,i_2)\in I_n^2}\left(\mathbf{a}_{i_1}\mathbf{a}_{i_1}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_2}^T\right)\right\| \\ &= 2(n-1)\left\|\sum_i \mathbf{a}_i\mathbf{a}_i^T\right\| = 2(n-1),\end{aligned}$$

where the last equality follows from the fact that $\{\mathbf{a}_1,\ldots,\mathbf{a}_n\}$ are orthonormal.

For Example 3.6, we similarly obtain that

$$\begin{aligned}(GG^*)_{ii} &= \sum_{j:j\neq i}c_{i,j}^2\left(\mathbf{a}_i\mathbf{a}_j^T+\mathbf{a}_j\mathbf{a}_i^T\right)^2 \\ &= \left(\sum_{j:j\neq i}c_{i,j}^2\right)\mathbf{a}_i\mathbf{a}_i^T+\sum_{j:j\neq i}c_{i,j}^2\mathbf{a}_j\mathbf{a}_j^T = \mathbf{a}_i\mathbf{a}_i^T+\sum_{j:j\neq i}c_{i,j}^2\mathbf{a}_j\mathbf{a}_j^T, \\ (GG^*)_{ij} &= \sum_{k\neq i,j}c_{i,k}c_{j,k}\left(\mathbf{a}_i\mathbf{a}_k^T+\mathbf{a}_k\mathbf{a}_i^T\right)\left(\mathbf{a}_j\mathbf{a}_k^T+\mathbf{a}_k\mathbf{a}_j^T\right)=0,\ i\neq j,\end{aligned}$$

hence $\|GG^*\|=\max_i\|(GG^*)_{ii}\|=1$. On the other hand,

$$\begin{aligned}\left\|\sum_{(i_1,i_2)\in I_n^2}A_{i_1,i_2}^2\right\| &= \left\|\sum_{(i_1,i_2)\in I_n^2}c_{i_1,i_2}^2\left(\mathbf{a}_{i_1}\mathbf{a}_{i_2}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_1}^T\right)^2\right\| \\ &= \left\|\sum_{(i_1,i_2)\in I_n^2}c_{i_1,i_2}^2\left(\mathbf{a}_{i_1}\mathbf{a}_{i_1}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_2}^T\right)\right\| = 2\left\|\sum_i\mathbf{a}_i\mathbf{a}_i^T\right\|=2,\end{aligned}$$

and

$$\begin{aligned}\sum_{i_1}\left\|\sum_{i_2:i_2\neq i_1}A_{i_1,i_2}^2\right\| &= \sum_{i_1}\left\|\sum_{i_2\neq i_1}c_{i_1,i_2}^2\left(\mathbf{a}_{i_1}\mathbf{a}_{i_2}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_1}^T\right)^2\right\| \\ &= \sum_{i_1}\left\|\sum_{i_2:i_2\neq i_1}c_{i_1,i_2}^2\left(\mathbf{a}_{i_1}\mathbf{a}_{i_1}^T+\mathbf{a}_{i_2}\mathbf{a}_{i_2}^T\right)\right\| = \sum_{i_1}\left\|\mathbf{a}_{i_1}\mathbf{a}_{i_1}^T+\sum_{i_2:i_2\neq i_1}c_{i_1,i_2}^2\mathbf{a}_{i_2}\mathbf{a}_{i_2}^T\right\|=n.\end{aligned}$$

References

[1] Radoslaw Adamczak, *Moment inequalities for U-statistics*, Ann. Probab. **34** (2006), no. 6, 2288–2314. MR-2294982

[2] Radoslaw Adamczak and Rafal Latala, *The LIL for U-statistics in Hilbert spaces*, Journal of Theoretical Probability **21** (2008), no. 3, 704–744. MR-2425366

[3] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, 2013. MR-3185193

[4] Jean-Christophe Bourin and Eun-Young Lee, *Unitary orbits of hermitian operators with convex or concave functions*, Bulletin of the London Mathematical Society **44** (2012), no. 6, 1085–1102. MR-3007642

[5] Richard Y. Chen, Alex Gittens, and Joel A. Tropp, *The masked sample covariance estimator: an analysis using matrix concentration inequalities*, Information and Inference (2012). MR-3311439

[6] Xiaohui Chen, *On bootstrap approximations for high-dimensional U-statistics and random quadratic forms*, arXiv preprint arXiv:1610.00032 (2016).

[7] V. de la Pena and E. Gine, *Decoupling: From dependence to independence*, Springer-Verlag, New York, 1999. MR-1666908

[8] V. de la Pena and S. J. Montgomery-Smith, *Decoupling inequalities for the tail probabilities of multivariate U-statistics*, Annals of Probability **23** (1995), no. 2, 806–816. MR-1334173

[9] Sjoerd Dirksen et al., *Tail bounds via generic chaining*, Electronic Journal of Probability **20** (2015). MR-3354613

[10] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, vol. 1, Birkhäuser Basel, 2013. MR-3100033

[11] Evarist Giné, Stanislaw Kwapien, Rafal Latala, and Joel Zinn, *The LIL for canonical U-statistics of order 2*, Annals of Probability (2001), 520–557. MR-1825163

[12] Evarist Gine, Rafal Latala, and Joel Zinn, *Exponential and moment inequalities for U-statistics*, High Dimensional Probability II (2000), 13–38. MR-1857312

[13] Evarist Giné and Richard Nickl, *Mathematical foundations of infinite-dimensional statistical models*, vol. 40, Cambridge University Press, 2016. MR-3588285

[14] Evarist Giné and Joel Zinn, *On Hoffmann-Jørgensen's inequality for U-processes*, Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference, Springer, 1992, pp. 80–91. MR-1227611

[15] Fang Han and Han Liu, *Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution*, Bernoulli: Official Journal of the Bernoulli Society for Mathematical Statistics and Probability **23** (2017), no. 1, 23. MR-3556765

[16] Wassily Hoeffding, *A class of statistics with asymptotically normal distribution*, The Annals of Mathematical Statistics (1948), 293–325. MR-0026294

[17] Christian Houdré and Patricia Reynaud-Bouret, *Exponential inequalities, with constants, for U-statistics of order two*, Stochastic inequalities and applications, Springer, 2003, pp. 55–69. MR-2073426

[18] Rustam Ibragimov and Sh. Sharakhmetov, *Analogues of Khintchine, Marcinkiewicz–Zygmund and Rosenthal inequalities for symmetric statistics*, Scandinavian Journal of Statistics **26** (1999), no. 4, 621–633. MR-1734266

[19] Marius Junge, Qiang Zeng, et al., *Noncommutative Bennett and Rosenthal inequalities*, The Annals of Probability **41** (2013), no. 6, 4287–4316. MR-3161475

[20] Vladimir S. Korolyuk and Yu V. Borovskich, *Theory of U-statistics*, vol. 273, Springer Science & Business Media, 2013. MR-1472486

[21] Jeanne Kowalski and Xin M. Tu, *Modern applied U-statistics*, vol. 714, John Wiley & Sons, 2008. MR-2368050

[22] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer-Verlag, Berlin, 2011. MR-2814399

[23] Françoise Lust-Piquard, *Inégalités de Khintchine dans c_p $(1 < p < \infty)$*, CR Acad. Sci. Paris **303** (1986), 289–292. MR-0859804

[24] Françoise Lust-Piquard and Gilles Pisier, *Noncommutative Khintchine and Paley inequalities*, Arkiv för Matematik **29** (1991), no. 1, 241–260. MR-1150376

[25] Lester Mackey, Michael I. Jordan, Richard Y. Chen, Brendan Farrell, Joel A. Tropp, et al., *Matrix concentration inequalities via the method of exchangeable pairs*, The Annals of Probability **42** (2014), no. 3, 906–945. MR-3189061

[26] Stanislav Minsker and Xiaohan Wei, *Robust modifications of U-statistics and applications to covariance estimation problems*, arXiv preprint arXiv:1801.05565 (2018).

[27] Gilles Pisier, *Non-commutative vector-valued l_p spaces and completely p-summing maps*, Asterisque – Societe Mathematique de France **247** (1998). MR-1648908

[28] H. Rauhut, *Compressive sensing and structured random matrices*, Radon Series Comp. Appl. Math (2012), 1–94. MR-2731597

[29] Holger Rauhut, *Circulant and Toeplitz matrices in compressed sensing*, arXiv preprint arXiv:0902.4394 (2009).

[30] Robert J. Serfling, *Approximation theorems of mathematical statistics*, vol. 162, John Wiley & Sons, 2009. MR-0595165

[31] J. A. Tropp, *An introduction to matrix concentration inequalities*, Foundations and Trends in Machine Learning (2015), 1–230. MR-3311439

[32] Joel Tropp, *On the conditioning of random subdictionaries*, Applied and Computational Harmonic Analysis **25** (2008), no. 1, 1–24. MR-2419702

[33] Joel A. Tropp, *User-friendly tail bounds for sums of random matrices*, Foundations of Computational Mathematics **12** (2012), no. 4, 389–434. MR-2946459

[34] Joel A. Tropp, *The expected norm of a sum of independent random matrices: an elementary approach*, High Dimensional Probability VII, Springer, 2016, pp. 173–202. MR-3565264

[35] Joel A. Tropp, *Second-order matrix concentration inequalities*, Applied and Computational Harmonic Analysis (2016). MR-3768857

[36] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, arXiv preprint arXiv:1011.3027 (2010). MR-2963170

[37] Marten Wegkamp and Yue Zhao, *Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas*, Bernoulli **22** (2016), no. 2, 1184–1226. MR-3449812