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

Stanislav Minsker*† and 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.

1. Introduction.

Since being introduced by W. Hoeffding (1948), 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 the excellent monographs by de la Pena and Gine (1999); Korolyuk and Borovskich (2013); Serfling (2009); Kowalski and Tu (2008).

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 of de la Pena and Montgomery-Smith (1995); de la Pena and Gine (1999); Gine, Latala and Zinn (2000); Giné and Zinn (1992); Ibragimov and Sharakhmetov (1999); Giné et al. (2001); Houdré and Reynaud-Bouret (2003), among others. However, 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 (Wegkamp and Zhao, 2016; Han and Liu, 2017).

The goal of this 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. The emphasis is made on expressing the upper bounds in terms of computable parameters. Our results extend the (variant of) matrix Rosenthal inequality for the sums of independent random matrices due to Chen, Gittens and Tropp (2012) (see also Junge et al., 2013; Mackey et al., 2014) to the framework of 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 Adamczak (2006), 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 by Chen (2016), 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 by Minsker and Wei (2018) investigates robust estimators of covariance matrices based on U-statistics, but deals only with the case of non-degenerate U-statistics 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.

*Department of Mathematics, University of Southern California
†Department of Electrical Engineering, University of Southern California
‡Authors gratefully acknowledge support by the National Science Foundation grant DMS-1712956.
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 \( \| A \|_p = (\text{tr}(A^*A)^{p/2})^{1/p} \). 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 \( \| \cdot \|_F \); 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 \( A \succeq B \) (or \( A > B \)) if \( A - B \) 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 the rectangular matrix \( A \in \mathbb{C}^{d_1 \times d_2} \), the Hermitian dilation \( \mathcal{D} : \mathbb{C}^{d_1 \times d_2} \rightarrow \mathbb{C}^{(d_1 + d_2) \times (d_1 + d_2)} \) is defined as

\[
\mathcal{D}(A) = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}.
\]

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 the 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 6 contains the technical background and the 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 \( H_{i_1, \ldots, i_m} : \mathcal{S}^m \rightarrow \mathbb{H}^d, (i_1, \ldots, i_m) \in I_n^m, 2 \leq m \leq n \), are \( \mathcal{S}^m \)-measurable, permutation symmetric kernels, meaning that \( H_{i_1, \ldots, i_m}(x_1, \ldots, x_m) = H_{i_1, \ldots, i_m}(x_{\pi_1}, \ldots, x_{\pi_m}) \) for any \( (x_1, \ldots, x_m) \in \mathcal{S}^m \) and any permutation \( \pi \). The generalized U-statistic is defined as (de la Pena and Gine, 1999)

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

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

\[
U'_n = \sum_{(i_1, \ldots, i_m) \in I_n^m} H_{i_1, \ldots, i_m}(X_{i_1}^{(1)}, \ldots, X_{i_m}^{(m)}),
\]
where \( \{ X_i^{(k)} \}_{i=1}^n, k = 1, \ldots, m \) are independent copies of \( X_1, \ldots, X_n \). 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 \( \pi_{m,k} \) \((k \leq m)\) is defined as

\[
\pi_{m,k} H(x_{i_1}, \ldots, x_{i_k}) := (\delta_{x_{i_1}} - P) \ldots (\delta_{x_{i_k}} - P) P^{m-k} H,
\]

where

\[
Q^m H := \int \ldots \int H(y_1, \ldots, y_m) dQ(y_1) \ldots dQ(y_m),
\]

for any probability measure \( Q \) in \((S, B)\), where \( B \) denotes the Borel \( \sigma \)-algebra on \( S \), and \( \delta_x \) is a Dirac measure concentrated at \( x \in S \). For example, \( \pi_{m,1} H(x) = \mathbb{E} [H(X_1, \ldots, X_m) | X_1 = x] - \mathbb{E} H(X_1, \ldots, X_m) \).

**Definition 2.2.** An \( S^m \)-measurable function \( F : S^m \rightarrow \mathbb{R}^d \) is \( P \)-degenerate of order \( r \) \((1 \leq r < m)\), if

\[
\mathbb{E} F(x_1, \ldots, x_r, X_{r+1}, \ldots, X_m) = 0, \quad \forall x_1, \ldots, x_r \in S,
\]

and \( \mathbb{E} F(x_1, \ldots, x_r, X_{r+1}, X_{r+2}, \ldots, X_m) \) 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 \( m - 1 \), that it is called completely degenerate. From now on, we will only consider generalized U-statistics of order \( m = 2 \) with completely degenerate (that is, of order 1) kernels. The case of non-degenerate U-statistics is easily reduced to the degenerate case via the Hoeffding’s decomposition; see (de la Pena and Gine, 1999, page 137) for details.

### 3. Main results.

Rosenthal-type moment inequalities for sum of independent matrices have appeared in a number of works including (Chen, Gittens and Tropp, 2012; Mackey et al., 2014; Tropp, 2015). Specifically, the following inequality follows from Theorem A.1 of (Chen, Gittens and Tropp, 2012):

**Lemma 3.1** (Matrix Rosenthal inequality). Suppose that \( q \geq 1 \) is an integer and fix \( r \geq q \sqrt{\log d} \). Consider a finite sequence of \( \{Y_i\} \) of independent copies of \( Y \in \mathbb{R}^d \). Then

\[
\left( \mathbb{E} \left[ \left( \sum_i (Y_i - \mathbb{E} Y_i) \right)^2 \right] \right)^{1/2q} \leq 2\sqrt{er} \left( \sum_i \mathbb{E} (Y_i - \mathbb{E} Y_i)^2 \right)^{1/2} + 4\sqrt{2er} \left( \max_i \mathbb{E} Y_i - \mathbb{E} \mathbb{E} Y_i \right)^2 \right)^{1/2q}.
\]

The bound above improves upon the moment inequality that follows from the matrix Bernstein’s inequality (Tropp, 2015):

**Lemma 3.2** (Matrix Bernstein’s inequality). Consider a finite sequence of \( \{Y_i\} \) of independent \( \mathbb{R}^d \)-valued random matrices such that \( \| Y_i - \mathbb{E} Y_i \| \leq B \). Then

\[
\text{Pr} \left( \sum_i (Y_i - \mathbb{E} Y_i) \geq \sigma \sqrt{u} + Bu \right) \leq 2e^{-u},
\]

where \( \sigma^2 := \sum_i \mathbb{E} (Y_i - \mathbb{E} Y_i)^2 \).

Indeed, Lemma 6.8 implies, after some simple algebra, that

\[
\left( \mathbb{E} \left[ \left( \sum_i Y_i - \mathbb{E} Y_i \right)^q \right] \right)^{1/q} \leq C \left( \sqrt{q + \log(2d)} \sigma + (q + \log(2d))B \right),
\]

where
for an absolute constant $C > 0$ and all $q \geq 1$. This bound is weaker than (3) as it requires almost sure boundedness of $\|Y_i - \mathbb{E}Y_i\|$ for all $i$. One the the main goal of this work is to obtain operator norm bounds similar to (3) for $\mathbb{H}^d$-valued U-statistics of order 2.

### 3.1. Degenerate U-statistics of order 2.

The moment bounds for scalar U-statistics are well-known, see for example the work by Gine, Latala and Zinn (2000) and references therein. Moment inequalities for general Banach-space valued U-statistics were obtained by Adamczak (2006). 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 Adamczak 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 matrix Rademacher series, see (Tropp, 2015, 2016a,b; Vershynin, 2010). 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

\[
\left( \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i A_i \right\|^2 \right)^{1/2} \leq \sqrt{c} (1 + 2 \log d) \cdot \left\| \sum_{i=1}^{n} A_i^2 \right\|^{1/2}. \tag{4}
\]

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 $\left\| \sum_{i=1}^{n} A_i^2 \right\|$.

**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 $\{\varepsilon_j^{(i)}\}_{j \in \mathbb{N}}$, $i = 1, 2$, is a sequence 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 $q \geq 1$,

\[
\max \left\{ \left\| GG^T \right\|, \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\| \right\}^{1/2} \leq \left( \mathbb{E} \|X\|^{2q} \right)^{1/(2q)} \leq \frac{4}{\sqrt{e}} \cdot \max \left\{ \left\| GG^T \right\|, \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\| \right\}^{1/2}, \tag{5}
\]

where $r := q \vee \log d$, and the matrix $G \in \mathbb{R}^{(nd) \times (nd)}$ is defined as

\[
G := \begin{pmatrix}
0 & A_{1,2} & \cdots & A_{1,n} \\
A_{2,1} & 0 & \cdots & A_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
A_{n,1} & A_{n,2} & \cdots & 0
\end{pmatrix}.
\]

**Remark 3.1** (Constants in Lemma 3.3). Matrix Rademacher chaos of order 2 has been studied previously by Rauhut (2009), Pisier (1998) and Rauhut (2012), where Schatten-$p$ norm upper
bound were obtained by iterating the Khintchine’s inequality for Rademacher series. Specifically, the following bound holds for all \( p \geq 1 \) (see Lemma 6.4 for the details):

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

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

\[
\left( \mathbb{E} \|X\|^{2p} \right)^{1/(2q)} \leq C \max(q, \log(nd)) \max \left\{ \left\| GG^T \right\|, \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 \right\}^{1/2}
\]

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

One may wonder if the term \( \|GG^T\| \) 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 \( \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 \geq \|GG^T\| \). However, a more careful examination shows that there is no strict dominance among \( \|GG^T\| \) and \( \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 \). The following example demonstrates a situation where \( \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 < \|GG^T\| \).

**Example 1.** Assume that \( d \geq n \geq 2 \), let \( \{a_1, \ldots, a_d\} \) be any orthonormal basis in \( \mathbb{R}^d \), and \( a := [a_1^T, \ldots, a_n^T] \in \mathbb{R}^{nd} \) be the “vertical concatenation” of \( a_1, \ldots, a_n \). Define

\[
A_{i_1,i_2} := a_{i_1}a_{i_2}^T + a_{i_2}a_{i_1}^T, \quad i_1, i_2 \in \{1, 2, \ldots, n\},
\]

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 \( \|GG^T\| \geq (n - 2)\|a\|^2 = (n - 2)n \), and \( \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 = 2(n - 1) \). Details are outlined in Section 6.4.

It follows from Lemma 6.1 that

\[
\|GG^T\| \leq \sum_{i_1} \left\| \sum_{i_2,i_2 \neq i_1} A_{i_1,i_2}^2 \right\|.
\]

Often, this inequality yields a “computable” upper bound for the right-hand side of the inequality (5), however, in some cases it results in the loss of precision, as the following example demonstrates.

**Example 2.** Assume that \( n \) is even, \( d \geq n \geq 2 \), let \( \{a_1, \ldots, a_d\} \) be an orthonormal basis in \( \mathbb{R}^d \), and let \( C \in \mathbb{R}^{n \times n} \) be an orthogonal matrix with entries \( c_{i,j} \) such that \( c_{i,i} = 0 \) for all \( i \). Define

\[
A_{i_1,i_2} = c_{i_1,i_2} \left( a_{i_1}a_{i_2}^T + a_{i_2}a_{i_1}^T \right), \quad i_1, i_2 \in \{1, 2, \ldots, n\},
\]

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 \( \|GG^T\| = 1 \), \( \sum_{(i_1,i_2)\in I_n^2} A_{i_1,i_2}^2 = 2 \), but

\[
\sum_{i_1} \left\| \sum_{i_2,i_2 \neq i_1} A_{i_1,i_2}^2 \right\| = n.
\]

Details are outlined in Section 6.4.
3.3. Moment inequalities for degenerate U-statistics of order 2.

Let $H_{i_1,i_2} : S \times S \mapsto \mathbb{R}^d$, $(i_1,i_2) \in I_n^2$, be a sequence of degenerate kernels, for example, $H_{i_1,i_2}(x_1,x_2) = \pi_{2,2} \hat{H}_{i_1,i_2}(x_1,x_2)$ for some non-degenerate permutation-symmetric $\hat{H}_{i_1,i_2}$. Everywhere below, $E_j[\cdot]$, $j = 1,2$, stands for the expectation with respect to $\{X^{(j)}_i\}_{i=1}^n$ 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 3.1.** Let $\{X^{(j)}_i\}_{i=1}^n$, $j = 1,2$, be $S$-valued i.i.d. random variables, $H_{i,j} : S \times S \mapsto \mathbb{R}^d$ – permutation-symmetric degenerate kernels. Then for all $q \geq 1$ and $r = \max(q,\log d)$,

\[
\left( E \left\| U_n \right\|^{2q} \right)^{1/2q} \leq 4 \left( E \left\| \sum_{i_1,i_2 \in I_n^2} H_{i_1,i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{2q} \right)^{1/2q} \\
\leq 128/\sqrt{e} \left[ 2\sqrt{2er}^{3/2} \left( E \max_{i_1} \left\| \sum_{i_2 \neq i_1} H^2_{i_1,i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\| q \right]^{1/(2q)} \\
+ r \left\| \sum_{i_1,i_2 \in I_n^2} E H^2_{i_1,i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} + r \left( E \left\| \sum_i E \tilde{G}_{i} \tilde{G}_i^T \right\|^{q} \right)^{1/2q} ,
\]

where $\tilde{G}_i$ is the the $i$-th column of the matrix $\tilde{G} \in \mathbb{R}^{nd}$ defined as

\[
\tilde{G} := \begin{pmatrix}
0 & H_{1,2} \left( X^{(1)}_1, X^{(2)}_2 \right) & \cdots & H_{1,n} \left( X^{(1)}_1, X^{(2)}_n \right) \\
H_{2,1} \left( X^{(1)}_2, X^{(2)}_1 \right) & 0 & \cdots & H_{2,n} \left( X^{(1)}_2, X^{(2)}_n \right) \\
\vdots & \vdots & \ddots & \vdots \\
H_{n,1} \left( X^{(1)}_n, X^{(2)}_1 \right) & H_{n,2} \left( X^{(1)}_n, X^{(2)}_2 \right) & \cdots & 0
\end{pmatrix}.
\] (7)

**Proof.** See Section 6.2.3. □

The following lower bound (proven in Section 6.2.4) demonstrates that all the terms in the bound of Theorem 3.1 are necessary.

**Lemma 3.4.** Under the assumptions of Theorem 3.1,

\[
\left( E \left\| U_n(H) \right\|^{2q} \right)^{1/2q} \geq C \left[ \left( E \max_{i_1} \left\| \sum_{i_2 \neq i_1} H^2_{i_1,i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\| q \right]^{1/(2q)} \\
+ \left( E \left\| \sum_i E \tilde{G}_i \tilde{G}_i^T \right\|^{q} \right)^{1/2q} + \left( E \left\| \sum_{i_1,i_2 \in I_n^2} E H^2_{i_1,i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\| q \right]^{1/(2q)} \right]
\]

where $C > 0$ is an absolute constant.

Our next goal is to obtain more “user-friendly” versions of the upper bound, and we first focus on the term $E \left\| \sum_{i=1}^n E \tilde{G}_i \tilde{G}_i^T \right\|^{q} = E \left\| \tilde{G} \tilde{G}^T \right\|^q$ appearing in Theorem 3.1 that might be difficult to deal with directly. It is easy to see that the $(i,j)$-th block of the matrix $E \tilde{G} \tilde{G}^T$ is

\[
\left( E \tilde{G} \tilde{G}^T \right)_{i,j} = \sum_{k \neq i,j} E \left[ H_{i,k}(X^{(1)}_i, X^{(2)}_k)H_{j,k}(X^{(1)}_j, X^{(2)}_k) \right].
\]
It follows from Lemma 6.1 that
\[
\|E_2 \tilde{G} \tilde{G}^T\| \leq \sum_i \left\| \left( E_2 \tilde{G} \tilde{G}^T \right)_{ii} \right\| = \sum_i \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|, \tag{8}
\]
hence
\[
\left( E \left\| E_2 \tilde{G} \tilde{G}^T \right\|^q \right)^{1/2q} \leq \left( E \left( \sum_i \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^q \right)^{1/2q} \\
\leq \left( \sum_i E \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^{1/2} \\
+ 2\sqrt{2\epsilon q} \left( E \max_i \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/2q},
\]
where we used Rosenthal’s inequality (Lemma 6.5 applied with \( d = 1 \)) in the last step. Together with the fact that \( E H^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \leq E E_2 H^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \) for all \( i, j \), and
\[
\left( E \max_i \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/2q} \leq \left( E \max_i \left\| \sum_{i_2:i_2 \neq i_1} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)},
\]
we obtain the following inequality.

**Corollary 3.1.** Under the assumptions of Theorem 3.1,
\[
\left( E \left\| U_n \right\|^2 \right)^{1/2q} \leq 256/\sqrt{\epsilon} \left[ r \left( \sum_i E \left\| \sum_{i_2:i_2 \neq i_1} E_2 H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^{1/2} \\
+ 2\sqrt{2\epsilon r^{3/2}} \left( E \max_i \left\| \sum_{i_2:i_2 \neq i_1} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)} \right].
\]

**Remark 3.2.** Assume that \( H_{i,j} = H \) is independent of \( i, j \) and is such that \( \|H(x_1, x_2)\| \leq M \) for all \( x_1, x_2 \in S \). Then
\[
E \max_i \left\| \sum_{i_2:i_2 \neq i_1} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \leq (n - 1)^q M^{2q},
\]
and it immediately follows from Lemma 6.7 and Corollary 3.1 that for all \( t \geq 1 \) and an absolute constant \( C > 0 \),
\[
\Pr \left( \left\| U_n \right\| \geq C \left( \sqrt{E E_2 H^2 \left( X_{1}^{(1)}, X_{2}^{(2)} \right)} (t + \log d) \cdot n + M \sqrt{n} (t + \log d)^{3/2} \right) \right) \leq e^{-t}. \tag{9}
\]
Next, we obtain further refinements of the result that follow from estimating the term
\[
r^{3/2} \left( E \max_i \left\| \sum_{i_2:i_2 \neq i_1} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)}.
\]
Lemma 3.5. Under the assumptions of Theorem 3.1,

\[ r^{3/2} \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2: i_2 \neq i_1} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{q} \right)^{1/(2q)} \leq \frac{2e\sqrt{2}}{q} \sqrt{1 + \frac{\log d}{q}} \left( \mathbb{E} \left\| \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} + \frac{1}{\| I \|} \mathbb{E} \left\| \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) \]

for some absolute constant $C$. On the other hand, direct application of the non-commutative Khintchine inequality (4) followed by Rosenthal inequality (Lemma 6.5) only gives that

\[ \mathbb{E} \| U_n \| \leq C \log d \left( \mathbb{E} \left\| \mathbb{E} \tilde{G} \tilde{G}^T \right\|^{1/2} + \mathbb{E} \left\| \max_{i_1} \left\| \sum_{i_2: i_2 \neq i_1} \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) \]

\[ \leq C \log d \left( \left\| \mathbb{E} \left\| \mathbb{E} \tilde{G} \tilde{G}^T \right\|^{1/2} + \mathbb{E} \left\| \max_{i_1} \left\| \sum_{i_2: i_2 \neq i_1} \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) \right) \]

and it is easy to see that the right-hand side of (10) is never worse than the bound (11). To see that it can be strictly better, consider the framework of Example 2, where it is easy to see (following the same calculations as those given in Section 6.4) that

\[ \left( \mathbb{E} \left\| \mathbb{E} \tilde{G} \tilde{G}^T \right\|^{1/2} \right) = 1, \quad \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2: i_2 \neq i_1} \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) = 1, \]

\[ \left( \mathbb{E} \left\| \sum_{i_1} \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) = 2, \quad \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2: i_2 \neq i_1} \mathbb{E} H^2_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right\|^{1/2} \right) = \sqrt{n}. \]

Remark 3.3 (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 (1) of $U_n$, namely

\[ \mathbb{D} \left( U_n \right) = \sum_{(i_1, i_2) \in I_n^2} \mathbb{D} \left( H_{i_1, i_2} \left( X^{(1)}_{i_1}, X^{(2)}_{i_2} \right) \right) \in \mathbb{H}^{d_1 + d_2}, \]

and observing that $\| U_n \| = \| \mathbb{D} (U_n) \|$; we omit the general statements and only consider the special case of vector-valued U-statistics in Section 5.
4. Adamczak’s moment inequality for U-statistics.

Adamczak (2006) developed moment inequalities for general Banach space-valued completely degenerated U-statistics of arbitrary order. More specifically, application of Theorem 1 in (Adamczak, 2006) to our scenario $\mathcal{B} = (\mathbb{H}^d, \| \cdot \|)$ and $m = 2$ yields the following bounds for all $q \geq 1$ and $t \geq 2$:

\[
(\mathbb{E}\|U_n\|^{2q})^{1/(2q)} \leq C \left( \mathbb{E}\|U_n\| + \sqrt{q} \cdot A + q \cdot B + q^{3/2} \cdot \Gamma + q^2 \cdot D \right), \tag{12}
\]

\[
\Pr \left( \|U_n\| \geq C \left( \mathbb{E}\|U_n\| + \sqrt{t} \cdot A + t \cdot B + t^{3/2} \cdot \Gamma + t^2 \cdot D \right) \right) \leq e^{-t},
\]

where $C$ is an absolute constant, and the quantities $A, B, \Gamma, D$ will be specified below (see Section 6.3 for the complete statement of Adamczak’s result). Notice that inequality (12) 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 (12) and the results of Theorem 3.1 and its corollaries, such as inequality (9): while (12) describes the deviations of $\|U_n\|$ from its expectation, (9) states that $U_n$ is close to its expectation as a random matrix; similar connections exist between the Matrix Bernstein inequality (Tropp, 2012) and Talagrand concentration inequality (e.g., Boucheron, Lugosi and Massart, 2013). It particular, (12) can be combined with a bound (10) for $\mathbb{E}\|U_n\|$ to obtain a moment inequality that is superior (in the certain range of $q$) to the results derived from Theorem 3.1.

**Theorem 4.1.** Inequalities (12) hold with the following choice of $A, B, \Gamma$ and $D$:

\[
A = \sqrt{\log(de)} \left( \mathbb{E} \sum_{i=1}^{n} \mathbb{E}_2 \mathcal{G}_i \mathcal{G}_i^T \right)^{1/2} + \log(de) \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2:i_2 \neq i_1} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^{1/2},
\]

\[
B = \left( \sup_{z \in \mathcal{C}^d: \|z\|_2 \leq 1} \sum_{(i_1,i_2) \in I_n^d} \mathbb{E} \left( z^* H_{i_1,i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) z \right)^2 \right)^{1/2} \leq \left( \sum_{(i_1,i_2) \in I_n^d} \mathbb{E} H_{i_1,i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/2},
\]

\[
\Gamma = \sqrt{1 + \frac{\log d}{q} \left( \sum_{i_1} \mathbb{E} \max_{i_2:i_2 \neq i_1} \left\| H_{i_1,i_2}^{(1)} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)} + \left( 1 + \frac{\log d}{q} \right) \left( \sum_{i_1} \mathbb{E} \max_{i_2:i_2 \neq i_1} \left\| H_{i_1,i_2}^T \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)},
\]

where $\mathcal{G}_i$ were defined in (7).

**Proof.** See Section 6.3. □

It is possible to further simplify the bounds for $A$ (via Lemma 3.5) and $D$ to deduce that one
can choose
\[
A = \log(d) \left( \mathbb{E} \left\| \sum_{i=1}^{n} E_2 \tilde{G}_i \tilde{G}_i^T \right\| + \left\| \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{1/2} \right),
\]
\[
B = \left( \sup_{z \in \mathbb{C}^d : \|z\|_2 \leq 1} \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} \left( z^* H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) z \right)^{1/2} \right),
\]
\[
\Gamma = (\log(d))^{3/2} \left( \sum_{i_1} \mathbb{E} \left\| \sum_{i_2 : i_2 \neq i_1} \mathbb{E} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{1/2q} \right),
\]
\[
D = \log(d) \left( \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} \left\| H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^{(1/(2q))}.
\]

The upper bound for \( A \) can be modified even further as in (8), using the fact that
\[
\mathbb{E} \left\| \sum_{i=1}^{n} E_2 \tilde{G}_i \tilde{G}_i^T \right\| \leq \sum_{i_1} \mathbb{E} \left\| \sum_{i_2 : i_2 \neq i_1} E_2 H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|.
\]

5. Bounds for vector-valued U-statistics.

In this section, we show that obtained results easily adapt to the case of Hilbert space-valued U-statistics, that is, when \( H_{i_1, i_2} \in \mathbb{C}^d \) for all \( i_1, i_2 \); for brevity, we will write \( H_{i_1, i_2} \) for \( H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \). Note that the Hermitian dilation of \( H_{i_1, i_2} \) satisfies
\[
D^2 (H_{i_1, i_2}) = \left( \begin{array}{cc}
H_{i_1, i_2} H_{i_1, i_2}^* & 0 \\
0 & H_{i_1, i_2}^* H_{i_1, i_2}
\end{array} \right),
\]

hence \( \|D^2 (H_{i_1, i_2})\| = \|H_{i_1, i_2}\|_2^2 \) and
\[
\sum_{i_1} \mathbb{E} \left\| \sum_{i_2 : i_2 \neq i_1} E_2 D^2 (H_{i_1, i_2}) \right\| = \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} D^2 (H_{i_1, i_2}) = \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} \|H_{i_1, i_2}\|_2^2.
\]

Moreover, it is easy to see that
\[
\mathbb{E} \max_{i_1} \left\| \sum_{i_2 : i_2 \neq i_1} E_2 D^2 (H_{i_1, i_2}) \right\| = \mathbb{E} \max_{i_1} \left( \sum_{i_2 : i_2 \neq i_1} \mathbb{E} \|H_{i_1, i_2}\|_2^2 \right)^q
\]

and
\[
\sum_{i_1} \mathbb{E} \max_{i_2 : i_2 \neq i_1} \|D^2 (H_{i_1, i_2})\|^q = \sum_{i_1} \mathbb{E} \max_{i_2 : i_2 \neq i_1} \|H_{i_1, i_2}\|_2^{2q}.
\]

Theorem 3.1 and Lemma 3.5 imply that for all \( q \gg 1 \) and \( r = q \vee \log d \),
\[
\left( \mathbb{E} \|U_n\|_2^{2q} \right)^{1/2q} \leq C \sqrt{\frac{1 + \frac{\log d}{q} \left( \mathbb{E} \sum_{(i_1, i_2) \in I_n^2} \|H_{i_1, i_2}\|_2^2 \right)^{1/2}}{r}},
\]
\[
+ r^{3/2} \left( \mathbb{E} \max_{i_1} \left( \sum_{i_2 : i_2 \neq i_1} \mathbb{E} \|H_{i_1, i_2}\|_2^2 \right)^q \right)^{1/(2q)} + r^2 \left( \sum_{i_1} \mathbb{E} \max_{i_2 : i_2 \neq i_1} \|H_{i_1, i_2}\|_2^{2q} \right)^{1/2q}.
\]
If moreover $m, M$ are such that $\|H_{i_1,i_2}\|_2 \leq M$ and $\sqrt{\mathbb{E}_2 \|H_{i_1,i_2}\|^2_2} \leq m$ almost surely for all $i_1, i_2$, then Lemma 6.7 implies that for an absolute constant $C > 0$ and all $t \geq 1$,

$$
\Pr \left( \|U_n\|_2 \geq C \sqrt{1 + \log(d)/t} \left( \frac{\mathbb{E} \sum_{(i_1,i_2) \in I^2_n} \|H_{i_1,i_2}\|^2_2}{2} \right)^{1/2} (t + \log d) + m \sqrt{n}(t^{3/2} + \log^{3/2}(d)) + M n^{1/2t}(t^2 + \log^2 d) \right) \leq e^{-t}. \tag{14}
$$

To provide comparison and illustrate the improvements achievable via Theorem 4.1, observe that (13) and Lemma 6.7 imply (after some simple algebra) the following bound for all $t \geq 1$:

$$
\Pr \left( \|U_n\|_2 \geq C \left( \log(de) \left( \frac{\mathbb{E} \sum_{(i_1,i_2) \in I^2_n} \|H_{i_1,i_2}\|^2_2}{2} \right)^{1/2} \right) \sqrt{t} + \mathbb{E} \sum_{(i_1,i_2) \in I^2_n} H_{i_1,i_2}^* H_{i_1,i_2} \right)^{1/2} t + m \sqrt{n} \log^{3/2}(de) t^{3/2} + M \log(de) n^{1/2t} t^2 \right) \leq e^{-t},
$$

which is better than (14) for small values of $t$.

6. Proofs.

6.1. 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 6.1.** Let $M \in \mathbb{H}^{d_1+d_2}$ be 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

$$
\|M\| \leq \|A\| + \|B\|
$$

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

**Proof.** It follows from a result by Bourin and Lee (2012) that 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. \hfill \Box

The second result is the well-known decoupling inequality for U-statistics due to de la Pena and Montgomery-Smith (1995).

**Lemma 6.2.** Let $\{X_i\}_{i=1}^n$ be a sequence of independent random variables with values in a measurable space $(S, \mathcal{S})$ equipped with the probability measure $P$, 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 $(i_1, \ldots, i_m) \in I^m_n$, let $H_{i_1,...,i_m} : S^m \to B$ be a measurable function. Moreover, let $\Phi : [0, \infty) \to [0, \infty)$ be a convex nondecreasing function such that

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

Then

$$
\mathbb{E}\Phi(\frac{1}{n} \sum_{i_1,...,i_m} H_{i_1,...,i_m}(X_{i_1}, \ldots, X_{i_m})) \leq \mathbb{E}\Phi(\frac{1}{n} \sum_{i_1,...,i_m} H_{i_1,...,i_m}(X_{i_1}^{(k)}, \ldots, X_{i_m}^{(k)}))
$$

for every $k = 1, 2, \ldots, m$.
for all \((i_1, \ldots, i_m) \in I^n_m\). Then
\[
\mathbb{E} \left( \left\| \sum_{(i_1, \ldots, i_m) \in I^n_m} H_{i_1, \ldots, i_m}(X_{i_1}, \ldots, X_{i_m}) \right\| \right) \leq \mathbb{E} \left( C_m \left\| \sum_{(i_1, \ldots, i_m) \in I^n_m} H_{i_1, \ldots, i_m}(X_{i_1}^{(1)}, \ldots, X_{i_m}^{(m)}) \right\| \right),
\]
where \(C_m := 2^n(m^m - 1) \cdot ((m - 1)^m - 1) \cdots 3\). 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 \(t > 0\),
\[
\Pr \left( \left\| \sum_{(i_1, \ldots, i_m) \in I^n_m} H_{i_1, \ldots, i_m}(X_{i_1}, \ldots, X_{i_m}) \right\| \geq t \right) \leq D_m \Pr \left( D_m \left\| \sum_{(i_1, \ldots, i_m) \in I^n_m} H_{i_1, \ldots, i_m}(X_{i_1}^{(1)}, \ldots, X_{i_m}^{(m)}) \right\| \geq t \right).
\]
Furthermore, if \(H_{i_1, \ldots, i_m}\) 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 inequalities for the Rademacher sums and the Rademacher chaos with explicit constants, see Theorems 6.14, 6.22 in (Rauhut, 2012) and Corollary 20 in (Tropp, 2008).

**Lemma 6.3.** Let \(B_j \in \mathbb{C}^{r \times t}, j = 1, \ldots, n\) 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 \(p \geq 1\),
\[
\mathbb{E} \left( \left\| \sum_{j=1}^n \varepsilon_j B_j \right\|_{S_{2p}}^{2p} \right) \leq \left( 2\sqrt[2]{2} e \right)^p \cdot \max \left\{ \left\| \sum_{j=1}^n B_j B_j^* \right\|_{S_{2p}}^{1/2} \left\| \left( \sum_{j=1}^n B_j^* B_j \right) \right\|_{S_{2p}}^{1/2} \right\}.
\]

**Lemma 6.4.** Let \(\{A_{i_1,i_2}^{(k)}\}_{i_1,i_2=1}^n\) be a sequence of Hermitian matrices of the same dimension, and let \(\{\varepsilon_i^{(k)}\}_{i=1}^n, k = 1, 2,\) be i.i.d. Rademacher random variables. Then for any \(p \geq 1\),
\[
\mathbb{E} \left( \left\| \sum_{i_1=1}^n \sum_{i_2=1}^n A_{i_1,i_2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \right\|_{S_{2p}}^{2p} \right) \leq 2 \left( 2\sqrt[2]{2} e \right)^{2p} \cdot \max \left\{ \left\| G G^T \right\|_{S_{2p}}^{1/2} \left\| \sum_{i_1,i_2=1}^n A_{i_1,i_2}^2 \right\|_{S_{2p}}^{1/2} \right\},
\]
where the matrix \(G \in \mathbb{R}^{(nd) \times (nd)}\) is defined as
\[
G := \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}.
\]

The following result by Chen, Gittens and Tropp (2012) is a variant of matrix Rosenthal inequality for nonnegative definite matrices.

**Lemma 6.5.** Let \(Y_1, \ldots, Y_n\) be a sequence of independent nonnegative definite \(d \times d\) random matrices. Then for all \(q \geq 1\) and \(r = \max(q, \log(d))\),
\[
\left( \mathbb{E} \left\| \sum_{j} Y_j \right\|_{q}^{2q} \right)^{1/2q} \leq \sum_{j} \mathbb{E} Y_j^{1/2} + 2\sqrt{2er} \left( \mathbb{E} \max_{j} \|Y_j\|^{q} \right)^{1/2q}.
\]
Next inequality, due to de la Peña and Gine (1999), allows to replace the sum of moments of nonnegative random variables with maxima.

**Lemma 6.6.** Let $\xi_1, \ldots, \xi_n$ be independent random variables. Then for all $q > 1$ and $\alpha \geq 0$,

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

Finally, the following inequalities allow transitioning between moment and tail bounds; see Proposition 7.11 in (Foucart and Rauhut, 2013) and Lemma A.2 in (Dirksen et al., 2015).

**Lemma 6.7.** Let $X$ be a random variable satisfying $(\mathbb{E} |X|^p)^{1/p} \leq a_0 p^2 + a_1 p^{3/2} + a_2 p + a_3 \sqrt{p} + a_4$ for all $p \geq 2$ and some positive real numbers $a_j$, $j = 0, \ldots, 3$. Then for any $u \geq 2$,

$$\Pr \left( |X| \geq e(a_0 u^2 + a_1 u^{3/2} + a_2 u + a_3 \sqrt{u} + a_4) \right) \leq \exp \left( -u \right).$$

**Lemma 6.8.** Let $X$ be a random variable such that $\Pr (|X| \geq a_1 u + a_2 \sqrt{u}) \leq e^{-u}$ for all $u \geq 1$ and some $0 \leq a_1, a_2 < \infty$. Then

$$(\mathbb{E} |X|^p)^{1/p} \leq 2e^{1/(2e)}(\sqrt{2\pi e} c^{1/(12p)})^{1/p} e^{-1} a_1 p + 2(2e)^{-1/2} e^{1/(2e)}(\sqrt{\pi e} c^{1/(6p)})^{1/p} a_2 \sqrt{p}$$

for all $p \geq 1$.

Next, we will use Lemma 6.2 combined with a well-known argument to obtain the symmetrization inequality for degenerate U-statistics.

**Lemma 6.9.** Let $H_{i_1,i_2} : S \times S \rightarrow \mathbb{R}^d$ be degenerate kernels, $X_1, \ldots, X_n$ – independent $S$-valued random variables, and assume that $\{X_i^{(k)}\}_{i=1}^{n}$, $k = 1, 2$, are independent copies of this sequence. Moreover, let $\{\varepsilon_i^{(k)}\}_{i=1}^{n}$, $k = 1, 2$, be i.i.d. Rademacher random variables. Define

$$U_n' := \sum_{(i_1,i_2) \in \mathcal{I}_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}).$$

Then for any $p \geq 1$,

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

**Proof.** Note that

$$\mathbb{E} \|U_n\|^p = \mathbb{E} \left\| \sum_{(i_1,i_2) \in \mathcal{I}_n^2} H_{i_1,i_2}(X_{i_1}, X_{i_2}) \right\|^p \leq \mathbb{E} \left\| 2^2 \sum_{(i_1,i_2) \in \mathcal{I}_n^2} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^p,$$

where the inequality follows from the fact that $H_{i_1,i_2}$ is $\mathcal{P}$-canonical, hence Lemma 6.2 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 $\{X_{j}^{(i)}, \varepsilon_{j}^{(i)}\}_{j \geq 1}$ only (that is, conditionally on $\{X_{j}^{(k)}, \varepsilon_{j}^{(k)}\}_{j \geq 1}$, $k \neq i$). Using iterative expectations and the symmetrization
inequality for the Rademacher sums twice (see Lemma 6.3 in (Ledoux and Talagrand, 1991)),
we deduce that
\[ E\|U_n\|^p \leq 4^p E \left[ \sum_{(i_1,i_2) \in I_n^2} H_{i_1,i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right]^p \]
\[ = 4^p E \left[ E_1 \left\| \sum_{i_1=1}^{n} \sum_{i_2 \neq i_1} H_{i_1,i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^p \right] \]
\[ \leq 4^p E \left[ E_2 \left\| \sum_{i_2=1}^{n} \sum_{i_1 \neq i_2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1,i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^p \right] \]
\[ \leq 4^p E \left[ \left( \sum_{(i_1,i_2) \in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1,i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^p \right]. \]

6.2. Proofs of results in Section 3.

6.2.1. Proof of Lemma 3.3.

Recall that
\[ X = \sum_{i_1=1}^{n} \sum_{i_2 \neq i_1} A_{i_1,i_2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)}, \]
and let \( C_p := 2 \left( \frac{2\sqrt{e}}{p} \right)^{2p} \). We will first establish the upper bound. Application of Lemma 6.4 (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
\[ \left( E\|X\|_{S_{2p}}^{2p} \right)^{1/2} \leq 2^{1/2} e^{1/2} \cdot p \cdot \max \left\{ \left\| (G^T)_{1/2} \right\|_{S_{2p}}, \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\|_{S_{2p}}^{1/2} \right\} \right\}, \tag{16} \]

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 (16) for \( p = \infty \). To this end, we need to find an upper bound for
\[ \inf_{p \geq q} \frac{p \cdot \max \left\{ \left\| (G^T)_{1/2} \right\|_{S_{2p}}, \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\|_{S_{2p}}^{1/2} \right\} }{\max \left\{ \left\| (G^T)_{1/2} \right\|_{S_{2p}}, \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\|_{S_{2p}}^{1/2} \right\} } \].
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 (16) is attained by the first or second term. Define

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

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

$$B_{i_2} = \sum_{i_1 : i_1 \neq i_2} \hat{B}_{i_1,i_2}.$$  \hfill (17)

Then it is easy to see that

$$GG^T = \sum_{i_2} B_{i_2}^T B_{i_2},$$

$$\sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 = \sum_{i_2} B_{i_2} B_{i_2}^T.$$  \hfill (18)

The following bound gives a key estimate.

**Lemma 6.10.** Let $M_1, \ldots, M_N$ be a sequence of $d \times nd$ matrices. Let $\lambda_1, \ldots, \lambda_{nd}$ be eigenvalues of $\sum_j M_j^T M_j$ and let $\nu_1, \ldots, \nu_d$ be eigenvalues of $\sum_j M_j M_j^T$. Then $\sum_{i=1}^{nd} \lambda_i = \sum_{j=1}^{d} \nu_j$. Furthermore, if $\max \lambda_i \leq \frac{1}{d} \sum_{j=1}^{d} \nu_j$, then

$$\left\| \left( \sum_j M_j M_j^T \right)^{1/2} \right\|_{S_2p}^{2p} \leq \left\| \left( \sum_j M_j^T M_j \right)^{1/2} \right\|_{S_2p}^{2p},$$

for any integer $p \geq 2$.

The proof of the Lemma is given in Section 6.2.2. We will apply this fact with $M_j = B_j$, $j = 1, \ldots, n$. Assuming that $\max \lambda_i \leq \frac{1}{d} \sum_{j=1}^{d} \nu_j$, it is easy to see that the second term in the maximum in (16) dominates, hence

$$\mathbb{E}[X]_{S_2p}^{2p} \leq C_p \left\| \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right\|_{S_2p}^{2p} \leq C_p \mathrm{tr} \left( \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right)^p \leq C_p \cdot d \cdot \left( \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 \right)^p \leq C_p \cdot d \cdot \sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2,$$  \hfill (18)

where the last equality follows from the fact that for any positive semidefinite matrix $H$, we have $\|H^p\| = \|H\|^p$.

On the other hand, when $\max \lambda_i > \frac{1}{d} \sum_{j=1}^{d} \lambda_j$, it is easy to see that for all $p \geq 1$,

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

which in turn implies that

$$d \left( \max \lambda_i \right)^p \geq \sum_{j=1}^{nd} \lambda_j^p,$$  \hfill (19)
Moreover,
\[
\left\| \sum_{i_2} B_{i_2} B_{i_2}^{T} \right\|_{S_2}^{1/2} 2^{p} = \text{tr} \left( \left( \sum_{i_2} B_{i_2} B_{i_2}^{T} \right)^{p} \right) = \sum_{i=1}^{n} X_{i}^{p}. \tag{20}
\]

Combining (19), (20), we deduce that
\[
\left\| \sum_{i_2} B_{i_2} B_{i_2}^{T} \right\|_{S_2}^{2^{p}} \leq d \left\| \left( \sum_{i_2} B_{i_2} B_{i_2}^{T} \right)^{p} \right\| = d \left\| (GG)^{p} \right\| = d \left\| G G^{T} \right\|^{p},
\]
where the second from the last equality follows again from the fact that for any positive semi-definite matrix \(H\), we have \(\|H^{p}\| = \|H\|^{p}\). Thus, combining the bound above with (16) and (18), we obtain
\[
\mathbb{E} \|X\|_{S_2}^{2^{p}} \leq d \cdot C_{p} \max \left\{ \|G G^{T}\|^{p}, \left\| \sum_{(i_1,i_2) \in I_2} A_{i_1,i_2}^{2} \right\|^{p} \right\}.
\]

Finally, set \(p = \max(q, \log(d))\) and note that \(d^{1/2p} \leq \sqrt{e}\), hence
\[
\left( \mathbb{E} \|X\|^{2^{p}} \right)^{1/2q} \leq \left( \mathbb{E} \|X\|^{2^{p}} \right)^{1/2p} \leq \frac{4}{\sqrt{e}} \max\{\log d, q\} \cdot \max \left\{ \|G G^{T}\|^{p}, \left\| \sum_{(i_1,i_2) \in I_2} A_{i_1,i_2}^{2} \right\|^{p} \right\}^{1/2}.
\]
This finishes the proof of upper bound.

Now, we turn to the lower bound. Let \(\mathbb{E}_1[-]\) stand for the expectation with respect to \(\{\varepsilon_{i}^{(1)}\}_{i \geq 1}\) only. Then
\[
\left( \mathbb{E} \|X\|^{2^{p}} \right)^{1/(2p)} \geq \left( \mathbb{E} \|X\|^{2} \right)^{1/2} = \left( \mathbb{E} \mathbb{E}_1 \left\| \sum_{(i_1,i_2) \in I_2} A_{i_1,i_2}^{(1)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \right\|^{2} \right)^{1/2}
\]
\[
= \left( \mathbb{E} \mathbb{E}_1 \left\| \sum_{(i_1,i_2) \in I_2} A_{i_1,i_2}^{(1)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \right\|^{2} \right)^{1/2}
\]
\[
= \left( \mathbb{E} \left\| \sum_{i_2} \left( \sum_{i_1, i_2 \neq i_1} A_{i_1,i_2} \varepsilon_{i_2}^{(2)} \right)^{2} \right\|^{1/2} \right)
\]
It is easy to check that
\[
\sum_{i_1=1}^{n} \left( \sum_{i_2, i_2 \neq i_1} A_{i_1,i_2} \varepsilon_{i_2}^{(2)} \right)^{2} = \left( \sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right) \left( \sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right)^{T},
\]
where \(B_{i}\) were defined in (40). Hence
\[
\left( \mathbb{E} \|X\|^{2^{p}} \right)^{1/(2p)} \geq \left( \mathbb{E} \left\| \sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right\| \left( \sum_{i_2} \varepsilon_{i_2}^{(2)} B_{i_2} \right)^{T} \right)^{1/2}.
\]
Next, for any matrix \( A \in \mathbb{R}^{d_1 \times d_2} \),
\[
\left\| \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} A^T A & 0 \\ 0 & A A^T \end{pmatrix} \right\| = \max\{\|A^T A\|, \|A A^T\|\} = \|A A^T\|,
\]
where the last equality follows since \( \|A A^T\| = \|A^T A\| \). Taking \( A = \sum_{i_2} \varepsilon_{i_2}^2 B_{i_2} \) gives
\[
(\mathbb{E} \|X\|^{2p})^{1/(2p)} \geq \left( \mathbb{E} \left\| BB^T \right\|^2 \right)^{1/2} = \left( \mathbb{E} \left( \sum_{i_2} \varepsilon_{i_2}^2 \begin{pmatrix} 0 & B_{i_2}^T \\ B_{i_2} & 0 \end{pmatrix} \right)^2 \right)^{1/2} = \max \left\{ \sum_{i_2} B_{i_2}^T B_{i_2}, \sum_{i_2} A_{i_1,i_2}^2 \right\}^{1/2}.
\]

6.2.2. Proof of Lemma 6.10.

The equality of traces is obvious since
\[
\text{tr} \left( \sum_{j=1}^N M_j M_j^T \right) = \sum_{j=1}^N \text{tr} \left( M_j M_j^T \right) = \sum_{j=1}^N \text{tr} \left( M_j^T M_j \right) = \text{tr} \left( \sum_{j=1}^N M_j^T M_j \right).
\]

Set
\[
S := \sum_{i=1}^{nd} \lambda_i = \sum_{i=1}^d \nu_i.
\]

Note that
\[
\left( \sum_{j=1}^N M_j M_j^T \right)^{1/2} \overset{2p}{\leq} \text{tr} \left( \left( \sum_{j=1}^N M_j M_j^T \right)^p \right) = \sum_{i=1}^{nd} \lambda_i^p,
\]
\[
\left( \sum_{j=1}^N M_j M_j^T \right)^{1/2} \overset{2p}{=} \text{tr} \left( \left( \sum_{j=1}^N M_j M_j^T \right)^p \right) = \sum_{i=1}^d \nu_i^p.
\]

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
\[
\left( \sum_{j=1}^N M_j M_j^T \right)^{1/2} \overset{2p}{\leq} \max_{0 \leq \lambda_i \leq \frac{S}{d}, \sum_{i=1}^{nd} \lambda_i = S} \sum_{i=1}^{nd} \lambda_i^p,
\]
\[
\left( \sum_{j=1}^N M_j M_j^T \right)^{1/2} \overset{2p}{=} \min_{\nu_i \geq 0, \sum_{i=1}^d \nu_i = S} \sum_{i=1}^d \nu_i^p.
\]

Hence, it is enough to show that
\[
\max_{0 \leq \lambda_i \leq \frac{S}{d}, \sum_{i=1}^{nd} \lambda_i = S} \sum_{i=1}^{nd} \lambda_i^p \leq \min_{\nu_i \geq 0, \sum_{i=1}^d \nu_i = S} \sum_{i=1}^d \nu_i^p. \tag{21}
\]
The right hand side of the inequality (21) can be estimated via Jensen’s inequality as

\[
\min_{\nu_i \geq 0, \sum_{i=1}^{d} \nu_i = S} \min_{\frac{1}{d} \sum_{i=1}^{d} \nu_i = \hat{S}} \frac{1}{d} \sum_{i=1}^{d} \nu_i^p \geq \frac{1}{d} \sum_{i=1}^{d} \nu_i \left( \frac{1}{d} \sum_{i=1}^{d} \nu_i \right)^p = d \cdot \left( \frac{S}{d} \right)^p = \frac{S^p}{d^{p-1}}. \tag{22}
\]

Given positive integers \( K \geq d \) and \( K' \geq d' \), we will write \((K, d) > (K', d')\) if \( K \geq K' \), \( d \geq d' \) and at least one of the inequalities is strict. We will now prove by induction that for all \((K, d), K \geq d\) and any \( S > 0 \),

\[
\max_{\lambda_1, \ldots, \lambda_K \in \Lambda(K, d, S)} \sum_{i=1}^{K} \lambda_i^p = d \left( \frac{S}{d} \right)^p,
\]

where

\[
\Lambda(K, d, S) = \left\{ \lambda_1, \ldots, \lambda_K : 0 \leq \lambda_j \leq \frac{S}{d} \forall j, \sum_{i=1}^{K} \lambda_i = S \right\}.
\]

The inequality is obvious for all pairs \((K, d)\) with \( K = d \) or with \( d = 1 \). Fix \((K, d)\) with \( K \geq d > 1 \), and assume that the claim holds for all \((K', d')\) such that \( K' \geq d' \) and \((K, d) > (K', d')\).

It is easy to check that the only critical point of the Lagrangian

\[
F(\lambda_1, \ldots, \lambda_K, \tau) = \sum_{i=1}^{K} \lambda_i^p + \tau \left( \sum_{i=1}^{K} \lambda_i - S \right)
\]

in the relative interior of the set \( \Lambda(K, d, S) \) is \( \hat{\lambda}_1 = \ldots = \hat{\lambda}_K = S/K \) where the function achieves its minimum, hence the maximum is attained on the boundary of the set \( \Lambda(K, d, S) \). There are 2 possibilities:

1. Without loss of generality, \( \lambda_1 = 0 \). Then the situation is reduced to \((K', d') = (K - 1, d)\), whence we conclude that

\[
\max_{\lambda_1, \ldots, \lambda_{K-1} \in \Lambda(K-1, d, S)} \sum_{i=1}^{K-1} \lambda_i^p = d \left( \frac{S}{d} \right)^p.
\]

2. Without loss of generality, \( \lambda_1 = S/d \). Then the situation is reduced to \((K', d') = (K - 1, d - 1)\) for \( S' = S(d-1)/d \), hence

\[
(S/d)^p + \max_{\lambda_1, \ldots, \lambda_{K-1} \in \Lambda(K-1, d-1, S')} \sum_{i=1}^{K-1} \lambda_i^p = (S/d)^p + (d-1) \left( \frac{S'}{d-1} \right)^p = d(S/d)^p.
\]

6.2.3. Proof of Theorem 3.1.

By Lemma 6.9, we have

\[
\left( \mathbb{E} \left\| U_n \right\|^{2q} \right)^{1/2q} \leq 16 \left( \mathbb{E} \left\| U'_n \right\|^{2q} \right)^{1/2q}, \tag{23}
\]

where $U_n'$ was defined in (15). Applying Lemma 3.3 conditionally on $\{X^{(j)}_i\}_{i=1}^n$, $j = 1, 2$, we get

$$
\left( E \left\| U_n' \right\|^{2q} \right)^{1/2q} = \left( E \left\| \sum_{(i_1, i_2) \in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1, i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \right\|^{2q} \right)^{1/(2q)}
\leq 4e^{-1/2} \max(q, \log d) \left( E \max \left\{ \left\| \tilde{G}^T \right\|, \left\| \tilde{G}^T \tilde{G} \right\| \right\} \right)^{1/2q}, \quad (24)
$$

where $\tilde{G}$ was defined in (7). Let $\tilde{G}_i$ be the $i$-th column of $\tilde{G}$, then

$$
\tilde{G}^T \tilde{G}_i = \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T, \quad \sum_{(i_1, i_2) \in I_n^2} H^2(X^{(1)}_{i_1}, X^{(2)}_{i_2}) = \sum_{i=1}^n \tilde{G}_i^T \tilde{G}_i.
$$

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

$$
Q_i = \begin{pmatrix} 0 & \tilde{G}_i^T \\ \tilde{G}_i & 0 \end{pmatrix},
$$

so that

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

Inequality (24) implies that

$$
\left( E \left\| \sum_{(i_1, i_2) \in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H_{i_1, i_2}(X^{(1)}_{i_1}, X^{(2)}_{i_2}) \right\|^{2q} \right)^{1/(2q)} \leq 4e^{-1/2} \max(q, \log d) \left( E \sum_{i=1}^n Q_i^2 \right)^{1/2q}. \quad (25)
$$

Let $E_2[.]$ stand for the expectation with respect to $\{X^{(2)}_i\}_{i=1}^n$ only (that is, conditionally on $\{X^{(1)}_i\}_{i=1}^n$). Then Minkowski inequality followed by the symmetrization inequality imply that

$$
\left( E \sum_{i=1}^n Q_i^2 \right)^{1/(2q)} \leq \left( E \sum_{i=1}^n (Q_i^2 - E_2 Q_i^2) \right)^{1/(2q)} + \left( E \sum_{i=1}^n E_2 Q_i^2 \right)^{1/2q}
\leq \left( E \sum_{i=1}^n \varepsilon_i Q_i^2 \right)^{1/(2q)} + \left( E \sum_{i=1}^n E_2 Q_i^2 \right)^{1/2q}.
$$

Next, we obtain an upper bound for $\left( E \sum_{i=1}^n \varepsilon_i Q_i^2 \right)^{1/(2q)}$. To this end, we apply Khintchine inequality (Lemma 6.3). Denote $C_r := 2 \left( \frac{2 \pi^2}{r^2} \right)^2$, and let $E_\varepsilon[.]$ be the expectation with respect
to \{\varepsilon_i\}_{i=1}^n only. Then for \( r > q \) we deduce that

\[
\mathbb{E}_x \left\| \sum_{i=1}^n \varepsilon_i Q_i^2 \right\|_{S_2^r} \leq C_{r/2} \left\| \sum_{i=1}^n Q_i^4 \right\|_{S_2^r}^{1/2} \left( \sum_{i=1}^n Q_i^2 \right)^{2r} \leq C_{r/2} \text{tr} \left( \sum_{i=1}^n Q_i^2 \right)^r \leq C_{r/2} \max_i \left\{ Q_i^2 \right\}^r \left( \sum_{i=1}^n Q_i^2 \right)^{1/2} \left\| \sum_{i=1}^n Q_i^2 \right\|_{S_2^r}^{2r},
\]

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

\[
\left\| \sum_{i=1}^n Q_i^2 \right\|_{S_2^r}^{1/2} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)^r = \text{tr} \left( \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)^r \right) = \text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)^r.
\]

Applying Lemma 6.10 with \( M_j = \tilde{G}_j \), \( j = 1, \ldots, n \), we deduce that

- if \( \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\| \leq \frac{1}{\lambda} \text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right) \), then \( \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|_{S_2^r}^{1/2} \leq \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|_{S_2^r}^{2r} \),

which implies that \( \text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)^r \leq \sum_{i=1}^n \text{tr} \left( \tilde{G}_i \tilde{G}_i^T \right)^r \), and

\[
\left\| \sum_{i=1}^n Q_i^2 \right\|_{S_2^r}^{1/2} \leq 2d \cdot \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|^{2r}.
\]

- if \( \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\| > \frac{1}{\lambda} \text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right) \), let \( \lambda_j \) be the \( j \)-th eigenvalue of \( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \), and note that

\[
d > \frac{\text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)}{\text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)} = \text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right) = \sum_{i=1}^n \frac{\lambda_i}{\max_j \lambda_j} \geq \sum_{i=1}^n \frac{\lambda_i}{\max_j \lambda_j}^r,
\]

where \( r \geq 1 \). In turn, it implies that

\[
\text{tr} \left( \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right)^r < d \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|^{2r}.
\]

Thus

\[
\left\| \sum_{i=1}^n Q_i^2 \right\|_{S_2^r}^{1/2} \leq d \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|^{r} + \left\| \sum_{i=1}^n \tilde{G}_i \tilde{G}_i^T \right\|^{r}.
\]
Putting the bounds together, we obtain that
\[
E_{\varepsilon}\left\| \sum_{i=1}^{n} \varepsilon_i Q_i^2 \right\|_{S_{2r}}^{2r} \leq 2dC_r^{1/2} \max_i \left\| Q_i^2 \right\|^{r} \max \left\{ \left\| \sum_{i=1}^{n} \tilde{G}_i \tilde{G}_i^T \right\|^{r}, \left\| \sum_{i=1}^{n} \tilde{G}_i^T \tilde{G}_i \right\|^{r} \right\} \tag{27}
\]
\[
\leq 2dC_r^{1/2} \max_i \left\| Q_i^2 \right\|^{r} \left\| \sum_{i=1}^{n} Q_i^2 \right\|^{r} .
\]

Next, observe that for \( r \) such that \( 2r \geq q \), \( E_{\varepsilon} \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{q} \leq \left( E_{\varepsilon} \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{2r} \right)^{q/2r} \) by Hölder’s inequality, hence
\[
\left( E \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{q/2} \right)^{1/2q} \leq \left( E \left( E_{\varepsilon} \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{2r} \right) \right)^{1/2q} \leq \left( E \left( E_{\varepsilon} \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{2r} \right) \right)^{1/2q} \leq \left( 2dC_r^{1/2} \max \left\| Q_i^2 \right\|^{q/2} \left\| \sum_{i=1}^{n} Q_i^2 \right\|^{q/2} \right)^{1/2q} .
\]

Next, set \( r = q \vee \log d \) and apply Cauchy-Schwarz inequality to deduce that
\[
\left( E \left\| \sum_{j=1}^{n} \varepsilon_j Q_j^2 \right\|^{q/2} \right)^{1/2q} \leq (8r)^{1/4} \left( E \max_i \left\| Q_i^2 \right\|^{q/2} \right)^{1/4q} \left( E \left\| \sum_{i=1}^{n} Q_i^2 \right\|^{q} \right)^{1/(4q)} \tag{28}
\]

Substituting bound (28) into (26) and letting
\[
R_q := \left( E \left\| \sum_{i=1}^{n} Q_i^2 \right\|^{q} \right)^{1/(2q)},
\]
we obtain
\[
R_q \leq (8r)^{1/4} \sqrt{2R_q} \left( E \max_i \left\| Q_i^2 \right\|^{q/2} \right)^{1/(4q)} + \left( E \left\| \sum_{i=1}^{n} E_2 Q_i^2 \right\|^{q} \right)^{1/2q} .
\]

If \( x, a, b > 0 \) are such that \( x \leq a \sqrt{2} + b \), then \( x \leq 4a^2 \vee 2b \), hence
\[
R_q \leq 4\sqrt{2r} \left( E \max_i \left\| Q_i^2 \right\|^{q/2} \right)^{1/2q} + 2 \left( E \left\| \sum_{i=1}^{n} E_2 Q_i^2 \right\|^{q} \right)^{1/2q} .
\]
Finally, it follows from (25) that

$$
\left( \mathbb{E} \left\| \sum_{(i_1, i_2) \in I_n^2} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} H(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^q \right)^{1/(2q)} 
\leq 16 \sqrt{2} e^{-r/2} \left( \mathbb{E} \max_i \left\| Q_i \right\|^q \right)^{1/(2q)} + \frac{8}{\sqrt{e}} r \left( \mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 Q_i^2 \right\|^q \right)^{1/(2q)} 
\leq 16 \sqrt{2} e^{-r/2} \left( \mathbb{E} \max_i \left\| \tilde{G}_i^T \tilde{G}_i \right\|^q \right)^{1/(2q)} + \frac{8}{\sqrt{e}} r \left( \mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 \tilde{G}_i \tilde{G}_i^T \right\|^q \right)^{1/(2q)} 
= 16 \sqrt{2} e^{-r/2} \left( \mathbb{E} \max_i \left\| \sum_{i_2 : i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \right)^{1/(2q)} 
+ \frac{8}{\sqrt{e}} r \left( \mathbb{E} \left\| \sum_{i=1}^n \mathbb{E}_2 \tilde{G}_i \tilde{G}_i^T \right\|^q \right)^{1/(2q)} + \mathbb{E} \left\| \sum_{i=1}^n \left( \sum_{i_2 : i_2 \neq i_1} \mathbb{E}_2 H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right) \right\|^q \right)^{1/(2q)} 
$$

(29)

where the last equality follows from the definition of $\tilde{G}_i$. To bring the bound to its final form, we will apply Rosenthal’s inequality (Lemma 6.5) to the last term in (29) to get that

$$
\left( \mathbb{E} \left\| \sum_{i=1}^n \left( \sum_{i_2 : i_2 \neq i_1} \mathbb{E}_2 H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right) \right\|^q \right)^{1/2q} \leq \left\| \sum_{(i_1, i_2) \in I_n^2} \mathbb{E} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{1/2q} 
+ 2 \sqrt{2} e r \left( \mathbb{E} \max_i \left\| \sum_{i_2 : i_2 \neq i_1} \mathbb{E}_2 H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\| \right)^{1/2q}.
$$

Moreover, Jensen’s inequality implies that

$$
\mathbb{E} \max_i \left\| \sum_{i_2 : i_2 \neq i_1} \mathbb{E}_2 H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q \leq \mathbb{E} \max_i \left\| \sum_{i_2 : i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^q.
$$

hence this term can be combined with one of the terms in (29).

6.2.4. Proof of Lemma 3.4.

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 6.2 and
the desymmetrization inequality to get that for some absolute constant $C_0 > 0$

\[
\left( \mathbb{E} \left\| U_n \right\|^{2q} \right)^{1/(2q)} \geq \frac{1}{C_0} \left( \mathbb{E} \left\| \sum_{(i_1, i_2) \in I_n^2} H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{2q} \right)^{1/(2q)}
\]

\[
= \frac{1}{C_0} \left( \mathbb{E} \mathbb{E}_1 \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_2 \neq i_1} H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/(2q)}
\]

\[
\geq \frac{1}{2C_0} \left( \mathbb{E} \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_2 \neq i_1} H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/(2q)}
\]

\[
= \frac{1}{2C_0} \left( \mathbb{E} \mathbb{E}_2 \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_2 \neq i_1} \epsilon_{i_1} \epsilon_{i_2} H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/(2q)}
\]

\[
\geq \frac{1}{4C_0} \left( \mathbb{E} \sum_{(i_1, i_2) \in I_n^2} \epsilon_{i_1} \epsilon_{i_2} H_{i_1, i_2} \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/(2q)}.
\]

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

\[
\left( \mathbb{E} \left\| U_n \right\|^{2q} \right)^{1/(2q)} \geq \frac{1}{4C_0} \left( \mathbb{E} \max \left\{ \left\| \sum_i \tilde{G}_i \tilde{G}_i^T \right\|, \left\| \sum_i \tilde{G}_i \tilde{G}_i^T \right\| \right\} \right)^{1/(2q)}
\]

\[
\geq \frac{1}{4\sqrt{2C_0}} \left( \mathbb{E} \sum_i \tilde{G}_i \tilde{G}_i^T \right)^{1/(2q)} + \frac{1}{4\sqrt{2C_0}} \left( \mathbb{E} \sum_i \tilde{G}_i \tilde{G}_i^T \right)^{1/(2q)}
\]

\[
\geq \frac{1}{4\sqrt{2C_0}} \left( \mathbb{E} \sum_i \mathbb{E}_2 \tilde{G}_i \tilde{G}_i^T \right)^{1/(2q)} + \frac{1}{4\sqrt{2C_0}} \left( \mathbb{E} \sum_i \mathbb{E}_2 \tilde{G}_i \tilde{G}_i^T \right)^{1/(2q)},
\]

where \( \tilde{G}_i \) is the \( i \)-th column if the matrix \( \tilde{G} \) defined in (7). Let

\[
Q_i = \begin{pmatrix} 0 & \tilde{G}_i^T \\ \tilde{G}_i & 0 \end{pmatrix}.
\]

It follows from (30) that

\[
\left( \mathbb{E} \left\| U_n \right\|^{2q} \right)^{1/(2q)} \geq \left( \mathbb{E} \left\| \sum_{i=1}^n Q_i^2 \right\| \right)^{1/(2q)}.
\]

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

\[
\left( \mathbb{E} \left\| U_n \right\|^{2q} \right)^{1/(2q)} \geq \left( \mathbb{E} \sum_{i=1}^n Q_i^2 \right)^{1/(2q)}
\]

\[
\geq \left( \mathbb{E} \max_i \left\| Q_i^2 \right\| \right)^{1/(2q)} \geq \left( \mathbb{E} \max_i \left\| \tilde{G}_i \tilde{G}_i^T \right\| \right)^{1/(2q)},
\]

where the last equality holds since \( \left\| \tilde{G}_i^T \tilde{G}_i \right\| = \left\| \tilde{G}_i \tilde{G}_i^T \right\| \). The claim follows.
6.2.5. Proof of Lemma 3.5.

Note that

\[ r^{3/2} \left( \mathbb{E}_{i_1} \mathbb{E}_{i_2} \left[ \sum_{i_2, i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right] \right)^{q/2} \leq r^{3/2} \left( \mathbb{E}_{i_1} \mathbb{E}_{i_2} \left[ \sum_{i_2, i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right] \right)^{q/2} \]

Next, Lemma 6.5 implies that, for \( r = \max(q, \log(d)) \),

\[ \mathbb{E}_{i_1} \mathbb{E}_{i_2} \left[ \sum_{i_2, i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right]^{q} \leq 2^{2q-1} \left( \mathbb{E}_{i_2} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{q} + \left( 2 \sqrt{r} \right)^{2q} \mathbb{E}_{i_1} \mathbb{E}_{i_2} \left( H_{i_1, i_2} \right)^{2q}. \]

We will now apply Lemma 6.6 with \( \alpha = 1 \) and \( \xi_{i_1} := \sum_{i_2, i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \) to get that

\[ \sum_{i_1} \mathbb{E}_{i_1} \xi_{i_1} \leq 2(1 + q) \left( \mathbb{E}_{i_1} \sum_{i_2, i_2 \neq i_1} \mathbb{E}_{i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^q + \left( \mathbb{E} \max_{i_1} \mathbb{E}_{i_2} \mathbb{E}_{i_2} \left( H_{i_1, i_2} \right)^2 \right)^{q/2}. \]

Combining (31) with (32) and (33), we obtain

\[ r^{3/2} \left( \mathbb{E}_{i_1} \mathbb{E}_{i_2} \left[ \sum_{i_2, i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right] \right)^{q/2} \leq 2r \sqrt{2} \left( \mathbb{E} \max_{i_1} \mathbb{E}_{i_2} \mathbb{E}_{i_2} \left( H_{i_1, i_2} \right)^2 \right)^{q/2} \]

which yields the result.

6.3. Proof of Theorem 4.1.

Let \( J \subseteq I \subseteq \{1, 2\} \). We will write \( i \) to denote the multi-index \( (i_1, i_2) \in \{1, \ldots, n\}^2 \). We will also let \( i_I \) be the restriction of \( i \) onto its coordinates indexed by \( I \), and, for a fixed value of \( i_I \), let \( (H_i)_{i} \) be the array \( \{H_i, i \in \{1, \ldots, n\}^{|I|}\} \), where \( H_i := H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \). 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 of Adamczak (2006), we define

\[ \| (H_i)_{i} \|_{I, J} = \mathbb{E}_{i_J} \sup \left\{ \mathbb{E} \sum_{i_J} \langle \Phi, H_i \rangle \prod_{j \in J} f_{i_J}^{(j)}(X_{i_j}^{(j)}) : \| \Phi \|_* \leq 1, f_{i_J}^{(j)} : S \mapsto \mathbb{R} \text{ for all } i, j, \text{ and } \sum_i \mathbb{E} \left| f_{i_J}^{(j)}(X_{i_j}^{(j)}) \right|^2 \leq 1, j \in J \right\} \]

(35)
and \( \|H_1\|_{\mathcal{O},\mathcal{O}} := \|H_1\| \), where \( \langle A_1, A_2 \rangle := \text{tr}(A_1 A_2^\ast) \) for \( A_1, A_2 \in \mathbb{H}^d \) and \( \| \cdot \|_* \) denotes the nuclear norm. Theorem 1 in (Adamczak, 2006) states that for all \( q \geq 1 \),

\[
\left( \mathbb{E} \|U_n\|^{2q} \right)^{1/2q} \leq C \left[ \sum_{I \in \{1,2\}} \sum_{J \in I} q^{\lfloor |J|/2 + |I|} \left( \sum_{i,j} \mathbb{E} I_{ij} \|H_{ij}\|_{I,J}^{2q} \right)^{1/2q} \right],
\]

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

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

where

\[
A \leq 2 \mathbb{E}_1 \left( \sup_{\Phi: \|\Phi\|_* \leq 1} \sum_{i_2} \mathbb{E}_2 \left( \sum_{i_1} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \right)^{1/2},
\]

\[
B \leq \left( \sup_{\Phi: \|\Phi\|_* \leq 1} \sum_{(i_1,i_2) \in I_n^2} \mathbb{E} \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \right)^{1/2},
\]

\[
+ 2 \left( \sum_{i_2} \mathbb{E}_2 \left( \mathbb{E}_1 \left( \sum_{i_1} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^{2q} \right)^{1/2q} \right),
\]

\[
\Gamma \leq 2 \left( \sum_{i_2} \mathbb{E}_2 \left( \sup_{\Phi: \|\Phi\|_* \leq 1} \sum_{i_1} \mathbb{E}_1 \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \right)^{1/2q} \right),
\]

\[
D \leq \left( \sum_{(i_1,i_2) \in I_n^2} \mathbb{E} \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{2q} \right)^{1/2q}.
\]

The bounds for \( A, B, \Gamma \) above are obtained from (35) via the Cauchy-Schwarz inequality. Note that \( \Phi \mapsto \mathbb{E} \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \) is a convex function, hence its maximum over the convex set \( \{ \Phi \in \mathbb{H}^d : \|\Phi\|_* \leq 1 \} \) 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 \( \phi \phi^\ast \) for some \( \phi \in \mathbb{C}^d \). It implies that

\[
\sup_{\Phi: \|\Phi\|_* \leq 1} \mathbb{E} \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \leq \sup_{\|\phi\|_{\mathbb{C}^d} \leq 1} \mathbb{E} \left( H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \phi \phi^\ast) \right)^2 \leq \mathbb{E} H_{i_1,i_2}^2(X_{i_1}^{(1)}, X_{i_2}^{(2)}) .
\]

(36)

Moreover,

\[
\sum_{i_2} \mathbb{E}_2 \left( \mathbb{E}_1 \left( \sum_{i_1} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}, \Phi) \right)^2 \right)^{1/2q} = \sum_{i_2} \mathbb{E}_2 \left( \mathbb{E}_1 \left( \sum_{i_1} H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{2q} \right)^{1/2q}
\]

\[
\leq \sum_{i_2} \mathbb{E}_2 \left( \sum_{i_1} \left\| H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^{2q} \right)
\]

\[
\leq \sum_{i_2} \mathbb{E}_2 \left( 2^{2q} \mathbb{E}_1 \left( \sum_{i_1} H_{i_1,i_2}^2(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2} + 4 \mathbb{E}_2 \left( \mathbb{E}_1 \max_{i_1} \left\| H_{i_1,i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^{2q} \right)^{1/2q} \right)^{2q}.
\]

(37)
where we have used Lemma 3.1 in the last step, and \( r = q \lor \log d \). Combining (36), (37), we get that

\[
B \leq \left( \sum_{(i_1, i_2) \in F_d^2} \mathbb{E} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2} + 4\sqrt{e} r \left( \sum_{i_1} \mathbb{E} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2q}
+ 8\sqrt{2e} r \left( \sum_{i_1} \mathbb{E} \max_{i_1} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2q}.
\]

(38)

It is also easy to get the following bound for \( \Gamma \): since

\[
\sup_{\Phi : \|\Phi\|_{\alpha} \leq 1} \mathbb{E}_1 \left\langle H \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right), \Phi \right\rangle^2 \leq \left\| \mathbb{E}_1 H^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|
\]

we deduce that

\[
\Gamma \leq 2 \left( \sum_{i_2} \mathbb{E}_2 \left( \sup_{\Phi : \|\Phi\|_{\alpha} \leq 1} \sum_{i_1} \mathbb{E}_1 \left\langle H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}), \Phi \right\rangle^2 \right) \right)^{1/2q}
\leq 2 \left( \sum_{i_2} \mathbb{E}_2 \left\| \sum_{i_1} \mathbb{E}_1 H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^{1/2q} \right) \leq 2 \left( \sum_{i_2} \mathbb{E}_2 \sum_{i_1} \mathbb{E}_1 H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2q}.
\]

(39)

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

**Lemma 6.11.** The following inequality holds:

\[
A \leq 2 \mathbb{E}_1 \left( \sup_{\Phi : \|\Phi\|_{\alpha} \leq 1} \sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}), \Phi \right\rangle^2 \right)^{1/2}
\leq 64\sqrt{e} \log(de) \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2 : i_2 \neq i_1} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\|^{1/2} \right)
+ 8\sqrt{2e} \log(de) \left( \mathbb{E} \left| \sum_i \mathbb{E}_2 \tilde{G}_i G_i^T \right| \right) + \sum_{(i_1, i_2) \in F_d^2} \mathbb{E} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^{1/2},
\]

where \( \tilde{G} \) was defined in (7).

Combining the bounds (38), (39) and Lemma 6.11, and grouping the terms with the same power of \( q \), we get the result of Theorem 4.1.

It remains to prove Lemma 6.11. To this end, note that Jensen’s inequality and an argument similar to (36) imply that

\[
\mathbb{E}_1 \left( \sup_{\Phi : \|\Phi\|_{\alpha} \leq 1} \sum_{i_2} \mathbb{E}_2 \left\langle \sum_{i_1} H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}), \Phi \right\rangle^2 \right)^{1/2}
\leq \left( \mathbb{E} \sup_{\Phi : \|\Phi\|_{\alpha} \leq 1} \left\langle \sum_{i_1} H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}), \Phi \right\rangle^2 \right)^{1/2} \leq \left( \mathbb{E} \left\| \sum_{i_1} \sum_{i_2} H^2_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right\| \right)^{1/2}.
\]
Next, arguing as in the proof of Lemma 3.3, we define

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

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

\[ B_{i_2} = \sum_{i_1 : i_1 \neq i_2} \hat{B}_{i_1, i_2}. \tag{40} \]

Using the representation (40), we have

\[
\left( \mathbb{E} \left\| \sum_{i_2} \left( \sum_{i_1} H_{i_1, i_2}(X_{i_1}^{(1)}, X_{i_2}^{(2)}) \right)^2 \right\|^{1/2} \right)^2 = \left( \mathbb{E} \left\| \left( \sum_{i_2} B_{i_2} \right) \left( \sum_{i_2} B_{i_2}^T \right)^T \right\|^{1/2} \right)^2 \\
= \left( \mathbb{E} \left\| \sum_{i_2} \left( \begin{array}{cc} 0 & B_{i_2}^T \\ B_{i_2} & 0 \end{array} \right) \right\|^{1/2} \right)^2 \leq 2 \left( \mathbb{E} \left\| \sum_{i_2} \varepsilon_{i_2} \left( \begin{array}{cc} 0 & B_{i_2}^T \\ B_{i_2} & 0 \end{array} \right) \right\|^{1/2} \right)^2,
\]

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 (4) yields that

\[ A \leq 4 \sqrt{c(1 + 2 \log d)} \left( \mathbb{E} \left\| \sum_{i_2} \left( \begin{array}{cc} B_{i_2}^T & 0 \\ 0 & B_{i_2} B_{i_2}^T \end{array} \right) \right\|^{1/2} \right)^{1/2} \\
= 4 \sqrt{c(1 + 2 \log d)} \left( \mathbb{E} \max \left\{ \left\| \sum_{i_2} B_{i_2} B_{i_2} \right\|, \left\| \sum_{i_2} B_{i_2} B_{i_2}^T \right\| \right\} \right)^{1/2} \\
= 4 \sqrt{c(1 + 2 \log d)} \left( \mathbb{E} \max \left\{ \left\| \tilde{G} \tilde{G}^T \right\|, \sum_{(i_1, i_2) \in \mathcal{I}_n^2} \left( \sum_{i_1, i_2} \right) H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\} \right)^{1/2}.
\]

Note that the last expression is of the same form as equation (24) in the proof of Theorem 3.1 with \( q = 1 \). Repeating the same argument, one can show that

\[ A \leq 64 \sqrt{c \log(d)} \left( \mathbb{E} \max_{i_1} \left\| \sum_{i_2 \neq i_1} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right\|^{1/2} \right)^{1/2} \\
+ 8 \sqrt{c \log(d)} \left( \mathbb{E} \sum_{i} \left\| \tilde{G} \tilde{G}^T \right\| + \sum_{(i_1, i_2) \in \mathcal{I}_n^2} \mathbb{E} H_{i_1, i_2}^2 \left( X_{i_1}^{(1)}, X_{i_2}^{(2)} \right) \right)^{1/2},
\]

which is an analogue of (29).

6.4. Calculations related to Examples 1 and 2.

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

\[ (GG^T)_{ii} = \sum_{j \neq i} A_{i,j}^2 = \sum_{j \neq i} \left( a_i a_j^T + a_j a_i^T \right)^2 = (n - 1) a_i a_i^T + \sum_{j \neq i} a_j a_j^T. \]

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

\[ (GG^T)_{ij} = \sum_{k \neq i,j} A_{i,k} A_{j,k} = \sum_{k \neq i,j} \left( a_i a_k^T + a_k a_i^T \right) \left( a_j a_k^T + a_k a_j^T \right) = (n - 2) a_i a_j^T. \]
We thus obtain that
\[ GG^T = (n - 2)aa^T + \text{Diag} \left( \sum_{j=1}^{n} a_j a_j^T, \ldots, \sum_{j=1}^{n} a_j a_j^T \right), \]
where \( \text{Diag}(\cdot) \) denotes the block-diagonal matrix with diagonal blocks in the brackets. Since
\[ \text{Diag} \left( \sum_{j=1}^{n} a_j a_j^T, \ldots, \sum_{j=1}^{n} a_j a_j^T \right) \geq 0, \]
it follows that
\[ \|GG^T\| \geq (n - 2)\|a\|^2 = (n - 2)n. \]
On the other hand,
\[
\sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 = \sum_{(i_1,i_2) \in I_n^2} \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right)^2 = \sum_{(i_1,i_2) \in I_n^2} \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right) = 2(n - 1) \left\| a_i a_i^T \right\| = 2(n - 1).
\]
where the last equality follows from the fact that \( \{a_1, \ldots, a_n\} \) are orthonormal.

For Example 2, we similarly obtain that
\[
(GG)^{ii} = \sum_{j:j \neq i} c_{i,j}^2 \left( a_i a_j^T + a_j a_i^T \right)^2 = \left( \sum_{j:j \neq i} c_{i,j}^2 \right) a_i a_i^T + \sum_{j:j \neq i} c_{i,j}^2 a_j a_j^T = a_i a_i^T + \sum_{j:j \neq i} c_{i,j}^2 a_j a_j^T,
\]
\[
(GG)^{ij} = \sum_{k \neq i,j} c_{i,k} c_{j,k} \left( a_i a_k^T + a_k a_i^T \right) \left( a_j a_k^T + a_k a_j^T \right) = 0, \quad i \neq j,
\]
hence \( \|GG^T\| = \max_i \left\| (GG^T)^{ii} \right\| = 1. \) On the other hand,
\[
\sum_{(i_1,i_2) \in I_n^2} A_{i_1,i_2}^2 = \sum_{(i_1,i_2) \in I_n^2} \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right)^2 \]
\[
= \sum_{(i_1,i_2) \in I_n^2} c_{i_1,i_2}^2 \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right) = 2 \left\| a_i a_i^T \right\|^2 = 2,
\]
and
\[
\sum_{i_1} \sum_{i_2:i_2 \neq i_1} A_{i_1,i_2}^2 = \sum_{i_1} \sum_{i_2 \neq i_1} c_{i_1,i_2}^2 \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right)^2 \]
\[
= \sum_{i_1} \sum_{i_2:i_2 \neq i_1} c_{i_1,i_2}^2 \left( a_{i_1} a_{i_2}^T + a_{i_2} a_{i_1}^T \right) = \sum_{i_1} a_{i_1} a_{i_1}^T + \sum_{i_2:i_2 \neq i_1} c_{i_1,i_2}^2 a_{i_2} a_{i_2}^T = n.
\]
References

ADAMCZAK, R. (2006). Moment inequalities for U-statistics. The Annals of Probability 34 2288–2314.

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). Concentration inequalities: A nonasymptotic theory of independence. Oxford university press.

BOURIN, J.-C. and LEE, E.-Y. (2012). Unitary orbits of Hermitian operators with convex or concave functions. Bulletin of the London Mathematical Society 44 1085–1102.

CHEN, X. (2016). On bootstrap approximations for high-dimensional U-statistics and random quadratic forms. arXiv preprint arXiv:1610.00032.

CHEN, R. Y., GITTENS, A. and TROPP, J. A. (2012). The masked sample covariance estimator: an analysis using matrix concentration inequalities. Information and Inference.

DE LA PENA, V. and GINE, E. (1999). Decoupling: From dependence to independence. Springer-Verlag, New York.

DE LA PENA, V. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. Annals of Probability 23 806-816.

DIRKSEN, S. et al. (2015). Tail bounds via generic chaining. Electronic Journal of Probability 20.

FOUCART, S. and RAUHUT, H. (2013). A mathematical introduction to compressive sensing 1. Birkhäuser Basel.

GINE, E., LATALA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. High Dimensional Probability II 13–38.

GINE, E. and ZINN, J. (1992). On Hoffmann-Jørgensen’s inequality for U-processes. In Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference 80–91. Springer.

GINE, E., KWAPIEN, S., LATALA, R. and ZINN, J. (2001). The LIL for canonical U-statistics of order 2. Annals of probability 520–557.

HAN, F. and LIU, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability 23 23.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics 293–325.

HOUDRÉ, C. and REYNAUD-BOURET, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In Stochastic inequalities and applications 55–69. Springer.

IBRAMIMOV, R. and SHARAKHMETOV, S. (1999). Analogues of Khintchine, Marcinkiewicz–Zygmund and Rosenthal inequalities for symmetric statistics. Scandinavian journal of statistics 26 621–633.

JUNGE, M., ZENG, Q. et al. (2013). Noncommutative Bennett and Rosenthal inequalities. The Annals of Probability 41 4287–4316.

KOROLYUK, V. S. and BOROVSKICH, Y. V. (2013). Theory of U-statistics 273. Springer Science & Business Media.

KOWALSKI, J. and TU, X. M. (2008). Modern applied U-statistics 714. John Wiley & Sons.

LEDoux, M. and TALAGRAND, M. (1991). Probability in Banach Spaces: isoperimetry and processes. Springer-Verlag, Berlin.

MACKEY, L., JORDAN, M. I., CHEN, R. Y., FARRELL, B., TROPP, J. A. et al. (2014). Matrix concentration inequalities via the method of exchangeable pairs. The Annals of Probability 42 906–945.

MINSKER, S. and WEI, X. (2018). Robust modifications of U-statistics and applications to covariance estimation problems. arXiv preprint arXiv:1801.05565.

PISIER, G. (1998). Non-commutative vector-valued $L_p$ spaces and completely $p$-summing maps.
Rauhut, H. (2009). Circulant and Toeplitz matrices in compressed sensing. *arXiv preprint arXiv:0902.4394*.

Rauhut, H. (2012). Compressive sensing and structured random matrices. *Radon Series Comp. Appl. Math* 1-94.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics* 162. John Wiley & Sons.

Tropp, J. (2008). On the conditioning of random subdictionaries. *Applied and Computational Harmonic Analysis* 25 1–24.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics* 12 389–434.

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

Tropp, J. A. (2016a). The expected norm of a sum of independent random matrices: An elementary approach. In *High Dimensional Probability VII* 173–202. Springer.

Tropp, J. A. (2016b). Second-order matrix concentration inequalities. *Applied and Computational Harmonic Analysis*.

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

Wegkamp, M. and Zhao, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* 22 1184–1226.