High-Dimensional Central Limit Theorems for Homogeneous Sums

Yuta Koike[1]

Abstract

This paper develops a quantitative version of de Jong's central limit theorem for homogeneous sums in a high-dimensional setting. More precisely, under appropriate moment assumptions, we establish an upper bound for the Kolmogorov distance between a multi-dimensional vector of homogeneous sums and a Gaussian vector so that the bound depends polynomially on the logarithm of the dimension and is governed by the fourth cumulants and the maximal influences of the components. As a corollary, we obtain high-dimensional versions of fourth-moment theorems, universality results and Peccati–Tudor-type theorems for homogeneous sums. We also sharpen some existing (quantitative) central limit theorems by applications of our result.

1 Introduction

Let $X = (X_i)_{i=1}^{\infty}$ be a sequence of independent centered random variables with unit variance. A *homogeneous sum* is a random variable of the form

$$Q(f; X) = \sum_{i_1, \ldots, i_p=1}^{N} f(i_1, \ldots, i_q) X_{i_1} \cdots X_{i_q},$$

✉ Yuta Koike
  kyuta@ms.u-tokyo.ac.jp

[1] University of Tokyo and CREST JST, Tokyo, Japan

where $N, q \in \mathbb{N}$, $[N] := \{1, \ldots, N\}$ and $f : [N]^q \to \mathbb{R}$ is a symmetric function vanishing on diagonals, i.e., $f(i_1, \ldots, i_q) = 0$ unless $i_1, \ldots, i_q$ are mutually different. Studies of limit theorems for a sequence of homogeneous sums have some history in probability theory. Rotar' [53,54] investigated invariance principles for $Q(f; X)$ regarding the law of X. In the notable work of de Jong [22], the following striking result has been established: For every $n \in \mathbb{N}$, let $f_n : [N_n]^q \to \mathbb{R}$ be a symmetric function vanishing on diagonals with q fixed and $N_n \uparrow \infty$ as $n \to \infty$. Assume $E[X_i^4] < \infty$ for all i and $E[Q(f_n; X)^2] = 1$ for all n. Then, $Q(f_n; X)$ converges in law to the standard normal distribution, provided that the following two conditions hold true:

(i) $E[Q(f_n; X)^4] \to 3$ as $n \to \infty$.
(ii) $\max_{1 \leq i \leq N_n} \mathrm{Inf}_i(f_n) \to 0$ as $n \to \infty$, where $\mathrm{Inf}_i(f_n)$ is defined by

$$\mathrm{Inf}_i(f_n) := \sum_{i_2, \ldots, i_q = 1}^{N_n} f_n(i, i_2, \ldots, i_q)^2 \qquad (1.1)$$

and called the *influence* of the ith variable of f_n.

When $q = 1$, condition (ii) says that $\max_{1 \leq i \leq N_n} f_n(i)^2 \to 0$ as $n \to \infty$, which is equivalent to the celebrated *Lindeberg condition*. In this case condition (i) is always implied by (ii), and thus it is an extra one. In contrast, when $q \geq 2$, condition (ii) is no longer sufficient for the asymptotic normality of the sequence $(Q(f_n; X))_{n=1}^{\infty}$, so one needs an additional condition. The motivation of introducing condition (i) in [22] was that one can easily check condition (i) is equivalent to the asymptotic normality of $(Q(f_n; X))_{n=1}^{\infty}$ when $q = 2$ and X is Gaussian (see also [21]). Later on, this observation was significantly improved in the influential paper by Nualart & Peccati [49]: For any q, the asymptotic normality of $(Q(f_n; X))_{n=1}^{\infty}$ is implied just by condition (i) as long as X is Gaussian. Results of this type are nowadays called *fourth-moment theorems* and have been extensively studied in the past decade. In particular, further investigation of the fourth-moment theorem in [49] has led to the introduction of the so-called *Malliavin–Stein method* by Nourdin & Peccati [43], which have produced one of the most active research areas in the recent probabilistic literature. We refer the reader to the monograph [44] for an introduction to this subject and the survey [3] for recent developments.

Implication of the Malliavin–Stein method to de Jong's central limit theorem (CLT) for homogeneous sums has been investigated in the seminal work of Nourdin, Peccati & Reinert [47], where several important extensions of de Jong's result have been developed. The following three results are particularly relevant to our work:

(I) First, they have established a multi-dimensional extension of de Jong's CLT which shows multi-dimensional vectors of homogeneous sums enjoy a CLT if de Jong's criterion is satisfied component-wise. More precisely, let $d \in \mathbb{N}$ and, for every $j = 1, \ldots, d$, let $q_j \in \mathbb{N}$ and $f_{n,j} : [N_n]^{q_j} \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Also, let $\mathfrak{C} = (\mathfrak{C}_{jk})_{1 \leq j, k \leq d}$ be a $d \times d$ positive

semidefinite symmetric matrix and suppose that

$$\max_{1 \leq j,k \leq d} |E[Q(f_{n,j}; X)Q(f_{n,k}; X)] - \mathfrak{C}_{jk}| \to 0$$

as $n \to \infty$. Then, the d-dimensional random vector

$$\boldsymbol{Q}^{(n)}(\boldsymbol{X}) := (Q(f_{n,1}; \boldsymbol{X}), \ldots, Q(f_{n,d}; \boldsymbol{X}))$$

converges in law to the d-dimensional normal distribution $\mathcal{N}_d(0, \mathfrak{C})$ with mean 0 and covariance matrix $\mathfrak{C}$ as $n \to \infty$ if $E[Q(f_{n,j}; X)^4] - 3E[Q(f_{n,j}; X)^2]^2 \to 0$ and $\max_{1 \leq i \leq N_n} \mathrm{Inf}_i(f_{n,j}) \to 0$ as $n \to \infty$ for every $j = 1, \ldots, d$.

(II) Second, they have found the following *universality* of Gaussian variables in the context of homogeneous sums ([47, Theorem 1.2]): Assume

$$\sup_n \sum_{i_1,\ldots,i_q=1}^{N_n} f_{n,j}(i_1, i_2, \ldots, i_{q_j})^2 < \infty$$

and $\mathfrak{C}_{jj} > 0$ for every j. Then, if $\boldsymbol{Q}^{(n)}(\boldsymbol{G})$ converges in law to $\mathcal{N}_d(0, \mathfrak{C})$ as $n \to \infty$ for a sequence of standard Gaussian variables $\boldsymbol{G} = (G_i)_{i=1}^{\infty}$, then $\boldsymbol{Q}^{(n)}(\boldsymbol{X})$ converges in law to $\mathcal{N}_d(0, \mathfrak{C})$ as $n \to \infty$ for any sequence $\boldsymbol{X} = (X_i)_{i=1}^{\infty}$ of independent centered random variables with unit variance and such that $\sup_i E[|X_i|^3] < \infty$.

(III) Third, they have established some quantitative versions of de Jong's CLT for homogeneous sums; see Proposition 5.4 and Corollary 7.3 in [47] for details (see also Sect. 2.1.1).

We remark that these results have been generalized in various directions by subsequent studies. For example, the universality results analogous to (II) have also been established for Poisson variables in Peccati & Zheng [52] and i.i.d. variables with zero skewness and nonnegative excess kurtosis in Nourdin et al. [45,46], respectively. Also, the recent work of Döbler & Peccati [28] has extended (I) and (II) to more general degenerate U-statistics which were originally treated in [22].

As the title of the paper suggests, the aim of this paper is to extend the above results to a high-dimensional setting where the dimension d depends on n and $d = d_n \to \infty$ as $n \to \infty$. Of course, in such a setting, the "asymptotic distribution" $\mathcal{N}_d(0, \mathfrak{C})$ also depends on n and, even worse, it is typically no longer tight. Therefore, we need to properly reformulate the above statements in this setting. In this paper, we adopt the so-called *metric approach* to accomplish this purpose: We try to establish the convergence of some metric between the laws of $\boldsymbol{Q}^{(n)}(\boldsymbol{X})$ and $\mathcal{N}_d(0, \mathfrak{C})$. Specifically, we take the *Kolmogorov distance* as the metric between the probability laws. Namely, letting $Z^{(n)}$ be a d_n-dimensional centered Gaussian vector with covariance matrix $\mathfrak{C}_n$ for each n, we aim at proving the following convergence:

$$\sup_{x \in \mathbb{R}^{d_n}} |P(\boldsymbol{Q}^{(n)}(\boldsymbol{X}) \leq x) - P(Z^{(n)} \leq x)| \to 0 \quad \text{as } n \to \infty.$$

Here, for vectors $x = (x_1, \ldots, x_{d_n}) \in \mathbb{R}^{d_n}$ and $y = (y_1, \ldots, y_{d_n}) \in \mathbb{R}^{d_n}$, we write $x \leq y$ to express $x_j \leq y_j$ for every $j = 1, \ldots, d_n$. In addition, we are particularly interested in a situation where the dimension $d = d_n$ increases extremely faster than the "standard" convergence rate of Gaussian approximation for a sequence of univariate homogeneous sums. Given that both $\sqrt{|E[Q(f_n; X)^4] - 3E[Q(f_n; X)^2]^2|}$ and $\max_{1 \leq i \leq N_n} \sqrt{\text{Inf}_i(f_n)}$ can be the optimal convergence rates of the Gaussian approximation of $Q(f_n; X)$ in the Kolmogorov distance (see [42, Proposition 3.8] for the former and [30, Remark 1] for the latter), we might consider the quantity

$$\delta_n := \max_{1 \leq j \leq d_n} \sqrt{|E[Q(f_{n,j}; X)^4] - 3E[Q(f_{n,j}; X)^2]^2| + \max_{1 \leq i \leq N_n} \text{Inf}_i(f_{n,j})}$$

as an appropriate definition of the "standard" convergence rate. Then, we aim at proving

$$\sup_{x \in \mathbb{R}^{d_n}} |P(\boldsymbol{Q}^{(n)}(\boldsymbol{X}) \leq x) - P(Z^{(n)} \leq x)| \leq C (\log d_n)^a \delta_n^b \qquad (1.2)$$

for all $n \in \mathbb{N}$, where $a, b, C > 0$ are constants which do not depend on n (here and below we assume $d_n \geq 2$). As a byproduct, results of this type enable us to extend fourth-moment theorems and universality results for homogeneous sums to a high-dimensional setting (see Theorem 2.2 for the precise statement).

Our formulation of a high-dimensional extension of CLTs for homogeneous sums is motivated by the recent path-breaking work of Chernozhukov, Chetverikov & Kato [13,18], where results analogous to (1.2) have been established for sums of independent random vectors. More formally, let $(\xi_{n,i})_{i=1}^n$ be a sequence of independent centered d_n-dimensional random vectors. Set $S_n := n^{-1/2} \sum_{i=1}^n \xi_{n,i}$ and assume $\mathfrak{C}_n = E[S_n S_n^\top]$ ($\top$ denotes the transpose of a matrix). Then, under an appropriate assumption on moments, we have

$$\sup_{x \in \mathbb{R}^{d_n}} |P(S_n \leq x) - P(Z^{(n)} \leq x)| \leq C' \left(\frac{\log^7(d_n n)}{n} \right)^{1/6}, \qquad (1.3)$$

where $C' > 0$ is a constant which does not depend on n (see Proposition 2.1 in [18] for the precise statement). Here, we shall remark that the bound in (1.3) depends on n through $n^{-1/6}$, which is suboptimal when the dimension d_n is fixed. However, in [18, Remark 2.1(ii)] it is conjectured that the rate $n^{-1/6}$ is nearly optimal in a minimax sense when d_n is extremely larger than n (see also [10, Remark 1]). This conjecture is motivated by the fact that the rate $n^{-1/6}$ is minimax optimal in CLTs for sums of independent random variables taking values in an infinite-dimensional Banach space (see, e.g., [8, Theorem 2.6]). Given that high-dimensional CLTs of type (1.3) are closely related to Gaussian approximation of the suprema of empirical processes (see, e.g., [15,17]), it would be worth mentioning that a duality argument enables us to translate the minimax rate for CLTs in a Banach space to the one for Gaussian approximation of the suprema of empirical processes with a specific class of functions in the Kolmogorov distance; see [50] for details. For this reason, we also conjecture

that $b = 1/3$ would give an optimal dependence on δ_n of the bound in (1.2) (note that the rate $n^{-1/2}$ is the standard convergence rate of CLTs for sums of independent one-dimensional random variables). In this paper, we indeed establish that the bound of type (1.2) holds true with $b = 1/3$ under a moment assumption on X when q_j's do not depend on j (see Theorem 2.1 and Remark 2.1).

We remark that there are a number of articles which extend the scope of the Chernozhukov–Chetverikov–Kato theory (CCK theory for short) in various directions. We refer the reader to the survey [6] for recent developments. Nevertheless, most studies focus on *linear* statistics (i.e., sums of random variables) and there are only a few articles concerned with *nonlinear* statistics. Two exceptions are U-statistics developed in [10–12,56] and Wiener functionals developed in [34,35]. On the one hand, however, the former are mainly concerned with non-degenerate U-statistics which are approximately linear statistics via Hoeffding decomposition (Chen & Kato [11] also handle degenerate U-statistics, but they focus on the randomized incomplete versions that are still approximately linear statistics). On the other hand, although the latter deal with essentially nonlinear statistics, they must be functionals of a (possibly infinite-dimensional) Gaussian process, except for [35, Theorem 3.2] that is a version of our result with $q_j \equiv 2$ (see Sect. 2.1.2 for more details). In this sense, our result would be the first extension of CCK-type results to essentially nonlinear statistics based on possibly non-Gaussian variables.

Finally, we mention that the main results of this paper have potential applications to statistics. In fact, the original motivation of this paper is to improve the Gaussian approximation result for maxima of high-dimensional vectors of random quadratic forms given by [35, Theorem 3.2], which is used to ensure the validity of the bootstrap testing procedure proposed in [35, Section 4.1] (see Sect. 2.2). Another potential application might be specification test for parametric form in nonparametric regression. In this area, to derive the null distributions of test statistics, one sometimes needs to approximate the maximum of (essentially degenerate) quadratic forms; see [25,32,40] for instance.

This paper is organized as follows. Section 2 presents the main results obtained in the paper, while Sects. 3–7 are devoted to the proof of the main results: Sect. 3 demonstrates a basic scheme of the CCK theory to prove high-dimensional CLTs. Subsequently, Sect. 4 presents a connection of this scheme to Stein's method. Based on this observation, Sect. 5 develops a high-dimensional CLT of the form (1.2) for homogeneous sums based on normal and gamma variables. Then, Sect. 6 establishes a kind of invariance principle for high-dimensional homogeneous sums using a randomized version of the Lindeberg method. Finally, Sect. 7 completes the proof of the main results.

Notation

$\mathbb{Z}_+$ denotes the set of all nonnegative integers. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we define $\|x\|_{\ell_\infty} := \max_{1 \le j \le d} |x_j|$. For $N \in \mathbb{N}$, we set $[N] := \{1, \ldots, N\}$. We set $\sum_{i=p}^{q} \equiv 0$ if $p > q$ by convention. For $q \in \mathbb{N}$, we denote by $\mathfrak{S}_q$ the set of all permutations of $[q]$, i.e., the symmetric group of degree q. For a function $f : [N]^q \to \mathbb{R}$, we set

$\mathcal{M}(f) := \max_{1 \leq i \leq N} \mathrm{Inf}_i(f)$ (recall that $\mathrm{Inf}_i(f)$ is defined according to (1.1)). We also set $\|f\|_{\ell_2} := \sqrt{\sum_{i_1,\dots,i_q=1}^{N} f(i_1, \dots, i_q)^2}$. For a function $h : \mathbb{R}^d \to \mathbb{R}$, we set $\|h\|_{\infty} := \sup_{x \in \mathbb{R}^d} |h(x)|$. We write $C_b^m(\mathbb{R}^d)$ for the set of all real-valued C^m functions on $\mathbb{R}^d$ all of whose partial derivatives are bounded. We write $\partial_{j_1 \dots j_m} = \frac{\partial^m}{\partial x_{j_1} \cdots \partial x_{j_m}}$ for short. Throughout the paper, $Z = (Z_1, \dots, Z_d)$ denotes a d-dimensional centered Gaussian random vector with covariance matrix $\mathfrak{C} = (\mathfrak{C}_{ij})_{1 \leq i,j \leq d}$ (note that we do not assume that $\mathfrak{C}$ is positive definite in general). Also, $(q_j)_{j=1}^{\infty}$ stands for a sequence of positive integers. Throughout the paper, we will regard $(q_j)_{j=1}^{\infty}$ as fixed, i.e., it does not vary when we consider asymptotic results. Given a probability distribution μ, we write $X \sim \mu$ to express that X is a random variable with distribution μ. For $\nu > 0$, we write $\gamma(\nu)$ for the gamma distribution with shape ν and rate 1. If S is a topological space, $\mathcal{B}(S)$ denotes the Borel σ-field of S.

Given a random variable X, we set $\|X\|_p := \{\mathrm{E}[|X|^p]\}^{1/p}$ for every $p > 0$. When X satisfies $\mathrm{E}[X^4] < \infty$, we denote the fourth cumulant of X by $\kappa_4(X)$. Note that $\kappa_4(X) = \mathrm{E}[X^4] - 3\mathrm{E}[X^2]^2$ if X is centered. For $\alpha > 0$, we define the ψ_α-norm of X by $\|X\|_{\psi_\alpha} := \inf\{C > 0 : \mathrm{E}[\psi_\alpha(|X|/C)] \leq 1\}$, where $\psi_\alpha(x) := \exp(x^\alpha) - 1$. Note that $\|\cdot\|_{\psi_\alpha}$ is indeed a norm (on a suitable space) if and only if $\alpha \geq 1$. Some useful properties of the ψ_α-norm are collected in Appendix A.

2 Main Results

Our first main result is a high-dimensional version of de Jong's CLT for homogeneous sums:

Theorem 2.1 *Let $X = (X_i)_{i=1}^{N}$ be a sequence of independent centered random variables with unit variance. Set $w = \frac{1}{2}$ if $\mathrm{E}[X_i^3] = 0$ for every $i \in [N]$ and $w = 1$ otherwise. For every $j \in [d]$, let $f_j : [N]^{q_j} \to \mathbb{R}$ be a symmetric function vanishing on diagonals, and set $Q(X) := (Q(f_1; X), \dots, Q(f_d; X))$. Suppose that $d \geq 2$, $\underline{\sigma} := \min_{1 \leq j \leq d} \|Z_j\|_2 > 0$ and $\max_{1 \leq i \leq N} \|X_i\|_{\psi_\alpha} < \infty$ for some $\alpha \in (0, w^{-1}]$. Then,*

$$
\sup_{x \in \mathbb{R}^d} |P(Q(X) \leq x) - P(Z \leq x)|
$$

$$
\leq C(1 + \underline{\sigma}^{-1}) \left\{ (\log d)^{\frac{2}{3}} \delta_0[Q(X)]^{\frac{1}{3}} \right.
$$

$$
\left. + (\log d)^{\mu + \frac{1}{2}} \delta_1[Q(X)]^{\frac{1}{3}} + (\log d)^{\frac{2\overline{q}_d - 1}{\alpha} + \frac{3}{2}} \max_{1 \leq k \leq d} \overline{B}_N^{q_k} \sqrt{\mathcal{M}(f_k)} \right\}, \quad (2.1)
$$

where $\overline{B}_N := \max_{1 \leq i \leq N}(\|X_i\|_{\psi_\alpha} \vee |\mathrm{E}[X_i^3]|)$, $\overline{q}_d := \max_{1 \leq j \leq d} q_j$, $\mu := \max\{\frac{2}{3}w\overline{q}_d - \frac{1}{6}, \frac{2(\overline{q}_d - 1)}{3\alpha} + \frac{1}{3}\}$, $C > 0$ depends only on $\alpha, \overline{q}_d$ and

$$
\delta_0[Q(X)] := \max_{1 \leq j,k \leq d} \left| \mathrm{E}[Q(f_j; X)Q(f_k; X)] - \mathfrak{C}_{jk} \right|,
$$

$$\delta_1[\boldsymbol{Q}(X)] := \overline{A}_N^{2w\overline{q}_d-1} \max_{1 \le j,k \le d} \left\{ 1_{\{q_j=q_k\}} \sqrt{|\kappa_4(Q(f_k;X))| + \overline{A}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2} \right.$$

$$\left. + 1_{\{q_j<q_k\}} \overline{A}_N^{q_j} \|f_j\|_{\ell_2} \left(|\kappa_4(Q(f_k;X))| + \overline{A}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2 \right)^{1/4} \right\}$$

with $\overline{A}_N := \max_{1 \le i \le N}(|\mathrm{E}[X_i^3]| \vee \|X_i\|_4)$.

Remark 2.1 (a) Since $\sum_{i=1}^N \mathrm{Inf}_i(f_k)^2 \le \|f_k\|_{\ell_2}^2 \mathcal{M}(f_k)$, Theorem 2.1 gives the bound of the form (1.2) under reasonable assumptions when $q_1 = \cdots = q_d$. For example, this is the case when $\mathrm{E}[Q(f_j;X)Q(f_k;X)] = \mathfrak{C}_{jk}$ for all $j,k \in [d]$, $\sup_i \|X_i\|_{\psi_\alpha} < \infty$ and $\sup_j \|f_j\|_{\ell_2} < \infty$. Here, we keep $\sum_{i=1}^N \mathrm{Inf}_i(f_k)^2$ rather than $\|f_k\|_{\ell_2}^2 \mathcal{M}(f_k)$ for the convenience of application.

(b) When $q_j < q_k$ for some $j,k \in [d]$, the exponents of $|\kappa_4(Q(f_k;X))|$ and $\mathcal{M}(f_k)$ appearing in the bound of (2.1) are 1/12, which are halves of those for the case $q_j = q_k$. This phenomenon is not specific to the high-dimensional setting but common in fourth-moment-type theorems. See Remark 1.9(a) in [29] for more details.

(c) In Sect. 2.1, we compare Theorem 2.1 to two existing results in some detail. The results therein show the dependence of the bound in (2.1) on the dimension d is as sharp as (and sometimes sharper than) the previous results.

We can easily extend Theorem 2.1 to a high-dimensional CLT for homogeneous sums in hyperrectangles as follows. Let $\mathcal{A}^{\mathrm{re}}(d)$ be the set of all hyperrectangles in $\mathbb{R}^d$, i.e., $\mathcal{A}^{\mathrm{re}}(d)$ consists of all sets A of the form

$$A = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : a_j \le x_j \le b_j \text{ for all } j = 1, \ldots, d\}$$

for some $-\infty \le a_j \le b_j \le \infty$, $j = 1, \ldots, d$.

Corollary 2.1 *Under the assumptions of Theorem 2.1, we have*

$$\sup_{A \in \mathcal{A}^{\mathrm{re}}(d)} |P(\boldsymbol{Q}(X) \in A) - P(Z \in A)|$$

$$\le C'(1 + \underline{\sigma}^{-1}) \left\{ (\log d)^{\frac{2}{3}} \delta_0[\boldsymbol{Q}(X)]^{\frac{1}{3}} \right.$$

$$\left. + (\log d)^{\mu+\frac{1}{2}} \delta_1[\boldsymbol{Q}(X)]^{\frac{1}{3}} + (\log d)^{\frac{2\overline{q}_d-1}{\alpha}+\frac{3}{2}} \max_{1 \le k \le d} \overline{B}_N^{q_k} \sqrt{\mathcal{M}(f_k)} \right\},$$

where $C' > 0$ *depends only on* $\alpha, \overline{q}_d$.

For application, it is often useful to restate Theorem 2.1 in an asymptotic form as follows.

Corollary 2.2 *Let* $X = (X_i)_{i=1}^\infty$ *be a sequence of independent centered random variables with unit variance. Set* $w = \frac{1}{2}$ *if* $\mathrm{E}[X_i^3] = 0$ *for every* $i \in \mathbb{N}$

and $w = 1$ otherwise. For every $n \in \mathbb{N}$, let $N_n, d_n \in \mathbb{N} \setminus \{1\}$ and $f_{n,k} : [N_n]^{q_k} \to \mathbb{R}$ ($k = 1, \ldots, d_n$) be symmetric functions vanishing on diagonals, and set $\mathbf{Q}^{(n)}(\mathbf{X}) := (Q(f_{n,1}; \mathbf{X}), \ldots, Q(f_{n,d}; \mathbf{X}))$. Moreover, for every $n \in \mathbb{N}$, let $Z^{(n)} = (Z_{n,1}, \ldots, Z_{n,d_n})$ be a d_n-dimensional centered Gaussian vector with covariance matrix $\mathfrak{C}_n = (\mathfrak{C}_{n,kl})_{1 \leq k,l \leq d_n}$. Suppose that $\overline{q}_\infty := \sup_{j \in \mathbb{N}} q_j < \infty$, $\inf_{n \in \mathbb{N}} \min_{1 \leq k \leq d_n} \|Z_{n,k}\|_2 > 0$, $\sup_{i \in \mathbb{N}} \|X_i\|_{\psi_\alpha} < \infty$ for some $\alpha \in (0, w^{-1}]$ and

$$(\log d_n)^2 \max_{1 \leq k,l \leq d_n} |\mathrm{E}[Q(f_{n,k}; \mathbf{X})Q(f_{n,l}; \mathbf{X})] - \mathfrak{C}_{n,kl}| \to 0 \tag{2.2}$$

as $n \to \infty$. Moreover, setting $a_1 := (4w\overline{q}_\infty - 2) \vee (4\alpha^{-1}(\overline{q}_\infty - 1) + 5)$ and $a_2 := 2\alpha^{-1}(2\overline{q}_\infty - 1) + 3$, we suppose that either one of the following conditions is satisfied:

(i) $(\log d_n)^{2a_1} \max_{1 \leq j \leq d_n} |\kappa_4(Q(f_{n,j}; \mathbf{X}))| \to 0$ and $(\log d_n)^{2a_1 \vee a_2} \max_{1 \leq j \leq d_n} \mathcal{M}(f_{n,j}) \to 0$ as $n \to \infty$.

(ii) $(\log d_n)^{a_1} \max_{1 \leq j \leq d_n} |\kappa_4(Q(f_{n,j}; \mathbf{X}))| \to 0$ and $(\log d_n)^{a_1 \vee a_2} \max_{1 \leq j \leq d_n} \mathcal{M}(f_{n,j}) \to 0$ as $n \to \infty$ and $q_1 = q_2 = \cdots$.

Then, we have $\sup_{A \in \mathcal{A}^{re}(d_n)} |P(\mathbf{Q}^{(n)}(\mathbf{X}) \in A) - P(Z^{(n)} \in A)| \to 0$ as $n \to \infty$.

Our second main result gives high-dimensional versions of fourth-moment theorems, universality results and Peccati–Tudor-type theorems for homogeneous sums:

Theorem 2.2 *Let us keep the same notation as in Corollary 2.2. Suppose that one of the following conditions is satisfied:*

(A) $\mathbf{X}$ *is a sequence of independent copies of a random variable X such that $\|X\|_{\psi_\alpha} < \infty$ for some $\alpha > 0$ and $\mathrm{E}[X^3] = 0$ and $\mathrm{E}[X^4] \geq 3$.*

(B) *For every i, X_i is a standardized Poisson random variable with intensity $\lambda_i > 0$, i.e., $\lambda_i + \sqrt{\lambda_i} X_i$ is a Poisson random variable with intensity λ_i. Moreover, $\inf_{i \in \mathbb{N}} \lambda_i > 0$.*

(C) *For every i, X_i is a standardized gamma random variable with shape $\nu_i > 0$ and unit rate, i.e., $\nu_i + \sqrt{\nu_i} X_i \sim \gamma(\nu_i)$. Moreover, $\inf_{i \in \mathbb{N}} \nu_i > 0$.*

Suppose also $2 \leq \inf_{j \in \mathbb{N}} q_j \leq \sup_{j \in \mathbb{N}} q_j < \infty$, $0 < \inf_{n \in \mathbb{N}} \min_{1 \leq j \leq d_n} \mathfrak{C}_{jj}^{(n)} \leq \sup_{n \in \mathbb{N}} \max_{1 \leq j \leq d_n} \mathfrak{C}_{jj}^{(n)} < \infty$ and

$$(\log d_n)^a \max_{1 \leq j,k \leq d_n} |\mathrm{E}[Q(f_{n,j}; \mathbf{X})Q(f_{n,k}; \mathbf{X})] - \mathfrak{C}_{n,jk}| \to 0$$

as $n \to \infty$ for every $a > 0$. Then, we have $\kappa_4(Q(f; \mathbf{X})) \geq 0$ for any symmetric function $f : [N]^q \to \mathbb{R}$ vanishing on diagonals. Moreover, the following conditions are equivalent:

(i) $(\log d_n)^a \max_{1 \leq j \leq d_n} \kappa_4(Q(f_{n,j}; \mathbf{X})) \to 0$ *as $n \to \infty$ for every $a > 0$.*

(ii) $(\log d_n)^a \max_{1 \leq j \leq d_n} \sup_{x \in \mathbb{R}} |P(Q(f_{n,j}; \mathbf{X}) \leq x) - P(Z_{n,j} \leq x)| \to 0$ *as $n \to \infty$ for every $a > 0$.*

(iii) $(\log d_n)^a \sup_{x \in \mathbb{R}^{d_n}} |P(\boldsymbol{Q}^{(n)}(\boldsymbol{X}) \le x) - P(\boldsymbol{Z}^{(n)} \le x)| \to 0$ *as* $n \to \infty$ *for every* $a > 0$.

(iv) $(\log d_n)^a \sup_{x \in \mathbb{R}^{d_n}} |P(\boldsymbol{Q}^{(n)}(\boldsymbol{Y}) \le x) - P(\boldsymbol{Z}^{(n)} \le x)| \to 0$ *as* $n \to \infty$ *for any* $a > 0$ *and sequence* $\boldsymbol{Y} = (Y_i)_{i \in \mathbb{N}}$ *of centered independent variables with unit variance such that* $\sup_{i \in \mathbb{N}} \|Y_i\|_{\psi_\alpha} < \infty$ *for some* $\alpha > 0$.

Remark 2.2 (a) The implications (i) $\Rightarrow$ (iii), (iii) $\Rightarrow$ (iv) and (ii) $\Rightarrow$ (iii) can be viewed as high-dimensional versions of fourth-moment theorems, universality results and Peccati–Tudor-type theorems for homogeneous sums, respectively. Here, *Peccati–Tudor-type theorems* refer to statements such that a joint CLT is implied by component-wise CLTs (Peccati & Tudor [51] have established such a result for multiple Wiener–Itô integrals with respect to an isonormal Gaussian process).

(b) The proof of Theorem 2.2 relies on the fact that condition (i) *automatically* yields $(\log d_n)^a \max_j \mathcal{M}(f_{n,j}) \to 0$ as $n \to \infty$ for every $a > 0$. On the one hand, this fact has already been established in the previous work for cases (A) and (B) (see the proof of Lemma 7.2). On the other hand, for case (C), this fact seems not to have appeared in the literature so far. Indeed, for case (C) we obtain it as a byproduct of the proof of Proposition 5.2 (see Lemma 5.4). As a consequence, Theorem 2.2 seems new for case (C) even in the fixed-dimensional case. We remark that the fourth-moment theorem for case (C) has been established by [1] in the univariate case, which inspired our discussions in Sect. 5 (see also [9]).

2.1 Comparison of Theorem 2.1 to Some Existing Results

2.1.1 Comparison to Corollary 7.3 in Nourdin, Peccati and Reinert [47]

First, we compare our result to the quantitative multi-dimensional CLT for homogeneous sums obtained in Nourdin et al. [47]. To state their result, we need to introduce the notion of *contraction*, which will also play an important role in Sect. 5.2. For two symmetric functions $f : [N]^p \to \mathbb{R}$, $g : [N]^q \to \mathbb{R}$ and $r \in \{0, 1 \dots, p \wedge q\}$, we define the contraction $f \star_r g : [N]^{p+q-2r} \to \mathbb{R}$ by

$$f \star_r g(i_1, \dots, i_{p+q-2r}) = \sum_{k_1, \dots, k_r = 1}^{N} f(i_1, \dots, i_{p-r}, k_1, \dots, k_r) g$$
$$(i_{p-r+1}, \dots, i_{p+q-2r}, k_1, \dots, k_r). \qquad (2.3)$$

In particular, we have

$$f \star_0 g(i_1, \dots, i_{p+q}) = f \otimes g(i_1, \dots, i_{p+q}) = f(i_1, \dots, i_p) g(i_{p+1}, \dots, i_{p+q}).$$

Now we are ready to state the result of [47]. To simplify the notation, we focus only on the identity covariance matrix case and do not keep the explicit dependence of constants on q_j's.

Proposition 2.1 *(Nourdin et al. [47], Corollary 7.3) Let us keep the same notation as in Theorem 2.1. Suppose that* $\mathfrak{C}_{jk} = \mathrm{E}[Q(f_j; \boldsymbol{X})Q(f_k; \boldsymbol{X})]$ *for all* $j, k \in [d]$ *and* $\mathfrak{C}$

is the identity matrix of size d. Suppose also that $\beta := \max_{1 \leq i \leq N} \mathrm{E}[|X_i|^3] < \infty$ *and* $q_d \geq \cdots \geq q_1 \geq 2$. *Then, we have*

$$
\sup_{A \in \mathcal{C}(\mathbb{R}^d)} |P(Q(X) \in A) - P(Z \in A)|
$$

$$
\leq K d^{3/8} \left\{ \overline{\Delta} + \mathsf{C}(\beta + 1) \left(\sum_{j=1}^d \beta^{(q_j-1)/3} \right)^3 \sqrt{\max_{1 \leq j \leq d} \mathcal{M}(f_j)} \right\}^{1/4}, \quad (2.4)
$$

where $\mathcal{C}(\mathbb{R}^d)$ *is the set of all convex Borel subsets of* $\mathbb{R}^d$, $K > 0$ *is a constant depending only on* $\overline{q}_d$, $\mathsf{C} := \sum_{i=1}^N \max_{1 \leq j \leq d} \mathrm{Inf}_i(f_j)$ *and*

$$
\overline{\Delta} := \sum_{1 \leq j \leq k \leq d} \left(\sum_{r=1}^{q_j-1} (\|f_j \star_{q_j-r} f_j\|_{\ell_2} + \|f_k \star_{q_k-r} f_k\|_{\ell_2}) + 1_{\{q_j < q_k\}} \sqrt{\|f_k \star_{q_k-q_j} f_k\|_{\ell_2}} \right).
$$

To compare Proposition 2.1 to our result, we need to bound the quantity $\overline{\Delta}$ by $|\kappa_4(Q(f_j; X))|$ and $\mathcal{M}(f_j)$, $j \in [d]$. This can be carried out by the following lemma (proved in Sect. 7.4):

Lemma 2.1 *Let* $X = (X_i)_{i=1}^N$ *be a sequence of independent centered random variables with unit variance and such that* $M := 1 + \max_{1 \leq i \leq N} \mathrm{E}[X_i^4] < \infty$. *Also, let* $q \geq 2$ *be an integer and* $f : [N]^q \to \mathbb{R}$ *be a symmetric function vanishing on diagonals. Then, we have*

$$
\max_{1 \leq r \leq q-1} \|f \star_r f\|_{\ell_2} \leq \sqrt{|\kappa_4(Q(f; X))| + CM \|f\|_{\ell_2}^2 \mathcal{M}(f)},
$$

where $C > 0$ *depends only on* q.

Remark 2.3 The bound in Lemma 2.1 is generally sharp. In fact, it is well known that $\sqrt{|\kappa_4(Q(f; X))|}$ has the same order as $\max_{1 \leq r \leq q-1} \|f \star_r f\|_{\ell_2}$ if X is Gaussian (see, e.g., Eq.(5.2.6) in [44]). Moreover, if $q = 2$ and $f(i, j) = N^{-1/2} 1_{\{|i-j|=1\}}$, then both $\|f \star_1 f\|_{\ell_2}$ and $\|f\|_{\ell_2} \sqrt{\mathcal{M}(f)}$ are of order $N^{-1/2}$.

With the help of Lemma 2.1, we observe that the bound in (2.4) typically has the same order as

$$
d^{3/8} \left\{ d^2 \max_{1 \leq j,k \leq d} \widehat{\Delta}_{jk} + d^3 \mathsf{C} \max_{1 \leq j \leq d} \sqrt{\mathcal{M}(f_j)} \right\}^{1/4},
$$

where

$$
\widehat{\Delta}_{jk} := 1_{\{q_j=q_k\}} \sqrt{|\kappa_4(Q(f_j; X))| + \mathcal{M}(f_j)}
$$
$$
+ 1_{\{q_j < q_k\}} \{|\kappa_4(Q(f_k; X))| + \mathcal{M}(f_k)\}^{1/4}.
$$

Thus, in the bound of (2.4), the dimension appears as a power of d, while the exponent of the "standard" convergence rate $\delta := \max_{1 \le j \le d} \sqrt{|\kappa_4(Q(f_j; X))| + \mathcal{M}(f_j)}$ is $1/4$. These are much improved in our result because the former appears as a power of $\log d$ and the latter is $1/3$. Nevertheless, we should note that the bound in (2.4) is given for the much stronger metric than the Kolmogorov distance. In fact, to the best of the author's knowledge, all the known bounds for this metric depend polynomially on the dimension even for sums of independent random variables; see [58, Section 1.1] and references therein.

Remark 2.4 (a) Roughly speaking, the exponent of δ is $1/4$ in the bound of (2.4) because this bound is transferred from an analogous quantitative CLT for the Gaussian counterpart by the Lindeberg method with matching moments *up to the second order*. To overcome this issue, we need to match moments *up to the third order* and thus we can no longer rely on the result analogous to Theorem 2.1 for the Gaussian counterpart, which is obtained in [35]. For this reason, we will develop a high-dimensional CLT for homogeneous sums based on normal and gamma variables in Sect. 5.

(b) It is worth noting that the quantity $\mathsf{C} = \sum_{i=1}^N \max_{1 \le j \le d} \mathrm{Inf}_i(f_j)$ in the bound of (2.4) can be much larger than $\max_{1 \le j \le d} \sum_{i=1}^N \mathrm{Inf}_i(f_j) = \max_{1 \le j \le d} \|f_j\|_{\ell_2}^2$ in high-dimensional situations (see Remark 2.5 for a concrete example). Indeed, naïve application of the Lindeberg method produces a quantity like C, which prevents us from using the Lindeberg method in its pure form (this is why Chernozhukov et al. [13,18] rely on Stein's method to prove their high-dimensional CLTs; see [14, Appendix L] for a detailed discussion). In Sect. 6, we will resolve this issue by *randomizing* the Lindeberg method as Deng & Zhang [24] have recently done in the context of sums of independent random vectors.

2.1.2 Comparison to Theorem 3.2 in Koike [35]

Next, we compare our result to the Gaussian approximation result for maxima of quadratic forms obtained in [35, Theorem 3.2]. Here, for an explicit comparison, we state this result with applying [35, Corollary 3.1]. For a function $f : [N]^2 \to \mathbb{R}$, we denote the $N \times N$ matrix $(f(i, j))_{1 \le i,j \le N}$ by $[f]$.

Proposition 2.2 *(Koike [35], Theorem 3.2 and Corollary 3.1) Let us keep the same notation as in Corollary 2.2. Suppose that $\inf_{n \in \mathbb{N}} \min_{1 \le k \le d_n} \|Z_{n,k}\|_2 > 0$, $q_j = 2$ for all j, $\sup_{i \in \mathbb{N}} \|X_i\|_{\psi_2} < \infty$ and (2.2) holds true as $n \to \infty$. Suppose also that*

$$(\log d_n)^3 \max_{1 \le k \le d_n} \sqrt{\mathrm{tr}\left([f_{n,k}]^4\right)}$$

$$+ (\log d_n)^5 \left(\max_{1 \le k \le d_n} \sqrt{\mathcal{M}(f_{n,k})} \right) \sum_{i=1}^{N_n} \max_{1 \le k \le d_n} \mathrm{Inf}_i(f_{n,k}) \to 0 \qquad (2.5)$$

as $n \to \infty$. Then, we have

$$\sup_{t \in \mathbb{R}} \left| P\left(\max_{1 \le k \le d_n} |Q(f_{n,k}; X)| \le t \right) - P\left(\max_{1 \le k \le d_n} |Z_{n,k}| \le t \right) \right| \to 0 \quad (2.6)$$

as $n \to \infty$.

When we apply our result to quadratic forms as above, we obtain the following result.

Proposition 2.3 *Let us keep the same notation as in Corollary 2.2. Set*

$$\Delta_n := \max_{1 \le k \le d_n} \sqrt{\operatorname{tr}\left([f_{n,k}]^4\right)} + \max_{1 \le k \le d_n} \sqrt{\mathcal{M}(f_{n,k})} \|f_{n,k}\|_{\ell_2}$$

for every n. Assume $q_1 = q_2 = \cdots = 2$, $\inf_{n \in \mathbb{N}} \min_{1 \le k \le d_n} \|Z_{n,k}\|_2 > 0$ and (2.2). Assume also that either one of the following conditions is satisfied:

(i) $\sup_{i \in \mathbb{N}} \|X_i\|_{\psi_1} < \infty$ *and* $(\log d_n)^5 \Delta_n \to 0$ *as $n \to \infty$.*
(ii) $\sup_{i \in \mathbb{N}} \|X_i\|_{\psi_2} < \infty$, $\mathrm{E}[X_i^3] = 0$ *for all i and* $(\log d_n)^3 \Delta_n \to 0$ *as $n \to \infty$.*

Then, we have $\sup_{A \in \mathcal{A}^{\mathrm{re}}(d_n)} |P(Q^{(n)}(X) \in A) - P(Z^{(n)} \in A)| \to 0$ as $n \to \infty$.

Remark 2.5 (a) Regarding the convergence rate of $\max_{1 \le k \le d_n} \operatorname{tr}\left([f_{n,k}]^4\right)$, condition (i) in Proposition 2.3 is stronger than the one in Proposition 2.2. However, the former imposes a weaker moment condition on X than the latter. More importantly, the second term of Δ_n is always smaller than or equal to the second term in (2.5), and the latter can be much larger than the former. For example, let us assume $N_n = d_n = n$ and consider the functions $f_{n,k}$ defined as follows:

$$f_{n,k}(i, j) = \begin{cases} n^{-1/2} & \text{if } |i - j| = 1, i \ne k, j \ne k, \\ n^{-1/4} & \text{if } |i - j| = 1, i = k \text{ or } j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Then, we have $\operatorname{Inf}_i(f_{n,k}) = (1 + 1_{\{1 < i < n\}}) n^{-1/2}$ if $i \in \{k, k \pm 1\}$ and $\operatorname{Inf}_i(f_{n,k}) = (1 + 1_{\{1 < i < n\}}) n^{-1}$ otherwise. Therefore, on the one hand

$$\left(\max_{1 \le k \le d_n} \sqrt{\mathcal{M}(f_{n,k})} \right) \sum_{i=1}^{N_n} \max_{1 \le k \le d_n} \operatorname{Inf}_i(f_{n,k})$$

does not converge to 0 as $n \to \infty$, but on the other hand $\max_{1 \le k \le d_n} \sqrt{\mathcal{M}(f_{n,k})} \|f_{n,k}\|_{\ell_2} = O(n^{-1/4})$ as $n \to \infty$. Note that in this case we have $\max_{1 \le k \le d_n} \sqrt{\operatorname{tr}\left([f_{n,k}]^4\right)} = O(n^{-1/4})$ and $\|f_{n,k}\|_{\ell_2} \to 1$ as $n \to \infty$, so (2.6) holds true due to Proposition 2.3.

(b) Condition (ii) in Proposition 2.3 requires the additional zero skewness assumption, but it always improves the assumption on the functions $f_{n,k}$ than the one in Proposition 2.2.

(c) We have $\Delta_n \leq 2 \max_k \|[f_{n,k}]\|_{sp} \|f_{n,k}\|_{\ell_2}$ with $\|\cdot\|_{sp}$ the spectral norm of matrices. So $(\log d_n)^a \Delta_n \to 0$ for some $a > 0$ is implied by $(\log d_n)^a \max_k \|[f_{n,k}]\|_{sp} \|f_{n,k}\|_{\ell_2} \to 0$.

2.2 Statistical Application: Bootstrap Test for the Absence of Lead–Lag Relationship

Let $W_t = (W_t^1, W_t^2)$ $(t \in \mathbb{R})$ be a two-sided bivariate standard Wiener process. Also let $\rho \in (-1, 1)$ and $\vartheta \in \mathbb{R}$ be two (unknown) parameters. We define the bivariate process $B_t = (B_t^1, B_t^2)$ $(t \in \mathbb{R})$ as $B_t^1 = W_t^1$ and $B_t^2 = \rho W_{t-\vartheta}^1 + \sqrt{1-\rho^2} W_t^2$. For each $\nu = 1, 2$, we consider the process $X^\nu = (X_t^\nu)_{t \geq 0}$ given by

$$X_t^\nu = X_0^\nu + \int_0^t \sigma_\nu(s) \mathrm{d}B_s^\nu, \qquad t \geq 0, \tag{2.7}$$

where $\sigma_\nu \in L^2(0, \infty)$ is nonnegative-valued and deterministic. If $\rho \neq 0$, there is a correlation between X^1 and X^2 with a time lag of ϑ. We aim to test for whether such a correlation really exists or not, given (possibly asynchronous) high-frequency observations of X^1 and X^2. Specifically, for each $\nu = 1, 2$, we observe the process X^ν on the interval $[0, T]$ at the deterministic sampling times $0 \leq t_0^\nu < t_1^\nu < \cdots < t_{n_\nu}^\nu \leq T$, which implicitly depend on the parameter $n \in \mathbb{N}$ such that

$$r_n := \max_{\nu=1,2} \max_{i=0,1,\dots,n_\nu+1} (t_i^\nu - t_{i-1}^\nu) \to 0$$

as $n \to \infty$, where we set $t_{-1}^\nu := 0$ and $t_{n_\nu+1}^\nu := T$ for each $\nu = 1, 2$. To test for the null hypothesis $H_0 : \rho = 0$ against the alternative $H_1 : \rho \neq 0$, Koike [35] proposed the test statistic given by $T_n = \sqrt{n} \max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$, where $\mathcal{G}_n$ is a finite subset of $\mathbb{R}$ and

$$U_n(\theta) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Delta_i^n X^1 \Delta_j^n X^2 K_\theta^{ij} \qquad \text{with } \Delta_i^n X^\nu = X_{t_i^\nu}^\nu - X_{t_{i-1}^\nu}^\nu \text{ and}$$

$$K_\theta^{ij} = 1_{\{(t_{i-1}^1, t_i^1] \cap (t_{j-1}^2 - \theta, t_j^2 - \theta] \neq \emptyset\}}.$$

The null distribution of T_n can be approximated by its Gaussian analog as follows:

Proposition 2.4 *([35], Proposition 4.1) For each $n \in \mathbb{N}$, let $(Z_n(\theta))_{\theta \in \mathcal{G}_n}$ be a family of centered Gaussian variables such that $\mathrm{E}[Z_n(\theta)Z_n(\theta')] = n \operatorname{Cov}[U_n(\theta), U_n(\theta')]$ for all $\theta, \theta' \in \mathcal{G}_n$. Suppose that $\sup_{t \in [0,T]}(\sigma_1(t) + \sigma_2(t)) < \infty$ and there are positive constants $\underline{v}, \overline{v}$ such that*

$$\underline{v} \leq n \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(\int_{t_{i-1}^1}^{t_i^1} \sigma_1(t)^2 \mathrm{d}t \right) \left(\int_{t_{j-1}^2}^{t_j^2} \sigma_2(t)^2 \mathrm{d}t \right) K_\theta^{ij} \leq \overline{v}$$

for all $n \in \mathbb{N}$ and $\theta \in \mathcal{G}_n$. Then, under the null hypothesis $\rho = 0$, we have

$$\sup_{x \in \mathbb{R}} \left| P\left(T_n \leq x \right) - P\left(\max_{\theta \in \mathcal{G}_n} |Z_n(\theta)| \leq x \right) \right| \to 0$$

as $n \to \infty$, provided that $nr_n^2 \log^6(\#\mathcal{G}_n) \to 0$.

Since the distribution of $\max_{\theta \in \mathcal{G}_n} |Z_n(\theta)|$ is analytically intractable, Koike [35] proposed a wild bootstrap procedure to approximate it. Formally, let $(w_i^1)_{i=1}^{\infty}$ and $(w_j^2)_{j=1}^{\infty}$ be mutually independent sequences of i.i.d. random variables independent of X^1 and X^2. Assume that $E[w_1^1] = E[w_1^2] = 0$, $\text{Var}[w_1^1] = \text{Var}[w_1^2] = 1$ and $\|w_1^1\|_{\psi_2} \vee \|w_1^2\|_{\psi_2} < \infty$. Define the bootstrapped test statistic as $T_n^* = \sqrt{n} \max_{\theta \in \mathcal{G}_n} |U_n^*(\theta)|$ where

$$U_n^*(\theta) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(w_i^1 \Delta_i^n X^1 \right) \left(w_j^2 \Delta_j^n X^2 \right) K_\theta^{ij}.$$

In [35, Proposition B.8], it is shown that

$$\sup_{x \in \mathbb{R}} \left| P\left(T_n^* \leq x \mid X \right) - P\left(\max_{\theta \in \mathcal{G}_n} |Z_n(\theta)| \leq x \right) \right| \to^p 0 \qquad (2.8)$$

as $n \to \infty$, provided that $r_n = O(n^{-3/4-\eta})$ and $\#\mathcal{G}_n = O(n^\gamma)$ for some $\eta, \gamma > 0$ in addition to the assumptions of Proposition 2.4. Our result allows us to relax the condition on r_n as follows:

Proposition 2.5 *Under the assumptions of Proposition 2.4, we have (2.8) as $n \to \infty$, provided that $r_n = O(n^{-1/2-\eta})$ and $\#\mathcal{G}_n = O(n^\gamma)$ for some $\eta, \gamma > 0$.*

3 Chernozhukov–Chetverikov–Kato Theory

In this section we demonstrate a basic scheme of the CCK theory to establish high-dimensional CLTs. One main ingredient of the CCK theory is the following smooth approximation of the maximum function: For each $\beta > 0$, we define the function $\Phi_\beta : \mathbb{R}^d \to \mathbb{R}$ by

$$\Phi_\beta(x) = \beta^{-1} \log \left(\sum_{j=1}^d e^{\beta x_j} \right), \qquad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Eq.(1) in [16] states that

$$0 \leq \Phi_\beta(x) - \max_{1 \leq j \leq d} x_j \leq \beta^{-1} \log d \qquad (3.1)$$

for any $x \in \mathbb{R}^d$. Therefore, the larger β is, the better Φ_β approximates the maximum function. The next lemma, which is a summary of [24, Lemmas 5–6], highlights the key properties of this smooth max function:

Lemma 3.1 *For any $\beta > 0$, $m \in \mathbb{N}$ and C^m function $h : \mathbb{R} \to \mathbb{R}$, there is an $\mathbb{R}^{\otimes m}$-valued function $\Upsilon_\beta(x) = (\Upsilon_\beta^{j_1,...,j_m}(x))_{1 \leq j_1,...,j_m \leq d}$ on $\mathbb{R}^d$ satisfying the following conditions:*

(i) *For any $x \in \mathbb{R}^d$ and $j_1, \ldots, j_m \in [d]$, we have $|\partial_{j_1...j_m}(h \circ \Phi_\beta)(x)| \leq \Upsilon_\beta^{j_1,...,j_m}(x)$.*

(ii) *For every $x \in \mathbb{R}^d$, we have*

$$\sum_{j_1,...,j_m=1}^{d} \Upsilon_\beta^{j_1,...,j_m}(x) \leq c_m \max_{1 \leq k \leq m} \beta^{m-k} \|h^{(k)}\|_\infty,$$

where $c_m > 0$ depends only on m.

(iii) *For any $x, t \in \mathbb{R}^d$ and $j_1, \ldots, j_m \in [d]$, we have*

$$e^{-8\|t\|_{\ell_\infty}\beta} \Upsilon_\beta^{j_1,...,j_m}(x+t) \leq \Upsilon_\beta^{j_1,...,j_m}(x) \leq e^{8\|t\|_{\ell_\infty}\beta} \Upsilon_\beta^{j_1,...,j_m}(x+t).$$

Remark 3.1 An explicit expression of the constant c_m in Lemma 3.1 can be derived from [24, Lemma 5]. In particular, we have $c_1 = 1$ and $c_2 = 3$.

Another important ingredient of the CCK theory is the so-called *anti-concentration inequality*. For our purpose, the following one is particularly useful (see [19] for the proof):

Lemma 3.2 *(Nazarov's inequality) If $\underline{\sigma} := \min_{1 \leq j \leq d} \|Z_j\|_2 > 0$, for any $x \in \mathbb{R}^d$ and $\varepsilon > 0$ we have*

$$P(Z \leq x + \varepsilon) - P(Z \leq x) \leq \frac{\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right).$$

These tools enable us to establish the following form of smoothing inequality:

Proposition 3.1 *Let $g_0 : \mathbb{R} \to [0, 1]$ be a measurable function such that $g_0(t) = 1$ for $t \leq 0$ and $g_0(t) = 0$ for $t \geq 1$. Also, let $\varepsilon > 0$ and set $\beta := \varepsilon^{-1} \log d$. Suppose that $\underline{\sigma} := \min_{1 \leq j \leq d} \|Z_j\|_2 > 0$. Then, for any d-dimensional random vector F, we have*

$$\sup_{x \in \mathbb{R}^d} |P(F \leq x) - P(Z \leq x)| \leq \Delta_\varepsilon(F, Z) + \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right), \quad (3.2)$$

where

$$\Delta_\varepsilon(F, Z) := \sup_{y \in \mathbb{R}^d} \left| \mathrm{E}[g_0(\varepsilon^{-1}\Phi_\beta(F - y))] - \mathrm{E}[g_0(\varepsilon^{-1}\Phi_\beta(Z - y))] \right|. \quad (3.3)$$

Proof This result has been essentially shown in Step 2 in the proof of [18, Lemma 5.1]. □

Remark 3.2 Proposition 3.1 can be seen as a special version of more general smoothing inequalities such as [7, Lemma 2.1]. An important feature of bound (3.2) is that the quantity $\Delta_\varepsilon(F, Z)$ contains only test functions of the form $x \mapsto g_0(\Phi_\beta(x - y))$ for some $y \in \mathbb{R}^d$. If g_0 is sufficiently smooth, derivatives of such a test function admit good estimates with respect to the dimension d, as seen from Lemma 3.1.

It might be worth mentioning that we can use Proposition 3.1 to derive a bound for the Kolmogorov distance by the Wasserstein distance. Let us recall the definition of the Wasserstein distance.

Definition 3.1 (Wasserstein distance) For d-dimensional random vectors F, G with integrable components, the *Wasserstein distance* between the laws of F and G is defined by

$$\mathcal{W}_1(F, G) := \sup_{h \in \mathcal{H}} |E[h(F)] - E[h(G)]|,$$

where $\mathcal{H}$ denotes the set of all functions $h : \mathbb{R}^d \to \mathbb{R}$ such that

$$\|h\|_{\mathrm{Lip}} := \sup_{x, y \in \mathbb{R}^d : x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|} \leq 1.$$

Here, $\| \cdot \|$ is the usual Euclidian norm on $\mathbb{R}^d$.

Corollary 3.1 *Under the assumptions of Proposition 3.1, we have*

$$\sup_{x \in \mathbb{R}^d} |P(F \leq x) - P(Z \leq x)| \leq \sqrt{\frac{2 \left(\sqrt{2 \log d} + 2\right)}{\underline{\sigma}} \mathcal{W}_1(F, Z)}.$$

Proof It suffices to consider the case $\mathcal{W}_1(F, Z) > 0$. Let us define the function $g_0 : \mathbb{R} \to [0, 1]$ by $g_0(x) = \min\{1, \max\{1-x, 0\}\}, x \in \mathbb{R}$. Then, for any $x, x', y \in \mathbb{R}^d$ and $\varepsilon > 0$, we have $|g_0(\varepsilon^{-1}\Phi_\beta(x - y) - g_0(\varepsilon^{-1}\Phi_\beta(x' - y)| \leq \varepsilon^{-1}\|x - x'\|_{\ell_\infty}$ by [13, Lemma A.3], so we obtain $\Delta_\varepsilon(F, Z) \leq \varepsilon^{-1}\mathcal{W}_1(F, Z)$. Now, setting $\varepsilon = \sqrt{\underline{\sigma}\mathcal{W}_1(F, Z)/(2\sqrt{2 \log d} + 4)}$, we infer the desired result from Proposition 3.1. □

When $d = 1$, Corollary 3.1 recovers the standard estimate (cf. Eq.(C.2.6) in [44]). We shall remark that a bound similar to the above (with a slightly different constant) has already appeared in [4, Theorem 3.1].

Remark 3.3 It is generally impossible to derive (1.2)-type bounds from the corresponding ones for the Wasserstein distance. To see this, let $F = (F_1, \ldots, F_d)$ be a d-dimensional random vector such that the laws of $F_1, \ldots, F_d$ are identical (and integrable). Also, let $G = (G_1, \ldots, G_d)$ be another d-dimensional random vector satisfying the same condition. Then, we can easily verify $\mathcal{W}_1(F, G) \geq \sqrt{d}\mathcal{W}_1(F_1, G_1)$ by definition.

4 Stein Kernels and High-Dimensional CLTs

In the rest of the paper, we fix a C^∞ function $g_0 : \mathbb{R} \to [0, 1]$ such that $g_0(t) = 1$ for $t \leq 0$ and $g_0(t) = 0$ for $t \geq 1$: For example, we can take it as $g_0(t) = f_0(1 - t)/\{f_0(t) + f_0(1 - t)\}$, where the function $f_0 : \mathbb{R} \to \mathbb{R}$ is defined by $f_0(t) = e^{-1/t}$ if $t > 0$ and $f_0(t) = 0$ otherwise.

To make Proposition 3.1 useful, we need to obtain a "good" upper bound for the quantity $\Delta_\varepsilon(F, Z)$. As briefly mentioned in Remark 2.4, Chernozhukov et al. [13] have pointed out that Stein's method effectively solves this task. Moreover, discussions in [16,35] implicitly suggest that the CCK theory would have a nice connection to Stein kernels. In this section, we illustrate this idea.

Definition 4.1 (Stein kernel) Let $F = (F_1, \ldots, F_d)$ be a centered d-dimensional random variable. A $d \times d$ matrix-valued measurable function $\tau_F = (\tau_F^{ij})_{1 \leq i, j \leq d}$ on $\mathbb{R}^d$ is called a *Stein kernel* for (the law of) F if $\max_{1 \leq i, j \leq d} \mathrm{E}[|\tau_F^{ij}(F)|] < \infty$ and

$$\sum_{j=1}^d \mathrm{E}[\partial_j \varphi(F) F_j] = \sum_{i,j=1}^d \mathrm{E}[\partial_{ij} \varphi(F) \tau_F^{ij}(F)] \tag{4.1}$$

for any $\varphi \in C_b^\infty(\mathbb{R}^d)$.

Remark 4.1 In this paper, we adopt $C_b^\infty(\mathbb{R}^d)$ as the class of test functions for which identity (4.1) holds true because of convenience, but other classes are also used in the literature; see [20] for instance.

Lemma 4.1 *Let $F = (F_1, \ldots, F_d)$ be a centered d-dimensional random vector. Also, let $\tau_F = (\tau_F^{ij})_{1 \leq i, j \leq d}$ be a Stein kernel for F. Then, we have*

$$\sup_{y \in \mathbb{R}^d} \left| \mathrm{E}\left[h\left(\Phi_\beta(F - y) \right) \right] - \mathrm{E}\left[h\left(\Phi_\beta(Z - y) \right) \right] \right| \leq \frac{3}{2} \max\{\|h''\|_\infty, \beta\|h'\|_\infty\}\Delta$$

for any $\beta > 0$ and $h \in C_b^\infty(\mathbb{R})$, where

$$\Delta := \mathrm{E}\left[\max_{1 \leq i, j \leq d} |\tau_F^{ij}(F) - \mathfrak{C}_{ij}| \right].$$

Proof The proof is essentially same as that of [16, Theorem 1] or [35, Proposition 2.1], so we omit it. □

Proposition 4.1 *Suppose that $d \geq 2$ and $\underline{\sigma} := \min_{1 \leq j \leq d} \|Z_j\|_2 > 0$. Under the assumptions of Lemma 4.1, there is a universal constant $C > 0$ such that*

$$\sup_{x \in \mathbb{R}^d} |P(F \leq x) - P(Z \leq x)| \leq C(1 + \underline{\sigma}^{-1})\Delta^{1/3}(\log d)^{2/3}. \tag{4.2}$$

Proof Thanks to [44, Lemma 4.1.3], it suffices to consider the case $\Delta > 0$. By Lemma 4.1, for any $\varepsilon > 0$ we have $\Delta_\varepsilon(F, Z) \leq C' \varepsilon^{-2} (\log d) \Delta$, where $C' > 0$ is a universal constant. Therefore, Proposition 3.1 yields

$$\sup_{x \in \mathbb{R}^d} |P(F \leq x) - P(Z \leq x)| \leq C' \varepsilon^{-2} (\log d) \Delta + \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right).$$

Now, setting $\varepsilon = \Delta^{1/3} (\log d)^{1/6}$, we obtain the desired result.   □

5 A High-Dimensional CLT for Normal-Gamma Homogeneous Sums

In view of the results in Sect. 4, we naturally seek a situation where a vector of homogeneous sums has a Stein kernel. This is the case when all the components are eigenfunctions of a Markov diffusion operator (cf. Proposition 5.1 in [39]). Moreover, as clarified in [1,9,38], only some spectral properties of the Markov diffusion operator are essential for deriving a fourth-moment-type bound for the variance of the corresponding Stein kernel. This spectral property is especially satisfied when each X_i is either a Gaussian or (standardized) gamma variable, so this section focuses on such a situation and derives a high-dimensional CLT for this special case.

For each $\nu > 0$, we denote by $\gamma_\pm(\nu)$ the distribution of the random variable $\pm(X - \nu)/\sqrt{\nu}$ with $X \sim \gamma(\nu)$. Also, for every $q \in \mathbb{N}$ we set $\mathfrak{c}_q := \sum_{r=1}^q r! \binom{q}{r}^2$.

Proposition 5.1 *Let us keep the same notation as in Theorem 2.1 and assume $d \geq 2$. Let $\boldsymbol{Y} = (Y_i)_{i=1}^N$ be a sequence of independent random variables such that the law of Y_i belongs to $\{\mathcal{N}(0,1)\} \cup \{\gamma_+(\nu) : \nu > 0\} \cup \{\gamma_-(\nu) : \nu > 0\}$ for all i. For every i, define the constants v_i and η_i by*

$$v_i := \begin{cases} 2 & \text{if } Y_i \sim \mathcal{N}(0,1), \\ 2(1 + \nu^{-1}) & \text{if } Y_i \sim \gamma_\pm(\nu), \end{cases} \qquad \eta_i := \begin{cases} 1 & \text{if } Y_i \sim \mathcal{N}(0,1), \\ 1 \wedge \sqrt{\nu} & \text{if } Y_i \sim \gamma_\pm(\nu). \end{cases}$$

We also set $w_* = 1/2$ if $Y_i \sim \mathcal{N}(0,1)$ for all i and $w_* = 1$ otherwise. Then, $\kappa_4(Q(f_j; \boldsymbol{Y})) \geq 0$ for all j and

$$\sup_{y \in \mathbb{R}^d} \left| \mathrm{E}[h\left(\Phi_\beta(\boldsymbol{Q}(\boldsymbol{Y}) - y)\right)] - \mathrm{E}[h\left(\Phi_\beta(Z - y)\right)] \right|$$

$$\leq \frac{3}{2} \max\{\|h''\|_\infty, \beta\|h'\|_\infty\} \left(\delta_0[\boldsymbol{Q}(\boldsymbol{Y})] + C\delta_2[\boldsymbol{Q}(\boldsymbol{Y})]\right) \tag{5.1}$$

for any $\beta > 0$ and $h \in C_b^\infty(\mathbb{R})$, where $C > 0$ depends only on $\overline{q}_d$ and

$$\delta_2[\boldsymbol{Q}(\boldsymbol{Y})] := \max_{1 \leq j,k \leq d} \{\eta_N^{-1}(\log d)\}^{w_*(q_j + q_k) - 1} \left\{ 1_{\{q_j < q_k\}} \|Q(f_j; \boldsymbol{Y})\|_4 \kappa_4(Q(f_k; \boldsymbol{Y}))^{1/4} \right.$$

$$\left. + 1_{\{q_j = q_k\}} \sqrt{2\kappa_4(Q(f_j; \boldsymbol{Y})) + \left(2^{-q_j} \overline{v}_N^{q_j} - 1\right)(2q_j)! \mathfrak{c}_{q_j} \sum_{i=1}^N \mathrm{Inf}_i(f_j)^2} \right\}$$

with $\overline{v}_N := \max_{1 \le i \le N} v_i$ *and* $\underline{\eta}_N := \min_{1 \le i \le N} \eta_i$.

The rest of this section is devoted to the proof of Proposition 5.1. In the remainder of this section, we assume that the probability space $(\Omega, \mathcal{F}, P)$ is given by the product probability space $(\prod_{i=1}^N \Omega_i, \bigotimes_{i=1}^N \mathcal{F}_i, \bigotimes_{i=1}^N P_i)$, where

$$(\Omega_i, \mathcal{F}_i, P_i) := \begin{cases} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(0,1)) & \text{if } Y_i \sim \mathcal{N}(0,1), \\ ((0,\infty), \mathcal{B}((0,\infty)), \gamma(v)) & \text{if } Y_i \sim \gamma_\pm(v). \end{cases}$$

Then, we realize the variables $Y_1, \ldots, Y_N$ as follows: For $\omega = (\omega_1, \ldots, \omega_N) \in \Omega$, we define

$$Y_i(\omega) := \begin{cases} \omega_i & \text{if } Y_i \sim \mathcal{N}(0,1), \\ \pm(\omega_i - v)/\sqrt{v} & \text{if } Y_i \sim \gamma_\pm(v). \end{cases}$$

5.1 Γ-Calculus

Our first aim is to construct a suitable Markov diffusion operator whose eigenspaces contain all the components of $Q(Y)$. In the following, for an open subset U of $\mathbb{R}^m$, we write $C_p^\infty(U)$ for the set of all real-valued C^∞ functions on U all of whose partial derivatives have at most polynomial growth.

First, we denote by $\mathrm{L_{OU}}$ the Ornstein–Uhlenbeck operator on $\mathbb{R}$. Next, for every $v > 0$, we write L_v for the Laguerre operator on $(0, \infty)$ with parameter v. We then define the operators $\mathcal{L}_1, \ldots, \mathcal{L}_N$ by

$$\mathcal{L}_i := \begin{cases} \mathrm{L_{OU}} & \text{if } Y_i \sim \mathcal{N}(0,1), \\ \mathrm{L}_v & \text{if } Y_i \sim \gamma_\pm(v). \end{cases}$$

Finally, we construct the densely defined symmetric operator L in $L^2(P)$ by tensorization of $\mathcal{L}_1, \ldots, \mathcal{L}_N$ (see Section 2.2 of [2] for details). We will use the following properties of L (cf. [1] and Section 2.2 of [2]):

(i) If F and G are eigenfunctions of $-\mathrm{L}$ associated with eigenvalues p and q, respectively, FG belongs to $\bigoplus_{k=0}^{p+q} \mathrm{Ker}(\mathrm{L} + k\,\mathrm{Id})$.

(ii) The eigenspaces of $\mathrm{L_{OU}}$ and L_v associated with eigenvalue $k \in \mathbb{Z}_+$ are given by $\mathrm{Ker}(\mathrm{L_{OU}} + k\,\mathrm{Id}) = \{aH_k : a \in \mathbb{R}\}$ and $\mathrm{Ker}(\mathrm{L}_v + k\,\mathrm{Id}) = \{aL_k^{(v-1)} : a \in \mathbb{R}\}$, respectively. Here, H_k and $L_k^{(\alpha)}$ denote the Hermite polynomial of degree k and Laguerre polynomial of degree k and parameter $\alpha > -1$, respectively.

(iii) The eigenspace of L associated with eigenvalue k is given by

$$\mathrm{Ker}(\mathrm{L} + k\,\mathrm{Id}) = \bigoplus_{\substack{k_1 + \cdots + k_N = k \\ k_1, \ldots, k_N \in \mathbb{Z}_+}} \mathrm{Ker}(\mathcal{L}_1 + k_1\,\mathrm{Id}) \otimes \cdots \otimes \mathrm{Ker}(\mathcal{L}_N + k_N\,\mathrm{Id}).$$

$$(5.2)$$

Let us write $\mathcal{S} = C_p^\infty(\Omega)$. We define the *carré du champ operator* of L by

$$\Gamma(F, G) = \frac{1}{2}\left(L(FG) - F\,L\,G - G\,L\,F\right)$$

for all $F, G \in \mathcal{S}$. The following lemma is a special case of [39, Proposition 5.1].

Lemma 5.1 *For every $(i, j) \in [d]^2$, define the function $\tau^{ij} : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ by*

$$\tau^{ij}(x) = \frac{1}{q_j}E[\Gamma(Q(f_i; \mathbf{Y}), Q(f_j; \mathbf{Y})) \mid \mathbf{Q}(\mathbf{Y}) = x], \quad x \in \mathbb{R}^d.$$

Then, $\tau = (\tau^{ij})_{1 \le i, j \le d}$ is a Stein kernel for $\mathbf{Q}(\mathbf{Y})$.

We refer to [5] for more details about these operators.

5.2 A Bound for the Variance of the Carré du Champ Operator

In view of Lemmas 4.1 and 5.1 , we obtain (5.1) once we show that

$$E\left[\max_{1 \le j, k \le d}\left|\frac{1}{q_k}\Gamma\left(Q(f_j; \mathbf{Y}), Q(f_k; \mathbf{Y})\right) - \mathfrak{C}_{jk}\right|\right] \le \delta_0[\mathbf{Q}(\mathbf{Y})] + C\delta_2[\mathbf{Q}(\mathbf{Y})], (5.3)$$

where $C > 0$ depends only on $\overline{q}_d$. As a first step, we estimate $\mathrm{Var}[\Gamma(Q(f_j; \mathbf{Y}), Q(f_k; \mathbf{Y}))]$ for every $(j, k) \in [d]^2$. More precisely, our aim here is to prove the following result:

Proposition 5.2 *Let $p \le q$ be two positive integers. Let $f : [N]^p \to \mathbb{R}$ and $g : [N]^q \to \mathbb{R}$ be symmetric functions vanishing on diagonals and set $F := Q(f; \mathbf{Y})$ and $G := Q(g; \mathbf{Y})$. Then, $\kappa_4(F) \ge 0$, $\kappa_4(G) \ge 0$ and*

$$\mathrm{Var}\left[\frac{1}{q}\Gamma(F, G)\right]$$

$$\le 1_{\{p<q\}}\sqrt{E[F^4]}\sqrt{\kappa_4(G)}$$

$$+ 1_{\{p=q\}}\left\{2\sqrt{\kappa_4(F)}\sqrt{\kappa_4(G)} + \left(2^{-p}\overline{v}_N^p - 1\right)(2p)!\mathfrak{c}_p\sqrt{\sum_{i=1}^N \mathrm{Inf}_i(f)^2}\sqrt{\sum_{i=1}^N \mathrm{Inf}_i(g)^2}\right\}.$$

$$(5.4)$$

Before starting the proof, we remark how this result is related to the preceding studies. When $f = g$, Azmoodeh et al. [1] have derived a better estimate than (5.4) in a more general setting. Their technique of the proof can also be applied to the case $f \ne g$, and this has been implemented in Campese et al. [9]. However, this leads to a bound containing the quantity $\mathrm{Cov}[F^2, G^2] - 2\,E[FG]^2$, so we need an additional argument to estimate it. For this reason, we take an alternative route for the proof, which is inspired by the discussions in Zheng [59] as well as [9, Proposition 3.6]. As a

byproduct of this strategy, we obtain inequality (5.11) which leads to the universality of gamma variables.

We begin by introducing some notation. We write J_k for the orthogonal projection of $L^2(P)$ onto the eigenspace $\mathrm{Ker}(L + k\,\mathrm{Id})$. For every i, we define the random variable $\mathfrak{p}_2(Y_i)$ by

$$\mathfrak{p}_2(Y_i) := \begin{cases} H_2(Y_i) & \text{if } Y_i \sim \mathcal{N}(0,1), \\ \pm\frac{2}{v} L_2^{(v-1)}(\pm\sqrt{v}(Y_i + 1)) & \text{if } Y_i \sim \gamma_\pm(v). \end{cases}$$

The following lemma can be proved by a straightforward computation.

Lemma 5.2 *For every i, $\mathrm{E}[\mathfrak{p}_2(Y_i)] = \mathrm{E}[Y_i\mathfrak{p}_2(Y_i)] = 0$ and $\mathrm{E}[\mathfrak{p}_2(Y_i)^2] = v_i$.*

Next, given $h' : [N]^r \to \mathbb{R}$, we define

$$\langle h, h' \rangle := \sum_{i_1,\ldots,i_r=1}^N h(i_1,\ldots,i_r) h'(i_1,\ldots,i_r).$$

Note that $\|h\|_{\ell_2}^2 = \langle h, h \rangle$. For every $r \in \{0, 1, \ldots, p \wedge q\}$, we define the function $f \, \widehat{\star}_r^0 \, g : [N]^{p+q-r} \to \mathbb{R}$ by

$$f \, \widehat{\star}_r^0 \, g(i_1, \ldots, i_{p+q-2r}, k_1, \ldots, k_r)$$
$$:= \frac{1}{(p+q-2r)!} \sum_{\sigma \in \mathfrak{S}_{p+q-2r}} f(i_{\sigma(1)}, \ldots, i_{\sigma(p-r)}, k_1, \ldots, k_r)$$
$$\times g(i_{\sigma(p-r+1)}, \ldots, i_{\sigma(p+q-2r)}, k_1, \ldots, k_r).$$

Note that we have

$$\widetilde{f \star_r g}(i_1, \ldots, i_{p+q-2r}) = \sum_{(k_1,\ldots,k_r) \in \Delta_r^N} f \, \widehat{\star}_r^0 \, g(i_1, \ldots, i_{p+q-2r}, k_1, \ldots, k_r), \quad (5.5)$$

where $\widetilde{f \star_r g}$ is the symmetrization of $f \star_r g$ (recall (2.3)). Finally, we set $\Delta_q^N := \{(i_1, \ldots, i_q) \in [N]^q : i_j \neq i_k \text{ if } j \neq k\}$.

The next lemma is a key part in our proof.

Lemma 5.3 *Under the assumptions of Proposition 5.1, we have*

$$\mathrm{E}[J_{p+q}(FG)^2] \geq (p+q)! \| f \, \widetilde{\otimes} \, g \|_{\ell_2}^2. \quad (5.6)$$

Moreover, if $p = q$, we have

$$\left| \mathrm{E}[J_{2p}(F^2) J_{2p}(G^2)] - (2p)! \langle f \, \widetilde{\otimes} \, f, g \, \widetilde{\otimes} \, g \rangle \right|$$
$$\leq \left(2^{-p} \overline{v}_N^p - 1\right) (2p)! \mathfrak{c}_p \sqrt{\sum_{i=1}^N \mathrm{Inf}_i(f)^2} \sqrt{\sum_{i=1}^N \mathrm{Inf}_i(g)^2}. \quad (5.7)$$

Proof We can deduce from [48, Proposition 2.9] and (5.5)

$$
J_{p+q}(FG) = \sum_{r=0}^{p\wedge q} r! \binom{p}{r}\binom{q}{r} \sum_{(i_1,\ldots,i_{p+q-r})\in\Delta_{p+q-r}^N} f\,\widehat{\star}_r^0\, g(i_1,\ldots,i_{p+q-r}) Y_{i_1}
$$
$$
\cdots Y_{i_{p+q-2r}}\,\mathfrak{p}_2(Y_{i_{p+q-2r+1}})\cdots \mathfrak{p}_2(Y_{i_{p+q-r}}). \tag{5.8}
$$

Next, let $(i_1,\ldots,i_{p+q-r}) \in \Delta_{p+q-r}^N$ and $(j_1,\ldots,j_{p+q-l}) \in \Delta_{p+q-l}^N$. Thanks to Lemma 5.2,

$$
\mathrm{E}[Y_{i_1}\cdots Y_{i_{p+q-2r}}\,\mathfrak{p}_2(Y_{i_{p+q-2r+1}})\cdots\mathfrak{p}_2(Y_{i_{p+q-r}})Y_{j_1}
$$
$$
\cdots Y_{j_{p+q-2l}}\,\mathfrak{p}_2(Y_{j_{p+q-2l+1}})\cdots\mathfrak{p}_2(Y_{j_{p+q-l}})]
$$

does not vanish if and only if the following condition is satisfied:

($\star$) $(i_1,\ldots,i_{p+q-2r})$ is a permutation of $(j_1,\ldots,j_{p+q-2l})$ and $(i_{p+q-2r+1},\ldots,$ $i_{p+q-r})$ is a permutation of $(j_{p+q-2l+1},\ldots,j_{p+q-l})$.

Note that the condition ($\star$) can hold true only if $r = l$. Moreover, if the condition ($\star$) is satisfied, we have

$$
\mathrm{E}[Y_{i_1}\cdots Y_{i_{p+q-2r}}\,\mathfrak{p}_2(Y_{i_{p+q-2r+1}})\cdots\mathfrak{p}_2(Y_{i_{p+q-r}})Y_{j_1}
$$
$$
\cdots Y_{j_{p+q-2l}}\,\mathfrak{p}_2(Y_{j_{p+q-2l+1}})\cdots\mathfrak{p}_2(Y_{j_{p+q-l}})] = v_{i_{p+q-2r+1}}\cdots v_{i_{p+q-r}}
$$

by Lemma 5.2. Since there are totally $(p+q-2r)!$ permutations of $(i_1,\ldots,i_{p+q-2r})$ and $r!$ permutations of $(i_{p+q-2r+1},\ldots,i_{p+q-r})$, respectively, (5.8) yields

$$
\mathrm{E}[J_{p+q}(FG)^2]
$$
$$
= \sum_{r=0}^{p\wedge q} r!^2 \binom{p}{r}^2\binom{q}{r}^2 (p+q-2r)!\,r!
$$
$$
\times \sum_{(i_1,\ldots,i_{p+q-r})\in\Delta_{p+q-r}^N} f\,\widehat{\star}_r^0\, g(i_1,\ldots,i_{p+q-r})^2 v_{i_{p+q-2r+1}}\cdots v_{i_{p+q-r}}. \tag{5.9}
$$

Now, (5.9) is especially true when all the Y_i's follow the standard normal distribution. Therefore, the product formula for multiple integrals with respect to an isonormal Gaussian process yields

$$
(p+q)!\|f\,\widetilde{\otimes}\, g\|_{\ell_2}^2 = \sum_{r=0}^{p\wedge q} r!^2 \binom{p}{r}^2\binom{q}{r}^2 (p+q-2r)!\,r!
$$
$$
\times \sum_{(i_1,\ldots,i_{p+q-r})\in\Delta_{p+q-r}^N} f\,\widehat{\star}_r^0\, g(i_1,\ldots,i_{p+q-r})^2 \cdot 2^r,
$$

where $f\,\widetilde{\otimes}\, g$ is the symmetrization of $f\otimes g$. Combining this formula with (5.9), we obtain (5.6). $\qquad\square$

Next, we prove (5.7). Similar arguments to the above yield

$$\left| \mathrm{E}[J_{2p}(F^2)J_{2p}(G^2)] - (2p)! \langle f \widetilde{\otimes} f, g \widetilde{\otimes} g \rangle \right|$$

$$\leq \left(2^{-p}\overline{v}_N^p - 1\right) \sum_{r=1}^{p} r!^2 \binom{p}{r}^4 (2p-2r)! r!$$

$$\times \sum_{(i_1,\ldots,i_{2p-r}) \in \Delta_{2p-r}^N} \left| f \widehat{\star_r^0} f(i_1,\ldots,i_{2p-r}) g \widehat{\star_r^0} g(i_1,\ldots,i_{2p-r}) \right| \cdot 2^r$$

and

$$\sum_{r=1}^{p} r!^2 \binom{p}{r}^4 (2p-2r)! r! \sum_{(i_1,\ldots,i_{2p-r}) \in \Delta_{2p-r}^N} f \widehat{\star_r^0} f(i_1,$$

$$\ldots, i_{2p-r}) g \widehat{\star_r^0} g(i_1,\ldots,i_{2p-r}) \cdot 2^r$$

$$= (2p)! \left\langle f \widetilde{\otimes} f, g \widetilde{\otimes} g 1_{(\Delta_{2p}^N)^c} \right\rangle.$$

Combining these results with the Schwarz inequality, we infer that

$$\left| \mathrm{E}[J_{2p}(F^2)J_{2p}(G^2)] - (2p)! \langle f \widetilde{\otimes} f, g \widetilde{\otimes} g \rangle \right|$$

$$\leq \left(2^{-p}\overline{v}_N^p - 1\right) \sqrt{(2p)! \| f \widetilde{\otimes} f 1_{(\Delta_{2p}^N)^c} \|_{\ell_2}^2} \sqrt{(2p)! \| g \widetilde{\otimes} g 1_{(\Delta_{2p}^N)^c} \|_{\ell_2}^2}.$$

The desired result now follows from Eq.(50) in [27] and Hölder's inequality. ☐

Lemma 5.4 *Under the assumptions of Proposition 5.1, we have*

$$\sum_{k=1}^{p+q-1} \mathrm{E}[J_k(FG)^2] \leq \mathrm{Cov}[F^2, G^2] - 2\mathrm{E}[FG]^2 \tag{5.10}$$

and

$$\sum_{k=1}^{2p-1} \mathrm{E}[J_k(F^2)^2] + p!^2 \sum_{r=1}^{p-1} \binom{p}{r}^2 \| f \star_r f \|_{\ell_2}^2 \leq \mathrm{E}[F^4] - 3\mathrm{E}[F^2]^2. \tag{5.11}$$

Proof The proof is parallel to that of [59, Lemma 3.1] but using Lemma 5.3 instead of [59, Lemma 2.1]. ☐

Proof of Proposition 5.2 The nonnegativity of $\kappa_4(F)$ and $\kappa_4(G)$ follows from (5.11). (5.4) can be shown in a similar manner to the proof of [59, Theorem 1.1] but using Lemmas 5.3 and 5.4 instead of Lemmas 2.1 and 3.1 in [59], respectively. ☐

5.3 Proof of Proposition 5.1

We have already established the nonnegativity of $\kappa_4(Q(f_j; Y))$'s in Proposition 5.2. The remaining claim of the proposition follows once we prove (5.3). By Proposition 5.2 and Lemma A.2, this follows once we show that, under the assumptions of Proposition 5.1,

$$\|\Gamma(F, G) - \mathrm{E}[\Gamma(F, G)]\|_{\psi_{(w_*(p+q)-1)^{-1}}} \leq C_{p,q} \underline{\eta}_N^{-(w_*(p+q)-1)} \sqrt{\mathrm{Var}[\Gamma(F, G)]}, \tag{5.12}$$

where $C_{p,q} > 0$ depends only on p, q. To prove (5.12), note that

$$\Gamma(F, G) = \frac{1}{2} \left(\mathrm{L}(FG) + qFG + pGF \right)$$

$$= \frac{p+q}{2} \mathrm{E}[FG] + \sum_{k=1}^{p+q-1} \frac{p+q-k}{2} J_k(FG). \tag{5.13}$$

Hence, using Lemma A.5, we can deduce (5.12) by a hypercontractivity argument similar to those in [33, Section 5] and [41, Section 3.2]. □

6 Randomized Lindeberg Method

For any $\varpi \geq 0$ and $x \geq 0$, we set

$$\chi_\varpi(x) = \begin{cases} \exp(-x^{1/\varpi}) & \text{if } \varpi > 0, \\ 1_{[0,1)}(x) & \text{if } \varpi = 0. \end{cases}$$

The aim of this section is to prove the following result.

Proposition 6.1 *Set* $\Lambda_i := (\log d)^{(\overline{q}_d-1)/\alpha} \max_{1 \leq k \leq d} M_N^{q_k-1} \sqrt{\mathrm{Inf}_i(f_k)}$ *for* $i \in [N]$. *Let* $X = (X_i)_{i=1}^N$ *and* $Y = (Y_i)_{i=1}^N$ *be two sequences of independent centered random variables with unit variance. Suppose that* $M_N := \max_{1 \leq i \leq N}(\|X_i\|_{\psi_\alpha} \vee \|Y_i\|_{\psi_\alpha}) < \infty$ *for some* $\alpha \in (0, 2]$. *Suppose also that there is an integer* $m \geq 3$ *such that* $\mathrm{E}[X_i^r] = \mathrm{E}[Y_i^r]$ *for all* $i \in [N]$ *and* $r \in [m-1]$. *Then, for any* $h \in C_b^m(\mathbb{R})$, $\beta > 0$ *and* $\tau, \rho \geq 0$ *with* $\tau \rho M_N \max_{1 \leq i \leq N} \Lambda_i \leq \beta^{-1}$, *we have*

$$\sup_{y \in \mathbb{R}^d} \left| \mathrm{E}\left[h\left(\Phi_\beta(Q(X) - y) \right) \right] - \mathrm{E}\left[h\left(\Phi_\beta(Q(Y) - y) \right) \right] \right|$$

$$\leq C \left(\max_{1 \leq j \leq m} \beta^{m-j} \|h^{(j)}\|_\infty \right) \left\{ (\log d)^{m(\overline{q}_d-1)/\alpha} \max_{1 \leq k \leq d} M_N^{mq_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^{m/2} \right.$$

$$\left. + \left(e^{-\left(\frac{\tau}{K_1}\right)^\alpha} + \chi_{(\overline{q}_d-1)/\alpha}\left(\frac{\rho}{K_2}\right) \right) (\rho \vee 1)^m + \exp\left(-\left(\frac{\tau\rho}{K_3}\right)^{\alpha/\overline{q}_d} \right) (\tau\rho \vee 1)^m \right)$$

$$\times M_N^m \sum_{i=1}^{N} \Lambda_i^m \Bigg\} , \tag{6.1}$$

where $C > 0$ depends only on $m, \alpha, \overline{q}_d$, K_1 depends only on α, and $K_2, K_3 > 0$ depend only on $\alpha, \overline{q}_d$.

Remark 6.1 Proposition 6.1 can be viewed as a version of [47, Theorem 7.1]. Apart from that we take account of higher moment matching, there are important differences between these two results. On the one hand, the latter takes all C^3 functions with bounded third-order partial derivatives as test functions, while the former focuses only on test functions of the form $x \mapsto h(\Phi_\beta(x - y))$ for some $h \in C_b^m(\mathbb{R})$ and $y \in \mathbb{R}^d$. On the other hand, in the bound of (6.1), terms like $\sum_{i=1}^{N} \max_{1 \le k \le d} \mathrm{Inf}_i(f_k)$ always appear with exponential factors, so we can remove such terms by appropriately selecting the parameters τ, ρ. In contrast, such a quantity appears (as the constant C) in the dominant term of the bound given by [47, Theorem 7.1]. As pointed out in Remark 2.4(b), this can be crucial in a high-dimensional setting, and this phenomenon originates from a (naïve) application of the Lindeberg method. To avoid this difficulty, we use a randomized version of the Lindeberg method, which was originally introduced in [24] for sums of independent random vectors.

For the proof, we need three auxiliary results. The first one is a generalization of [36, Lemma S.5.1]:

Lemma 6.1 Let ξ be a nonnegative random variable such that $P(\xi > x) \le Ae^{-(x/B)^\alpha}$ for all $x \ge 0$ and some constants $A, B, \alpha > 0$. Then, we have

$$\mathrm{E}\left[\xi^p 1_{\{\xi > t\}}\right] \le A\left(1 + \frac{2p - \alpha}{p - \alpha}\right)\left(t \vee \{(2(p/\alpha - 1))^{1/\alpha}B\}\right)^p e^{-(t/B)^\alpha}$$

for any $p > \alpha$ and $t > 0$.

Proof The proof is analogous to that of [36, Lemma S.5.1] and elementary, so we omit it.   □

The second one is a moment inequality for homogeneous sums with a sharp constant:

Lemma 6.2 Let $X = (X_i)_{i=1}^{N}$ be a sequence of independent centered random variables. Suppose that $M := \max_{1 \le i \le N} \|X_i\|_{\psi_\alpha} < \infty$ for some $\alpha \in (0, 2]$. Also, let $q \in \mathbb{N}$ and $f : [N]^q \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Then,

$$\|Q(f; X)\|_p \le K_{\alpha,q} p^{q/\alpha} M^q \|f\|_{\ell_2}$$

for any $p \ge 2$, where $K_{\alpha,q} > 0$ depends only on α, q.

Since we need additional lemmas to prove Lemma 6.2, we postpone its proof to Appendix B.

The third one is well known and immediately follows from the commutativity of addition, but it will deserve to be explicitly stated for later reference.

Lemma 6.3 *Let S be a finite set and φ be a real-valued function on S. Also, let $b :$ $S \to S$ be a bijection. Then, $\sum_{x \in A} \varphi(b(x)) = \sum_{x \in b(A)} \varphi(x)$ for any $A \subset S$.*

Now we turn to the main body of the proof. Throughout the proof, we will use the standard multi-index notation. For a multi-index $\lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{Z}_+^d$, we set $|\lambda| := \lambda_1 + \cdots + \lambda_d$, $\lambda! := \lambda_1! \cdots \lambda_d!$ and $\partial^\lambda := \partial_1^{\lambda_1} \cdots \partial_d^{\lambda_d}$ as usual. Also, given a vector $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we write $x^\lambda = x_1^{\lambda_1} \cdots x_d^{\lambda_d}$.

Proof of Proposition 6.1 Without loss of generality, we may assume X and Y are independent. Throughout the proof, for two real numbers a and b, the notation $a \lesssim b$ means that $a \leq cb$ for some constant $c > 0$ which depends only on $m, \alpha, \overline{q}_d$.

Take a vector $y \in \mathbb{R}^d$ and define the function $\Psi : \mathbb{R}^d \to \mathbb{R}$ by $\Psi(x) = h(\Phi_\beta(x-y))$ for $x \in \mathbb{R}^d$. For any $i \in [N], \sigma \in \mathfrak{S}_N$ and $k \in [d]$, we define

$$W_i^\sigma = (W_{i,1}^\sigma, \ldots, W_{i,N}^\sigma) := (X_{\sigma(1)}, \ldots, X_{\sigma(i)}, Y_{\sigma(i+1)}, \ldots, Y_{\sigma(N)})$$

and

$$U_{k,i}^\sigma := \sum_{\substack{i_1,\ldots,i_{q_k}=1 \\ i_1 \neq i,\ldots,i_{q_k} \neq i}}^N f_k(\sigma(i_1), \ldots, \sigma(i_{q_k})) W_{i,i_1}^\sigma \cdots W_{i,i_{q_k}}^\sigma,$$

$$V_{k,i}^\sigma := \sum_{\substack{i_1,\ldots,i_{q_k}=1 \\ \exists j : i_j = i}}^N f_k(\sigma(i_1), \ldots, \sigma(i_{q_k})) \prod_{l:i_l \neq i} W_{i,i_l}^\sigma.$$

Then, we set $U_i^\sigma = (U_{k,i}^\sigma)_{k=1}^d$ and $V_i^\sigma = (V_{k,i}^\sigma)_{k=1}^d$. By construction, U_i^σ and V_i^σ are independent of $X_{\sigma(i)}$ and $Y_{\sigma(i)}$. Moreover, we have $Q(f_k; W_{i-1}^\sigma) = U_{k,i}^\sigma + Y_{\sigma(i)} V_{k,i}^\sigma$ and $Q(f_k; W_i^\sigma) = U_{k,i}^\sigma + X_{\sigma(i)} V_{k,i}^\sigma$ (with $W_0^\sigma := (Y_{\sigma(1)}, \ldots, Y_{\sigma(N)})$). In particular, by Lemma 6.3 it holds that $Q(f_k; W_0^\sigma) = Q(f_k; Y)$ and $Q(f_k; W_N^\sigma) = Q(f_k; X)$. Therefore, we obtain

$$\left| \mathbb{E}\left[\Psi(Q(X))\right] - \mathbb{E}\left[\Psi(Q(Y))\right] \right|$$
$$= \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \left| \mathbb{E}\left[\Psi(Q(W_N^\sigma))\right] - \mathbb{E}\left[\Psi(Q(W_0^\sigma))\right] \right|$$
$$\leq \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \sum_{i=1}^N \left| \mathbb{E}\left[\Psi(Q(W_i^\sigma))\right] - \mathbb{E}\left[\Psi(Q(W_{i-1}^\sigma))\right] \right|. \qquad (6.2)$$

Taylor's theorem and the independence of $X_{\sigma(i)}$ and $Y_{\sigma(i)}$ from U_i^σ and V_i^σ yield

$$\mathbb{E}\left[\Psi(U_i^\sigma + \xi V_i^\sigma)\right] = \sum_{\lambda \in \mathbb{Z}_+^d : |\lambda| \leq m-1} \frac{1}{\lambda!} \mathbb{E}\left[\partial^\lambda \Psi(U_i^\sigma)\left(V_i^\sigma\right)^\lambda\right] \mathbb{E}\left[\xi^{|\lambda|}\right] + R_i^\sigma[\xi]$$

for $\xi \in \{X_{\sigma(i)}, Y_{\sigma(i)}\}$, where

$$R_i^\sigma[\xi] := \sum_{\lambda \in \mathbb{Z}_+^d : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1} \, \mathrm{E}\left[\partial^\lambda \Psi(U_i^\sigma + t\xi V_i^\sigma)\xi^m (V_i^\sigma)^\lambda\right] dt.$$

Since $\mathrm{E}[X_i^r] = \mathrm{E}[Y_i^r]$ for all $i \in [N]$ and $r \in [m-1]$ by assumption, we obtain

$$\left|\mathrm{E}\left[\Psi\left(\boldsymbol{Q}_i^\sigma\right)\right] - \mathrm{E}\left[\Psi(\boldsymbol{Q}_{i-1}^\sigma)\right]\right| \le |R_i^\sigma[X_{\sigma(i)}]| + |R_i^\sigma[Y_{\sigma(i)}]| \le \mathbf{I}_i^\sigma + \mathbf{II}_i^\sigma, \quad (6.3)$$

where $\mathbf{I}_i^\sigma := \mathbf{I}_i^\sigma[X_{\sigma(i)}] + \mathbf{I}_i^\sigma[Y_{\sigma(i)}]$, $\mathbf{II}_i^\sigma := \mathbf{II}_i^\sigma[X_{\sigma(i)}] + \mathbf{II}_i^\sigma[Y_{\sigma(i)}]$ and

$$\mathbf{I}_i^\sigma[\xi] := \sum_{\lambda \in \mathbb{Z}_+^d : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1} \, \mathrm{E}\left[|\partial^\lambda \Psi(U_i^\sigma + t\xi V_i^\sigma)||\xi|^m |(V_i^\sigma)^\lambda|; \mathcal{E}_{\sigma,i}\right] dt,$$

$$\mathbf{II}_i^\sigma[\xi] := \sum_{\lambda \in \mathbb{Z}_+^d : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1} \, \mathrm{E}\left[|\partial^\lambda \Psi(U_i^\sigma + t\xi V_i^\sigma)||\xi|^m |(V_i^\sigma)^\lambda|; \mathcal{E}_{\sigma,i}^c\right] dt$$

for $\xi \in \{X_{\sigma(i)}, Y_{\sigma(i)}\}$ and $\mathcal{E}_{\sigma,i} := \{(|X_{\sigma(i)}| + |Y_{\sigma(i)}|)\|V_i^\sigma\|_{\ell_\infty} \le \tau \rho M_N \Lambda_{\sigma(i)}\}$.

First, we consider $\mathbf{I}_i^\sigma$. Since $\tau \rho M_N \max_{1 \le i \le N} \Lambda_i \le \beta^{-1}$ by assumption, Lemma 3.1 and the independence of $X_{\sigma(i)}, Y_{\sigma(i)}$ from U_i^σ, V_i^σ imply that

$$\mathbf{I}_i^\sigma \le \frac{e^8}{m!} \sum_{j_1, \ldots, j_m = 1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1, \ldots, j_m}(U_i^\sigma - y)(|X_{\sigma(i)}|^m + |Y_{\sigma(i)}|^m)|V_{j_1, i}^\sigma|^m\right]$$

$$\le \frac{e^8}{m!} \sum_{j_1, \ldots, j_m = 1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1, \ldots, j_m}(U_i^\sigma - y)|V_{j_1, i}^\sigma|^m\right] \mathrm{E}\left[|X_{\sigma(i)}|^m + |Y_{\sigma(i)}|^m\right]$$

$$\le \frac{e^8}{m!} \sup_{1 \le i \le N} \mathrm{E}[|X_i|^m + |Y_i|^m] \left\{\mathbf{I}(1)_i^\sigma + \mathbf{I}(2)_i^\sigma + \mathbf{I}(3)_i^\sigma\right\}, \quad (6.4)$$

where

$$\mathbf{I}(1)_i^\sigma := \sum_{j_1, \ldots, j_m = 1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1, \ldots, j_m}(U_i^\sigma - y)|V_{j_1, i}^\sigma|^m; \mathcal{C}_{\sigma,i} \cap \mathcal{D}_{\sigma,i}\right],$$

$$\mathbf{I}(2)_i^\sigma := \sum_{j_1, \ldots, j_m = 1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1, \ldots, j_m}(U_i^\sigma - y)|V_{j_1, i}^\sigma|^m; \mathcal{C}_{\sigma,i}^c\right],$$

$$\mathbf{I}(3)_i^\sigma := \sum_{j_1, \ldots, j_m = 1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1, \ldots, j_m}(U_i^\sigma - y)|V_{j_1, i}^\sigma|^m; \mathcal{D}_{\sigma,i}^c\right]$$

and $\mathcal{C}_{\sigma,i} := \{|X_{\sigma(i)}| + |Y_{\sigma(i)}| \le \tau M_N\}$, $\mathcal{D}_{\sigma,i} := \{\|V_i^\sigma\|_{\ell_\infty} \le \rho \Lambda_{\sigma(i)}\}$.

We begin by estimating $\mathbf{I}(1)_i^\sigma$. Let $(\delta_i)_{i=1}^N$ be a sequence of i.i.d. Bernoulli variables independent of X and Y with $P(\delta_i = 1) = 1 - P(\delta_i = 0) = i/(N+1)$. We set $\zeta_{i,a} := \delta_i X_a + (1 - \delta_i) Y_a$ for all $i, a \in [N]$. Then, since $\|\zeta_{i,\sigma(i)} V_i^\sigma\|_{\ell_\infty} \le \tau \rho M_N \max_{1 \le i \le N} \Lambda_i \le \beta^{-1}$ on the set $\mathcal{C}_{\sigma,i} \cap \mathcal{D}_{\sigma,i}$, by Lemma 3.1 we obtain

$$\mathbf{I}(1)_i^\sigma \le e^8 \sum_{j_1,\dots,j_m=1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}(U_i^\sigma + \zeta_{i,\sigma(i)} V_i^\sigma - y) |V_{j_1,i}^\sigma|^m \right].$$

The subsequent discussions are inspired by the proof of [24, Lemma 2] and we introduce some notation analogous to theirs. For any $i, a \in [N]$, we set

$$\mathcal{A}_{i,a} = \{(A, B) : A \subset [N], B \subset [N], A \cup B = [N] \setminus \{a\}, \#A = i - 1, \#B = N - i\},$$

where $\#S$ denotes the number of elements in a set S. We also set

$$\mathcal{A}_i = \{(A, B) : A \subset [N], B \subset [N], A \cup B = [N], \#A = i, \#B = N - i\}$$

for every $i \in \{0, 1 \dots, N\}$. Moreover, for any $A, B \subset [N]$ with $A \cap B = \emptyset$ and $i \in A \cup B$, we define the random variable $W_i^{(A,B)}$ by $W_i^{(A,B)} := X_i$ if $i \in A$ and $W_i^{(A,B)} := Y_i$ if $i \in B$. Then, we define

$$Q_k^{(A,B)} := \sum_{i_1,\dots,i_{q_k}=1}^N f_k(i_1, \dots, i_{q_k}) W_{i_1}^{(A,B)} \cdots W_{i_{q_k}}^{(A,B)}$$

for any $k \in [d]$ and $(A, B) \in \bigcup_{i=0}^\infty \mathcal{A}_i$, and set $\boldsymbol{Q}^{(A,B)} := (Q_k^{(A,B)})_{k=1}^d$. We also define

$$U_{k,a}^{(A,B)} := \sum_{\substack{i_1,\dots,i_{q_k}=1 \\ i_1 \ne a, \dots, i_{q_k} \ne a}}^N f_k(i_1, \dots, i_{q_k}) W_{i_1}^{(A,B)} \cdots W_{i_{q_k}}^{(A,B)},$$

$$V_{k,a}^{(A,B)} := \sum_{\substack{i_1,\dots,i_{q_k}=1 \\ \exists j : i_j = a}}^N f_k(i_1, \dots, i_{q_k}) \prod_{l:i_l \ne a} W_{i_l}^{(A,B)}$$

for any $k \in [d]$, $i, a \in [N]$ and $(A, B) \in \bigcup_{j=1}^N \mathcal{A}_{j,a}$ and set $U_a^{(A,B)} := (U_{k,a}^{(A,B)})_{k=1}^d$ and $V_a^{(A,B)} := (V_{k,a}^{(A,B)})_{k=1}^d$. Finally, for any $\sigma \in \mathfrak{S}_N$ and $i \in [N]$ we set $A_i^\sigma := \{\sigma(1), \dots, \sigma(i-1)\}$ and $B_i^\sigma := \{\sigma(i+1), \dots, \sigma(N)\}$.

Now, since we have $W_{i,j}^\sigma = W_{\sigma(j)}^{(A_i^\sigma, B_i^\sigma)}$ for $j \in [N]\setminus\{i\}$, it holds that $U_i^\sigma = U_{\sigma(i)}^{(A_i^\sigma, B_i^\sigma)}$ and $V_i^\sigma = V_{\sigma(i)}^{(A_i^\sigma, B_i^\sigma)}$ by Lemma 6.3. Therefore, we obtain

$$
\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \mathbf{I}(1)_i^\sigma
$$

$$
\leq \frac{e^8}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \sum_{j_1,\ldots,j_m=1}^{d}
$$

$$
\mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(U_{\sigma(i)}^{(A_i^\sigma, B_i^\sigma)} + \zeta_{i,\sigma(i)} V_{\sigma(i)}^{(A_i^\sigma, B_i^\sigma)} - y \right) \left| V_{j_1,\sigma(i)}^{(A_i^\sigma, B_i^\sigma)} \right|^m \right]
$$

$$
= \frac{e^8}{N!} \sum_{i=1}^{N} \sum_{a=1}^{N} \sum_{\sigma \in \mathfrak{S}_N : \sigma(i)=a} \sum_{j_1,\ldots,j_m=1}^{d} \mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(U_a^{(A_i^\sigma, B_i^\sigma)} \right. \right.
$$

$$
\left. \left. + \zeta_{i,a} V_a^{(A_i^\sigma, B_i^\sigma)} - y \right) \left| V_{j_1,a}^{(A_i^\sigma, B_i^\sigma)} \right|^m \right]
$$

$$
= \frac{e^8}{N!} \sum_{i=1}^{N} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \#\{\sigma \in \mathfrak{S}_N : A_i^\sigma = A, \sigma(i) = a\}
$$

$$
\times \sum_{j_1,\ldots,j_m=1}^{d} \mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(U_a^{(A,B)} + \zeta_{i,a} V_a^{(A,B)} - y \right) \left| V_{j_1,a}^{(A,B)} \right|^m \right]
$$

$$
= e^8 \sum_{i=1}^{N} \frac{(i-1)!(N-i)!}{N!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \sum_{j_1,\ldots,j_m=1}^{d}
$$

$$
\mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(U_a^{(A,B)} + \zeta_{i,a} V_a^{(A,B)} - y \right) \left| V_{j_1,a}^{(A,B)} \right|^m \right].
$$

Now, for $(A, B) \in \mathcal{A}_{i,a}$ we have $U_a^{(A,B)} + X_a V_a^{(A,B)} = Q^{(A\cup\{a\}, B)}$ and $U_a^{(A,B)} + Y_a V_a^{(A,B)} = Q^{(A, B\cup\{a\})}$, so we obtain

$$
\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \mathbf{I}(1)_i^\sigma
$$

$$
\leq e^8 \sum_{i=1}^{N} \frac{(i-1)!(N-i)!}{N!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \sum_{j_1,\ldots,j_m=1}^{d}
$$

$$
\left\{ \frac{i}{N+1} \mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(Q^{(A\cup\{a\}, B)} - y \right) \left| V_{j_1,a}^{(A,B)} \right|^m \right] \right.
$$

$$
\left. + \frac{N+1-i}{N+1} \mathrm{E}\left[\Upsilon_\beta^{j_1,\ldots,j_m} \left(Q^{(A, B\cup\{a\})} - y \right) \left| V_{j_1,a}^{(A,B)} \right|^m \right] \right\}
$$

$$= e^8 \sum_{i=1}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A\cup\{a\},B)} - y\right)\left|V_{j_1,a}^{(A,B)}\right|^m\right]$$

$$+ e^8 \sum_{i=1}^{N} \frac{(i-1)!(N+1-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B\cup\{a\})} - y\right)\left|V_{j_1,a}^{(A,B)}\right|^m\right]$$

$$= e^8 \sum_{i=1}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i,a}} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A\cup\{a\},B)} - y\right)\left|V_{j_1,a}^{(A,B)}\right|^m\right]$$

$$+ e^8 \sum_{i=0}^{N-1} \frac{i!(N-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_{i+1,a}} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B\cup\{a\})} - y\right)\left|V_{j_1,a}^{(A,B)}\right|^m\right]$$

$$= e^8 \sum_{i=1}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_i:a\in A} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B)} - y\right)\left|V_{j_1,a}^{(A\setminus\{a\},B)}\right|^m\right]$$

$$+ e^8 \sum_{i=0}^{N-1} \frac{i!(N-i)!}{(N+1)!} \sum_{a=1}^{N} \sum_{(A,B)\in\mathcal{A}_i:a\in B} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B)} - y\right)\left|V_{j_1,a}^{(A,B\setminus\{a\})}\right|^m\right]$$

$$= e^8 \sum_{i=0}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{(A,B)\in\mathcal{A}_i} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B)} - y\right) \sum_{a=1}^{N}\left|V_{j_1,a}^{(A\setminus\{a\},B\setminus\{a\})}\right|^m\right]$$

$$\leq e^8 \sum_{i=0}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{(A,B)\in\mathcal{A}_i} \sum_{j_1,\dots,j_m=1}^{d}$$

$$\mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}\left(\boldsymbol{Q}^{(A,B)} - y\right) \max_{1\leq k\leq d} \sum_{a=1}^{N}\left|V_{k,a}^{(A\setminus\{a\},B\setminus\{a\})}\right|^m\right].$$

Hence, Lemma 3.1 yields

$$\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \mathbf{I}(1)_i^{\sigma} \lesssim \max_{1 \leq j \leq m} \beta^{m-j} \|h^{(j)}\|_\infty \sum_{i=0}^{N} \frac{i!(N-i)!}{(N+1)!} \sum_{(A,B) \in \mathcal{A}_i}$$
$$\mathrm{E}\left[\max_{1 \leq k \leq d} \sum_{a=1}^{N} \left| V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})} \right|^m \right].$$

Now, by Lemma 6.2 we have

$$\left\| V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})} \right\|_r \leq C_{\alpha, \overline{q}_d} r^{(q_k - 1)/\alpha} M_N^{q_k - 1} \sqrt{\mathrm{Inf}_a(f_k)} \qquad (6.5)$$

for any $r \geq 1$, where $C_{\alpha, \overline{q}_d} > 0$ depends only on $\alpha, \overline{q}_d$. Hence, the Minkowski inequality yields

$$\left\| \sum_{a=1}^{N} \left| V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})} \right|^m \right\|_r \leq C_{\alpha, \overline{q}_d} (mr)^{m(q_k - 1)/\alpha} M_N^{m(q_k - 1)} \sum_{a=1}^{N} \mathrm{Inf}_a(f_k)^{m/2}. \quad (6.6)$$

Thus, if $\overline{q}_d > 1$, Lemma A.5 yields

$$\left\| \sum_{a=1}^{N} \left| V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})} \right|^m \right\|_{\psi_{\alpha/\{m(\overline{q}_d - 1)\}}}$$
$$\lesssim M_N^{m(q_k - 1)} \sum_{a=1}^{N} \mathrm{Inf}_a(f_k)^{m/2}.$$

Therefore, by Lemmas A.2 and A.6 we conclude that

$$\mathrm{E}\left[\max_{1 \leq k \leq d} \sum_{a=1}^{N} \left| V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})} \right|^m \right] \lesssim (\log d)^{m(\overline{q}_d - 1)/\alpha} \max_{1 \leq k \leq d} M_N^{m(q_k - 1)} \sum_{a=1}^{N} \mathrm{Inf}_a(f_k)^{m/2}.$$

This inequality also holds true when $\overline{q}_d = 1$ because in this case $V_{k,a}^{(A\setminus\{a\}, B\setminus\{a\})}$'s are non-random and thus it is a direct consequence of (6.6). As a result, we obtain

$$\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \mathbf{I}(1)_i^{\sigma} \lesssim \left(\max_{1 \leq j \leq m} \beta^{m-j} \|h^{(j)}\|_\infty \right)$$
$$\times \left((\log d)^{m(\overline{q}_d - 1)/\alpha} \max_{1 \leq k \leq d} M_N^{m(q_k - 1)} \sum_{a=1}^{N} \mathrm{Inf}_a(f_k)^{m/2} \right).$$
$$(6.7)$$

Next, we estimate $\mathbf{I}(2)_i^\sigma$. Since $X_{\sigma(i)}$ and $Y_{\sigma(i)}$ are independent of U_i^σ and V_i^σ,

$$\mathbf{I}(2)_i^\sigma \le \sum_{j_1,\dots,j_m=1}^d \mathrm{E}\left[\Upsilon_\beta^{j_1,\dots,j_m}(U_i^\sigma - y)\|V_i^\sigma\|_{\ell_\infty}^m\right] P\left(\mathcal{C}_{\sigma,i}^c\right).$$

Hence, Lemma 3.1 yields

$$\mathbf{I}(2)_i^\sigma \lesssim \left(\max_{1 \le j \le m} \beta^{m-j}\|h^{(j)}\|_\infty\right) \mathrm{E}\left[\|V_i^\sigma\|_{\ell_\infty}^m\right] P\left(\mathcal{C}_{\sigma,i}^c\right).$$

Now, if $\overline{q}_d > 1$, (6.5) and Lemma A.5 yield

$$\|V_{k,i}^\sigma\|_{\psi_{\alpha/(\overline{q}_d-1)}} \le c_{\alpha,\overline{q}_d} M_N^{q_k-1}\sqrt{\mathrm{Inf}_{\sigma(i)}(f_k)},$$

where $c_{\alpha,\overline{q}_d} > 0$ depends only on $\alpha, \overline{q}_d$. Hence, Lemmas A.2 and A.6 yield

$$\|\|V_i^\sigma\|_{\ell_\infty}\|_r \le c'_{\alpha,\overline{q}_d} r^{(\overline{q}_d-1)/\alpha} \Lambda_{\sigma(i)} \tag{6.8}$$

for every $r \ge 1$ with $c'_{\alpha,\overline{q}_d} > 0$ depending only on $\alpha, \overline{q}_d$. This inequality also holds true when $\overline{q}_d = 1$ because in this case V_i^σ is non-random and thus it is a direct consequence of (6.5). Meanwhile, (A.3) and Lemma A.3 yield $P(\mathcal{C}_{\sigma,i}^c) \le 2e^{-(\tau/2^{1\vee\alpha^{-1}})^\alpha}$. Consequently, we obtain

$$\frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N}\sum_{i=1}^N \mathbf{I}(2)_i^\sigma \lesssim \left(\max_{1 \le j \le m}\beta^{m-j}\|h^{(j)}\|_\infty\right) e^{-(\tau/2^{1\vee\alpha^{-1}})^\alpha}\sum_{i=1}^N \Lambda_i^m. \tag{6.9}$$

Third, we estimate $\mathbf{I}(3)_i^\sigma$. Lemma 3.1 yields

$$\mathbf{I}(3)_i^\sigma \lesssim \left(\max_{1 \le j \le m}\beta^{m-j}\|h^{(j)}\|_\infty\right)\mathrm{E}\left[\|V_i^\sigma\|_{\ell_\infty}^m; \mathcal{D}_{\sigma,i}^c\right].$$

If $\overline{q}_d > 1$, (6.8) and Lemma A.4 yield

$$P\left(\|V_i^\sigma\|_{\ell_\infty} \ge x\right) \le e^{(\overline{q}_d-1)/\alpha}\exp\left(-\left(\frac{x}{K_{\alpha,\overline{q}_d}\Lambda_{\sigma(i)}}\right)^{\alpha/(\overline{q}_d-1)}\right)$$

for every $x > 0$ with $K_{\alpha,\overline{q}_d} > 0$ depending only on $\alpha, \overline{q}_d$. Hence, Lemma 6.1 yields

$$\mathrm{E}\left[\|V_i^\sigma\|_{\ell_\infty}^m; \mathcal{D}_{\sigma,i}^c\right] \lesssim (\rho \vee 1)^m \Lambda_{\sigma(i)}^m \exp\left(-\left(\frac{\rho}{K_{\alpha,\overline{q}_d}}\right)^{\alpha/(\overline{q}_d-1)}\right).$$

Meanwhile, if $\overline{q}_d = 1$, V_i^σ is non-random, so (6.8) yields $\mathrm{E}\left[\|V_i^\sigma\|_{\ell_\infty}^m; \mathcal{D}_{\sigma,i}^c\right] \lesssim \Lambda_{\sigma(i)}^m 1_{\{c'_{\alpha,\overline{q}_d} > \rho\}}$. Consequently, setting $K'_{\alpha,\overline{q}_d} := K_{\alpha,\overline{q}_d} \vee c'_{\alpha,\overline{q}_d}$, we obtain

$$
\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^N \mathbf{I}(3)_i^\sigma
$$
$$
\lesssim \left(\max_{1 \le j \le m} \beta^{m-j} \|h^{(j)}\|_\infty \right) \chi_{(\overline{q}_d-1)/\alpha} \left(\frac{\rho}{K'_{\alpha,\overline{q}_d}} \right) (\rho \vee 1)^m \sum_{i=1}^N \Lambda_i^m. \quad (6.10)
$$

Now, combining (6.4), (6.7), (6.9), (6.10) with Lemma A.6, we obtain

$$
\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^N \mathbf{I}_i^\sigma
$$
$$
\lesssim \left(\max_{1 \le j \le m} \beta^{m-j} \|h^{(j)}\|_\infty \right) \left\{ (\log d)^{m(\overline{q}_d-1)/\alpha} \max_{1 \le k \le d} M_N^{mq_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^{m/2} \right.
$$
$$
\left. + e^{-(\tau/2^{1\vee\alpha^{-1}})^\alpha} M_N^m \sum_{i=1}^N \Lambda_i^m + \chi_{(\overline{q}_d-1)/\alpha} \left(\frac{\rho}{K'_{\alpha,\overline{q}_d}} \right) (\rho \vee 1)^m M_N^m \sum_{i=1}^N \Lambda_i^m \right\}.
$$
$$
(6.11)
$$

Next, we consider $\mathbf{II}_i^\sigma$. Lemma 3.1 yields

$$
\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^N \mathbf{II}_i^\sigma \lesssim \frac{1}{N!} \left(\max_{1 \le j \le m} \beta^{m-j} \|h^{(j)}\|_\infty \right)
$$
$$
\times \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^N \mathrm{E}\left[(|X_{\sigma(i)}|^m + |Y_{\sigma(i)}|^m) \max_{1 \le k \le d} |V_{k,i}^\sigma|^m; \mathcal{E}_{\sigma,i}^c \right].
$$

Since $X_{\sigma(i)}$ and $Y_{\sigma(i)}$ are independent of V_i^σ, Lemma A.6 and (6.8) imply that

$$
\|(|X_{\sigma(i)}| + |Y_{\sigma(i)}|)\|V_i^\sigma\|_{\ell_\infty}\|_r \le L_{\alpha,\overline{q}_d} r^{\overline{q}_d/\alpha} M_N \Lambda_{\sigma(i)}
$$

for every $r \ge 1$ with $L_{\alpha,\overline{q}_d} > 0$ depending only on $\alpha, \overline{q}_d$. Thus, by Lemma A.4 we obtain

$$
P\left((|X_{\sigma(i)}| + |Y_{\sigma(i)}|)\|V_i^\sigma\|_{\ell_\infty} \ge x\right) \le e^{\overline{q}_d/\alpha} \cdot \exp\left(-\left(\frac{x}{L'_{\alpha,\overline{q}_d} M_N \Lambda_{\sigma(i)}}\right)^{\alpha/\overline{q}_d}\right)
$$

for every $x > 0$ with $L'_{\alpha, \overline{q}_d} > 0$ depending only on $\alpha, \overline{q}_d$. So Lemma 6.1 yields

$$\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} \mathbf{\Pi}_i^{\sigma}$$

$$\lesssim \left(\max_{1 \leq j \leq m} \beta^{m-j} \|h^{(j)}\|_{\infty} \right) \sum_{i=1}^{N} (\tau \rho \vee 1)^m \, M_N^m \Lambda_i^m \exp\left(-\left(\frac{\tau \rho}{L'_{\alpha, \overline{q}_d}} \right)^{\alpha/\overline{q}_d} \right).$$

Combining this inequality with (6.2), (6.3) and (6.11), we complete the proof. □

7 Proof of the Main Results

7.1 Proof of Theorem 2.1

The following result is a version of [47, Lemma 4.3]. The proof is a minor modification of the latter's, so we omit it.

Lemma 7.1 *Let $q \in \mathbb{N}$ and $f : [N]^q \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Also, let $X = (X_i)_{i=1}^{N}$ and $Y = (Y_i)_{i=1}^{N}$ be two sequences of independent centered random variables with unit variance. Suppose that there are integers $3 \leq m \leq l$ such that $M_N := \max_{1 \leq i \leq N}(\|X_i\|_l \vee \|Y_i\|_l) < \infty$ and $\mathrm{E}[X_i^r] = \mathrm{E}[Y_i^r]$ for all $i \in [N]$ and $r \in [m-1]$. Then, we have $Q(f; X), Q(f; Y) \in L^l(P)$ and*

$$|\mathrm{E}[Q(f; X)^l] - \mathrm{E}[Q(f; Y)^l]|$$

$$\leq C M_N^{ql} (1 \vee \|f\|_{\ell_2})^{l-m} \sum_{i=1}^{N} \max\{\mathrm{Inf}_i(f)^{\frac{m}{2}}, \mathrm{Inf}_i(f)^{\frac{l}{2}}\},$$

where $C > 0$ depends only on q, l.

Proof of Theorem 2.1 Throughout the proof, for two real numbers a and b, the notation $a \lesssim b$ means that $a \leq cb$ for some constant $c > 0$ which depends only on $\alpha, \overline{q}_d$. Moreover, if $(\log d)^{\mu + \frac{1}{2}} \delta_1 [Q(X)]^{\frac{1}{3}} \geq 1$, then the claim evidently holds true with $C = 1$, so we may assume $(\log d)^{\mu + \frac{1}{2}} \delta_1 [Q(X)]^{\frac{1}{3}} < 1$.

Set $s_i := \mathrm{E}[X_i^3]$ for every i. We take a sequence $Y = (Y_i)_{i=1}^{N}$ of independent random variables such that

$$Y_i \sim \begin{cases} \mathcal{N}(0, 1) & \text{if } s_i = 0, \\ \gamma_+(4/s_i^2) & \text{if } s_i > 0, \\ \gamma_-(4/s_i^2) & \text{if } s_i < 0. \end{cases}$$

By construction, we have $\mathrm{E}[X_i^r] = \mathrm{E}[Y_i^r]$ for any $i \in [N]$ and $r \in [3]$. Moreover, one can easily check that $\|Y_i\|_r \leq \overline{B}_N (r-1)^w$ for any $i \in [N]$ and $r \geq 2$. Hence, by

Lemma A.5 we have $\max_{1 \le i \le N} \|Y_i\|_{\psi_\alpha} \le c_\alpha \overline{B}_N$ with $c_\alpha \ge 1$ depending only on α. Therefore, applying Proposition 6.1 with $m = 4$, we obtain

$$
\begin{aligned}
&\Delta_\varepsilon(\boldsymbol{Q}(X), \boldsymbol{Q}(Y)) \\
&\le C_1 \varepsilon^{-4} (\log d)^3 \left\{ (\log d)^{4(\overline{q}_d - 1)/\alpha} \max_{1 \le k \le d} \overline{B}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2 \right. \\
&\quad + \left(e^{-\left(\frac{\tau}{K_1}\right)^\alpha} + \chi_{(\overline{q}_d - 1)/\alpha}\left(\frac{\rho}{K_2}\right) \right)(\rho \vee 1)^4 \\
&\quad \left. + \exp\left(-\left(\frac{\tau \rho}{K_3 \overline{B}_N}\right)^{\alpha/\overline{q}_d} \right)(\tau \rho \vee 1)^4 \, \overline{B}_N^4 \sum_{i=1}^N \Lambda_i^4 \right\} \\
&=: C_1 \varepsilon^{-4} (\log d)^3 \, (\mathbb{I} + \mathbb{III})
\end{aligned}
$$

for any $\varepsilon > 0$ and $\tau, \rho \ge 0$ with $\tau \rho c_\alpha \overline{B}_N \max_{1 \le i \le N} \Lambda_i \le \varepsilon / \log d$, where $C_1, K_1, K_2, K_3 > 0$ depend only on $\alpha, \overline{q}_d$, and $\Lambda_i := (\log d)^{(\overline{q}_d - 1)/\alpha} \max_{1 \le k \le d} \overline{B}_N^{q_k - 1} \sqrt{\mathrm{Inf}_i(f_k)}$. We apply this inequality with $\tau := (\log d^2)^{1/\alpha} \{K_1 \vee (K_3/K_2)\}$, $\rho := (\log d^2)^{(\overline{q}_d - 1)/\alpha} K_2$ and

$$
\varepsilon := (\log d)^{1/6} \delta_0[\boldsymbol{Q}(X)]^{1/3} + (\log d)^\mu \delta_1[\boldsymbol{Q}(X)]^{1/3} + (\log d) \tau \rho c_\alpha \overline{B}_N \max_{1 \le i \le N} \Lambda_i.
$$

By construction, we have

$$
\mathbb{III} \lesssim d^{-1} \sum_{i=1}^N \sum_{k=1}^d \overline{B}_N^{4q_k} \mathrm{Inf}_i(f_k)^2 \le \max_{1 \le k \le d} \overline{B}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2.
$$

Therefore, we obtain

$$
\begin{aligned}
\Delta_\varepsilon(\boldsymbol{Q}(X), \boldsymbol{Q}(Y)) &\lesssim \varepsilon^{-4} (\log d)^{3 + 4(\overline{q}_d - 1)/\alpha} \max_{1 \le k \le d} \overline{B}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2 \\
&\lesssim \varepsilon^{-4} (\log d)^{3 + 4(\overline{q}_d - 1)/\alpha} \delta_1[\boldsymbol{Q}(X)]^2 \\
&\le (\log d)^{3 + 4\{(\overline{q}_d - 1)/\alpha - \mu\}} \delta_1[\boldsymbol{Q}(X)]^{2/3}.
\end{aligned}
$$

Since $3 + 4\{(\overline{q}_d - 1)/\alpha - \mu\} \le \frac{4}{3\alpha}(\overline{q}_d - 1) + \frac{5}{3} \le 2\mu + 1$ and $(\log d)^{\mu + \frac{1}{2}} \delta_1[\boldsymbol{Q}(X)]^{\frac{1}{3}} < 1$, we conclude that

$$
\Delta_\varepsilon(\boldsymbol{Q}(X), \boldsymbol{Q}(Y)) \lesssim (\log d)^{2\mu + 1} \delta_1[\boldsymbol{Q}(X)]^{2/3} \le (\log d)^{\mu + \frac{1}{2}} \delta_1[\boldsymbol{Q}(X)]^{\frac{1}{3}}. \quad (7.1)
$$

Meanwhile, Proposition 5.1 yields

$$
\Delta_\varepsilon(\boldsymbol{Q}(Y), Z) \lesssim \varepsilon^{-2} (\log d) \left(\delta_0[\boldsymbol{Q}(Y)] + \delta_2[\boldsymbol{Q}(Y)] \right).
$$

Now, in the present situation, the constants $w_*, \overline{v}_N$ and $\overline{\eta}_N$ appearing in Proposition 5.1 satisfy $w_* = w$, $\overline{v}_N \le 2 + \overline{A}_N^2/2$ and $\overline{\eta}_N^{-1} \le \overline{A}_N/2$, so we have

$$
\begin{aligned}
&\delta_2[\boldsymbol{Q}(\boldsymbol{Y})] \\
&\quad \le (\overline{A}_N/2)^{2w\overline{q}_d-1} (\log d)^{2w\overline{q}_d-1} \max_{1 \le j,k \le d} \Big\{ 1_{\{q_j < q_k\}} \|Q(f_j; \boldsymbol{Y})\|_4 \kappa_4(Q(f_k; \boldsymbol{Y}))^{1/4} \\
&\qquad + 1_{\{q_j = q_k\}} \sqrt{2\kappa_4(Q(f_k; \boldsymbol{Y})) + \Big((1 + \overline{A}_N^2/4)^{q_k} - 1\Big)(2q_k)! \mathfrak{c}_{q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2} \Big\}.
\end{aligned}
$$

Moreover, by a standard hypercontractivity argument, we have $\|Q(f_j; \boldsymbol{Y})\|_4 \lesssim \overline{A}_N^{q_j} \|Q(f_j; \boldsymbol{Y})\|_2$ for every j. Also, since Lemma A.5 yields $\max_{1 \le i \le N} \|Y_i\|_{\psi_\alpha} \lesssim \underline{\eta}_N^{-1}$, by Lemma 7.1 (with $l = m = 4$) we obtain

$$
|\mathrm{E}[Q(f_k; \boldsymbol{X})^4] - \mathrm{E}[Q(f_k; \boldsymbol{Y})^4]| \lesssim \overline{A}_N^{4q_k} \sum_{i=1}^N \mathrm{Inf}_i(f_k)^2
$$

for every k. Since we have $\|Q(f_k; \boldsymbol{Y})\|_2 = \sqrt{q_k!} \|f_k\|_{\ell_2} = \|Q(f_k; \boldsymbol{X})\|_2$ for every k, it holds that $\delta_2[\boldsymbol{Q}(\boldsymbol{Y})] \lesssim (\log d)^{2w\overline{q}_d-1} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]$. Consequently, we obtain

$$
\Delta_\varepsilon(\boldsymbol{Q}(\boldsymbol{Y}), Z) \lesssim (\log d)^{2/3} \delta_0[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + (\log d)^{2(w\overline{q}_d-\mu)} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]^{1/3}.
$$

Since $2(w\overline{q}_d - \mu) \le \frac{2}{3} w\overline{q}_d + \frac{1}{3} \le \mu + \frac{1}{2}$, we conclude that

$$
\Delta_\varepsilon(\boldsymbol{Q}(\boldsymbol{Y}), Z) \lesssim (\log d)^{2/3} \delta_0[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + (\log d)^{\mu+\frac{1}{2}} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]^{1/3}. \tag{7.2}
$$

Now, (7.1)–(7.2) imply that

$$
\begin{aligned}
\Delta_\varepsilon(\boldsymbol{Q}(\boldsymbol{X}), Z) &\le \Delta_\varepsilon(\boldsymbol{Q}(\boldsymbol{X}), \boldsymbol{Q}(\boldsymbol{Y})) + \Delta_\varepsilon(\boldsymbol{Q}(\boldsymbol{Y}), Z) \\
&\lesssim (\log d)^{2/3} \delta_0[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + (\log d)^{\mu+\frac{1}{2}} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]^{1/3}.
\end{aligned}
$$

Therefore, Proposition 3.1 yields

$$
\begin{aligned}
\sup_{x \in \mathbb{R}^d} &|P(\boldsymbol{Q}(\boldsymbol{X}) \le x) - P(Z \le x)| \\
&\lesssim (\log d)^{2/3} \delta_0[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + (\log d)^{\mu+\frac{1}{2}} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + \underline{\sigma}^{-1} \varepsilon \sqrt{\log d} \\
&\lesssim (1 + \underline{\sigma}^{-1})\{(\log d)^{2/3} \delta_0[\boldsymbol{Q}(\boldsymbol{X})]^{1/3} + (\log d)^{\mu+\frac{1}{2}} \delta_1[\boldsymbol{Q}(\boldsymbol{X})]^{1/3}\} \\
&\quad + \underline{\sigma}^{-1} (\log d)^{(2\overline{q}_d-1)/\alpha+\frac{3}{2}} \max_{1 \le k \le d} \overline{B}_N^{q_k} \sqrt{\mathcal{M}(f_k)}.
\end{aligned}
$$

This completes the proof. $\square$

7.2 Proof of Corollaries 2.1 and 2.2

Corollary 2.1 can be shown in an analogous manner to the proof of [18, Corollary 5.1] with applying Theorem 2.1 instead of [18, Lemma 5.1]. Corollary 2.2 immediately follows from Corollary 2.1. □

7.3 Proof of Theorem 2.2

Lemma 7.2 *Let $q \geq 2$ and $f : [N]^q \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Suppose that the sequence X satisfies one of conditions (A)–(C). Then, we have $\kappa_4(Q(f; X)) \geq 0$ and*

$$\mathcal{M}(f) \leq \max_{1 \leq r \leq q-1} \|f \star_r f\|_{\ell_2} \leq \frac{1}{q \cdot q!} \sqrt{\kappa_4(Q(f; X))}. \tag{7.3}$$

Proof The first inequality in (7.3) is a consequence of Eq.(1.9) in [47] (note that they define $\mathrm{Inf}_i(f)$ with dividing ours by $(q-1)!$). To prove the second inequality in (7.3), first we suppose that X satisfies condition (A). Let $G = (G_i)_{i \in \mathbb{N}}$ be a sequence of independent standard normal variables. Then, by [45, Proposition 3.1] we have $\kappa_4(Q(f; X)) \geq \kappa_4(Q(f; G))$. Therefore, (5.11) yields the desired result. Next, when X satisfies condition (B), the desired result follows from Eq.(5.3) in [29]. Finally, when X satisfies condition (C), the desired result follows from (5.11). Hence, we complete the proof. □

Lemma 7.3 *Let F, G be two random variables such that $\|F\|_{\psi_\alpha} \vee \|G\|_{\psi_\alpha} < \infty$ for some $\alpha > 0$. Then, we have*

$$|\mathrm{E}[|F|^r] - \mathrm{E}[|G|^r]| \leq \frac{2r(\|F\|_{\psi_\alpha}^r + \|G\|_{\psi_\alpha}^r)}{\alpha b} \Gamma\left(\frac{r}{\alpha b}\right) \sup_{x \in \mathbb{R}} |P(F \leq x) - P(G \leq x)|^b$$

for any $r \geq 1$ and $b \in (0, 1)$, where Γ denotes the gamma function.

Proof This is an easy consequence of [55, Theorem 8.16] and Lemma A.3. □

Proof of Theorem 2.2 The inequality $\kappa_4(Q(f; X)) \geq 0$ is proved in Lemma 7.2. The implication (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) is obvious. The implication (i) $\Rightarrow$ (iv) follows from Corollary 2.2 and Lemma 7.2.

It remains to prove (ii) $\Rightarrow$ (i). In view of Lemma 7.3, it is enough to prove $\sup_{n \in \mathbb{N}} \max_{1 \leq j \leq d_n} (\|Q(f_{n,j}; X)\|_{\psi_\beta} + \|Z_{n,j}\|_{\psi_\beta}) < \infty$ for some $\beta > 0$. This follows from Lemmas 6.2 and A.5. □

7.4 Proof of Lemma 2.1

Let us define the sequence of random variables $(Y_i)_{i=1}^N$ in the same way as in the proof of Theorem 2.1. Then, since $\mathrm{E}[Y_i^4] = 3 + \frac{3}{2}|\mathrm{E}[X_i^3]| \leq \frac{9}{2}M$, Lemma 7.1 yields

$$|\mathrm{E}[Q(f; X)^4] - \mathrm{E}[Q(f; Y)^4]| \leq C_1 M^q \mathcal{M}(f) \|f\|_{\ell_2}^2,$$

where $C_1 > 0$ depends only on q. Now, since $\mathrm{E}[Q(f; X)^2] = q! \|f\|_{\ell_2}^2 = \mathrm{E}[Q(f; Y)^2]$ and $\sqrt{\kappa_4(Q(f; Y))} \geq q \cdot q! \max_{1 \leq r \leq q-1} \|f \star_r f\|_{\ell_2}$ by Lemma 7.2, we obtain the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

7.5 Proof of Proposition 2.3

Lemma 7.4 *Let $X = (X_i)_{i=1}^N$ be a sequence of independent centered random variables with unit variance and such that $M := \max_{1 \leq i \leq N} \mathrm{E}[X_i^4] < \infty$. Also, let $f : [N]^2 \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Then, we have*

$$|\kappa_4(Q(f; X))| \leq C \left\{ (1 + M)\|f\|_{\ell_2}^2 \mathcal{M}(f) + \mathrm{tr}([f]^4) \right\},$$

where $C > 0$ is a universal constant.

Proof By Proposition 3.1 and Eq.(3.1) in [21], we have

$$|\mathrm{E}[Q(f; X)^4] - 6G_{\mathrm{V}}| \leq G_{\mathrm{I}} + 18G_{\mathrm{II}} + 24|G_{\mathrm{IV}}|,$$

where

$$G_{\mathrm{I}} = 2^3 \sum_{(i,j) \in \Delta_2^N} f(i, j)^4 \mathrm{E}[X_i^4]\mathrm{E}[X_j^4],$$

$$G_{\mathrm{II}} = 2^3 \sum_{(i,j,k) \in \Delta_3^N} f(i, j)^2 f(i, k)^2 \mathrm{E}[X_i^4],$$

$$G_{\mathrm{IV}} = 2 \sum_{(i,j,k,l) \in \Delta_4^N} f(i, j) f(i, k) f(l, j) f(l, k),$$

$$G_{\mathrm{V}} = 2 \sum_{(i,j,k,l) \in \Delta_4^N} f(i, j)^2 f(k, l)^2.$$

Since we have $2\|f\|_{\ell_2}^4 - G_{\mathrm{V}} \leq 8\|f\|_{\ell_2}^2 \mathcal{M}(f)$, it holds that $|\kappa_4(Q(f; X))| \leq |\mathrm{E}[Q(f; X)^4] - 6G_{\mathrm{V}}| + 48\|f\|_{\ell_2}^2 \mathcal{M}(f)$. Meanwhile, a straightforward computation yields

$$\mathrm{tr}([f]^4) = \sum_{(i,j) \in \Delta_2^N} f(i, j)^4 + 2 \sum_{(i,j,k) \in \Delta_3^N} f(i, k)^2 f(j, k)^2$$
$$+ \sum_{(i,j,k,l) \in \Delta_4^N} f(i, k) f(j, k) f(i, l) f(j, l).$$

Hence, we obtain

$$|\mathrm{E}[Q(f; X)^4] - 6G_{\mathrm{V}}|$$

$$\leq C_1 \left\{ (1+M) \left(\sum_{i,k=1}^{N} f(i,k)^4 + \sum_{(i,j,k)\in\Delta_3^N} f(i,k)^2 f(j,k)^2 \right) \right.$$
$$\left. + \mathrm{tr}([f]^4) \right\},$$

where $C_1 > 0$ is a universal constant. Since it holds that

$$\max \left\{ \sum_{(i,j)\in\Delta_2^N} f(i,j)^4, \sum_{(i,j,k)\in\Delta_3^N} f(i,k)^2 f(j,k)^2 \right\} \leq \|f\|_{\ell_2}^2 \mathcal{M}(f),$$

we obtain the desired result.   □

Proof of Proposition 2.3 The desired result immediately follows from Corollary 2.2 and Lemma 7.4.   □

7.6 Proof of Proposition 2.5

Define the $n_1 \times n_2$ matrix $\Xi_n(\theta)$ by $\Xi_n(\theta) = (\frac{1}{2}\Delta_i^n X^1 K_\theta^{ij} \Delta_j^n X^2)_{i,j}$, and set

$$\widetilde{\Xi}_n(\theta) = \begin{pmatrix} O & \Xi_n(\theta) \\ \Xi_n(\theta)^\top & O \end{pmatrix}.$$

Note that $U_n^*(\theta) = \boldsymbol{w}^\top \widetilde{\Xi}_n(\theta) \boldsymbol{w}$ with $\boldsymbol{w} = ((w_i^1)_{i=1}^{n_1}, (w_j^2)_{j=1}^{n_2})^\top$. Hence, by Proposition 2.3, it suffices to prove

$$\log^2(\#\mathcal{G}_n) \cdot \sqrt{n} \max_{\theta,\theta'\in\mathcal{G}_n} \left| \mathrm{E}[U_n(\theta)U_n(\theta')] - \mathrm{E}[U_n^*(\theta)U_n^*(\theta') \mid X] \right| \to^p 0, \tag{7.4}$$

$$\log^5(\#\mathcal{G}_n) \max_{\theta\in\mathcal{G}_n} \sqrt{\mathrm{tr}\left(\widetilde{\Xi}_n(\theta)^4\right)} \to^p 0, \tag{7.5}$$

$$\log^5(\#\mathcal{G}_n) \cdot \max_{\theta\in\mathcal{G}_n}$$
$$\sqrt{\max_{1\leq i\leq n_1} n\sum_{j=1}^{n_2} (\Delta_i^n X^1)^2 (\Delta_j^n X^2)^2 K_\theta^{ij} + \max_{1\leq j\leq n_2} n\sum_{i=1}^{n_1} (\Delta_i^n X^1)^2 (\Delta_j^n X^2)^2 K_\theta^{ij}} \to^p 0. \tag{7.6}$$

(7.4) and (7.5) are established in the proof of [35, Proposition B.8] under the current assumptions. Moreover, as in bounding the quantity $\mathrm{E}[R_{n,1}^*]$ in the proof of [35, Proposition B.8], we deduce for any $p \geq 1$

$$\mathrm{E}\left[\left| \max_{\theta\in\mathcal{G}_n} \max_i n\sum_j (\Delta_i^n X^1)^2 (\Delta_j^n X^2)^2 K_\theta^{ij} \right|^p \right]$$

$$\leq n^p \sum_i \mathrm{E}\left[(\Delta_i^n X^1)^{2p} \max_{\theta \in \mathcal{G}_n} \left| \sum_j (\Delta_j^n X^2)^2 K_\theta^{ij} \right|^p \right]$$

$$\leq n^p \sum_i \sqrt{ \mathrm{E}\left[(\Delta_i^n X^1)^{4p} \right] \mathrm{E}\left[\max_{\theta \in \mathcal{G}_n} \left| \sum_j (\Delta_j^n X^2)^2 K_\theta^{ij} \right|^{2p} \right] }$$

$$= O\left(n^p r_n^{p-1} (r_n \log(\#\mathcal{G}_n))^p \right).$$

Exchanging X^1 and X^2, we obtain a similar estimate. Hence, (7.6) holds by assumption. $\qquad\square$

Availability of data and materials Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declaration

Conflict of interest The author declares that he has no conflict of interest.

Appendix

A Properties of the ψ_α-norm

Here we collect several properties of the ψ_α-norm used in this paper. Let α be a positive number. Recall that the ψ_α-norm of a random variable X is defined by

$$\|X\|_{\psi_\alpha} := \inf\{C > 0 : \mathrm{E}[\psi_\alpha(|X|/C)] \leq 1\}, \tag{A.1}$$

where $\psi_\alpha(x) := \exp(x^\alpha) - 1$. From the definition, we can easily deduce the following useful identity:

$$\|X\|_{\psi_\alpha} = \||X|^\alpha\|_{\psi_1}^{1/\alpha}. \tag{A.2}$$

Using this relation, we can derive the properties of the ψ_α-norm from those of the ψ_1-norm. This is very convenient because the latter ones are well-studied in the literature. For example, since $\|\cdot\|_{\psi_1}$ satisfies the triangle inequality, we have

$$\|X + Y\|_{\psi_\alpha} \le 2^{1 \vee \alpha^{-1} - 1}(\|X\|_{\psi_\alpha} + \|Y\|_{\psi_\alpha}) \tag{A.3}$$

for any random variables X, Y. Also, using Young's inequality for products and Hölder's inequality, one can prove $\|X\|_{\psi_1} \le (\log 2)^{1/p-1}\|X\|_{\psi_p}$ for any random variable X and $p > 1$. So we obtain $\|X\|_{\psi_\alpha} \le (\log 2)^{1/\beta - 1/\alpha}\|X\|_{\psi_\beta}$ for any $0 < \alpha \le \beta < \infty$. Other useful results can be obtained from [57, Lemmas 2.2.1–2.2.2]:

Lemma A.1 *Suppose that there are constants $C, K > 0$ such that $P(|X| > x) \le Ke^{-Cx^\alpha}$ for all $x > 0$. Then, we have $\|X\|_{\psi_\alpha} \le ((1 + K)/C)^{1/\alpha}$.*

Lemma A.2 *There is a universal constant $K > 0$ such that*

$$\left\| \max_{1 \le j \le d} |X_j| \right\|_{\psi_\alpha} \le K^{1/\alpha}(\log(d+1))^{1/\alpha} \max_{1 \le j \le d} \|X_j\|_{\psi_\alpha}$$

for any $\alpha > 0$ and random variables $X_1, \ldots, X_d$.

It is also easy to check that $\|X\|_{\psi_\alpha}$ attains the infimum in (A.1) if $\|X\|_{\psi_\alpha} < \infty$. That is, $E[\psi_\alpha(|X|/\|X\|_{\psi_\alpha})] \le 1$. Therefore, the Markov inequality yields the following converse of Lemma A.1:

Lemma A.3 *If $\|X\|_{\psi_\alpha} < \infty$, we have $P(|X| \ge x) \le 2e^{-(x/\|X\|_{\psi_\alpha})^\alpha}$ for every $x > 0$.*

Next, we investigate the relation between the ψ_α-norm and moment growth. First, [26, Lemma A.1] yields the following result:

Lemma A.4 *If there is a constant $A > 0$ such that $\|X\|_p \le Ap^{1/\alpha}$ for all $p \ge 1$, then $P(|X| \ge x) \le e^{1/\alpha}e^{-(\alpha e)^{-1}(x/A)^\alpha}$ for every $x > 0$.*

Combining Lemma A.4 with Lemma A.1, we obtain the following result:

Lemma A.5 *Suppose that there is a constant $A > 0$ such that $\|X\|_p \le Ap^{1/\alpha}$ for all $p \ge 1$. Then, we have $\|X\|_{\psi_\alpha} \le \left((1 + e^{1/\alpha})\alpha e\right)^{1/\alpha} A$.*

Lemma A.3 and [26, Lemma A.2] yield the following converse of Lemma A.5:

Lemma A.6 *For all $p \ge 1$, it holds that $\|X\|_p \le c_\alpha\|X\|_{\psi_\alpha}p^{1/\alpha}$ with $c_\alpha := e^{1/2e-1/\alpha}\alpha^{-1/\alpha} \max\left\{1, 2\sqrt{\frac{2\pi}{\alpha}}e^{\alpha/12}\right\}$.*

Finally, we have the following Hölder-type inequality for the ψ_α-norm:

Lemma A.7 *([36], Proposition S.3.2) Let X_1, X_2 be two random variables such that $\|X_1\|_{\psi_{\alpha_1}} + \|X_2\|_{\psi_{\alpha_2}} < \infty$ for some $\alpha_1, \alpha_2 > 0$. Then, we have $\|X_1 X_2\|_{\psi_\alpha} \leq \|X_1\|_{\psi_{\alpha_1}} \|X_2\|_{\psi_{\alpha_2}}$, where $\alpha > 0$ is defined by the equation $1/\alpha = 1/\alpha_1 + 1/\alpha_2$.*

B Proof of Lemma 6.2

Lemma B.1 *(Strong domination) Let $(\xi_i)_{i=1}^N$ and $(\theta_i)_{i=1}^N$ be two sequences of independent symmetric random variables. Suppose that there is an integer $k > 0$ such that $P(|\xi_i| > t) \leq k P(|\theta_i| > t)$ for all $i \in [N]$ and $t > 0$. Then, for any $p \geq 1$ and $a_1, \ldots, a_N \in \mathbb{R}$ we have*

$$\left\| \sum_{i=1}^N a_i \xi_i \right\|_p \leq (2k)^{1/p} k \left\| \sum_{i=1}^N a_i \theta_i \right\|_p .$$

Proof This is a consequence of Theorem 3.2.1 and Corollary 3.2.1 in [37]. $\qquad \square$

Lemma B.2 *Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent copies of a symmetric random variable ξ satisfying $P(|\xi| \geq t) = e^{-|t|^\alpha}$ for every $t \geq 0$ and some $0 < \alpha \leq 2$. Then, there is a constant $C_\alpha > 0$ which depends only on α such that*

$$\left\| \sum_{i=1}^N a_i \xi_i \right\|_p \leq C_\alpha p^{1/\alpha} \sqrt{\sum_{i=1}^N a_i^2}$$

for any $p \geq 2$, $N \in \mathbb{N}$ and $a_1, \ldots, a_N \in \mathbb{R}$.

Proof Note that, by Lemmas A.1 and A.6, there is a constant $K_\alpha > 0$ depending only on α such that $\|\xi\|_r \leq K_\alpha r^\alpha$ for all $r \geq 1$. Then, the result follows from [31, Theorem 1.1] when $\alpha \leq 1$ and the Gluskin–Kwapień inequality (cf. page 17 of [31]) when $\alpha > 1$. $\qquad \square$

Lemma B.3 *Let $(\zeta_i)_{i=1}^N$ be a sequence of independent centered random variables such that $M := \max_{1 \leq i \leq N} \|\zeta_i\|_{\psi_\alpha} < \infty$ for some $0 < \alpha \leq 2$. Then, we have*

$$\left\| \sum_{i=1}^N a_i \zeta_i \right\|_p \leq K_\alpha M p^{1/\alpha} \sqrt{\sum_{i=1}^N a_i^2}$$

for any $p \geq 1$ and $a_1, \ldots, a_N \in \mathbb{R}$, where $K_\alpha > 0$ depends only on α.

Proof Thanks to symmetrization inequalities (see, e.g., [57, Lemma 2.3.1]), it suffices to consider the case that ζ_i is symmetric for all i. Then, the result follows from Lemmas A.3, B.1 and B.2 . $\qquad \square$

Lemma B.4 *Let $(X_{i,j})_{1 \leq i \leq N, 1 \leq j \leq q}$ be an array of independent centered random variables. Assume $M := \max_{1 \leq i \leq N, 1 \leq j \leq q} \|X_{i,j}\|_{\psi_\alpha} < \infty$ for some $0 < \alpha \leq 2$. Then,*

$$\left\| \sum_{i_1,\ldots,i_q=1}^{N} f(i_1, \ldots, i_q) X_{i_1,1} \cdots X_{i_q,q} \right\|_p \leq K_\alpha^q \, p^{q/\alpha} M^q \|f\|_{\ell_2}$$

for any $p \geq 2$ and $f : [N]^q \to \mathbb{R}$, where $K_\alpha > 0$ is the same as in Lemma B.3.

Proof This can easily be shown by induction on q and using Lemma B.3.   □

Proof of Lemma 6.2 The claim is an immediate consequence of [23, Theorem 1], [55, Theorem 8.16] and Lemma B.4.   □

References

1. Azmoodeh, E., Campese, S., Poly, G.: Fourth Moment Theorems for Markov diffusion generators. J. Funct. Anal. **266**, 2341–2359 (2014)
2. Azmoodeh, E., Malicet, D., Mijoule, G., Poly, G.: Generalization of the Nualart–Peccati criterion. Ann. Probab. **44**, 924–954 (2016)
3. Azmoodeh, E., Peccati, G.: Malliavin–Stein method: a survey of recent developments. Working paper. arXiv:1809.01912 (2018)
4. Azmoodeh, E., Peccati, G., Poly, G.: The law of iterated logarithm for subordinated Gaussian sequences: uniform Wasserstein bounds. ALEA Lat. Am. J. Probab. Math. Stat. **13**, 659–686 (2016)
5. Bakry, D., Gentil, I., Ledoux, M.: Analysis and Geometry of Markov Diffusion Operators. Springer, Berlin (2014)
6. Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., Kato, K.: High-dimensional econometrics and regularized GMM. Working paper. (2018) arXiv:1806.01888
7. Bentkus, V.: On the dependence of the Berry–Esseen bound on dimension. J. Statist. Plann. Inference **113**, 385–402 (2003)
8. Bentkus, V., Götze, F., Paulauskas, V., Račkauskas, A.: The accuracy of Gaussian approximation in Banach spaces. In: Prokhorov, Y., Statulevicius, V. (eds.) Limit Theorems of Probability Theory, Chapter II, pp. 25–111. Springer, Berlin (2000)
9. Campese, S., Nourdin, I., Peccati, G., Poly, G.: Multivariate Gaussian approximations on Markov chaoses. Electron. Commun. Probab. **21**, 1–9 (2016)
10. Chen, X.: Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. Ann. Stat. **46**, 642–678 (2018)
11. Chen, X., Kato, K.: Randomized incomplete U-statistics in high dimensions. Ann. Stat. **47**, 3127–3156 (2019)
12. Chen, X., Kato, K.: Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. Probab. Theory Relat. Fields **176**, 1097–1163 (2020)
13. Chernozhukov, V., Chetverikov, D., Kato, K.: Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Stat. **41**, 2786–2819 (2013)
14. Chernozhukov, V., Chetverikov, D., Kato, K.: Supplement to "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". https://doi.org/10.1214/13-AOS1161SUPP (2013)
15. Chernozhukov, V., Chetverikov, D., Kato, K.: Gaussian approximation of suprema of empirical processes. Ann. Stat. **42**, 1564–1597 (2014)
16. Chernozhukov, V., Chetverikov, D., Kato, K.: Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Probab. Theory Relat. Fields **162**, 47–70 (2015)
17. Chernozhukov, V., Chetverikov, D., Kato, K.: Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. Stoch. Process. Appl. **126**, 3632–3651 (2016)

18. Chernozhukov, V., Chetverikov, D., Kato, K.: Central limit theorems and bootstrap in high dimensions. Ann. Probab. **45**, 2309–2353 (2017)
19. Chernozhukov, V., Chetverikov, D., Kato, K.: Detailed proof of Nazarov's inequality. Unpublished paper. arXiv:1711.10696 (2017)
20. Courtade, T.A., Fathi, M., Pananjady, A.: Existence of Stein kernels under a spectral gap, and discrepancy bounds. Ann. Inst. Henri Poincaré Probab. Stat. **55**, 777–790 (2019)
21. de Jong, P.: A central limit theorem for generalized quadratic forms. Probab. Theory Relat. Fields **75**, 261–277 (1987)
22. de Jong, P.: A central limit theorem for generalized multilinear forms. J. Multivariate Anal. **34**, 275–289 (1990)
23. de la Pena, V.H., Montgomery-Smith, S.J.: Decoupling inequalities for the tail probabilities of multivariate U-statistics. Ann. Probab. **23**, 806–816 (1995)
24. Deng, H., Zhang, C.-H.: Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. Ann. Stat. **48**, 3643–3671 (2020)
25. Dette, H., Hetzler, B.: Specification tests indexed by bandwidths. Sankhyā Indian J. Stat. **69**, 28–54 (2007)
26. Dirksen, S.: Tail bounds via generic chaining. Electron. J. Probab. **20**, 1–29 (2015)
27. Döbler, C., Krokowski, K.: On the fourth moment condition for Rademacher chaos. Ann. Inst. Henri Poincaré Probab. Stat. **55**, 61–97 (2019)
28. Döbler, C., Peccati, G.: Quantitative de Jong theorems in any dimension. Electron. J. Probab. **22**, 1–35 (2017)
29. Döbler, C., Vidotto, A., Zheng, G.: Fourth moment theorems on the Poisson space in any dimension. Electron. J. Probab. **23**, 1–27 (2018)
30. Götze, F., Tikhomirov, A.N.: Asymptotic distribution of quadratic forms. Ann. Probab. **27**, 1072–1098 (1999)
31. Hitczenko, P., Montgomery-Smith, S., Oleszkiewicz, K.: Moment inequalities for sums of certain independent symmetric random variables. Studia Math. **123**, 15–42 (1997)
32. Horowitz, J.L., Spokoiny, V.G.: An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica **69**, 599–631 (2001)
33. Janson, S.: Gaussian Hilbert Space. Cambridge University Press, Cambridge (1997)
34. Koike, Y.: Mixed-normal limit theorems for multiple Skorohod integrals in high-dimensions, with application to realized covariance. Electron. J. Stat. **13**, 1443–1522 (2019)
35. Koike, Y.: Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. Ann. Stat. **47**, 1663–1687 (2019)
36. Kuchibhotla, A.K., Chakrabortty, A.: Moving beyond sub-Gaussianity in high-dimensional statistics: applications in covariance estimation and linear regression. Working paper. arXiv:1804.02605 (2018)
37. Kwapień, S., Woyczyński, W.A.: Random Series and Stochastic Integrals: Single and Multiple. Birkhäuser, Basel (1992)
38. Ledoux, M.: Chaos of a Markov operator and the fourth moment condition. Ann. Probab. **40**, 2439–2459 (2012)
39. Ledoux, M., Nourdin, I., Peccati, G.: Stein's method, logarithmic Sobolev and transport inequalities. Geom. Funct. Anal. **25**, 256–306 (2015)
40. Liu, M., Shang, Z., Cheng, G.: Nonparametric testing under random projection. Working paper. arXiv:1802.06308 (2018)
41. Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: invariance and optimality. Ann. Math. **171**, 295–341 (2010)
42. Nourdin, I., Peccati, G.: Stein's method and exact Berry–Esseen asymptotics for functionals of Gaussian fields. Ann. Probab. **37**, 2231–2261 (2009)
43. Nourdin, I., Peccati, G.: Stein's method on Wiener chaos. Probab. Theory Relat. Fields **145**, 75–118 (2009)
44. Nourdin, I., Peccati, G.: Normal Approximations with Malliavin Calculus: From Stein's Method to Universality. Cambridge University Press, Cambridge (2012)
45. Nourdin, I., Peccati, G., Poly, G., Simone, R.: Classical and free fourth moment theorems: Universality and thresholds. J. Theoret. Probab. **29**, 653–680 (2016)
46. Nourdin, I., Peccati, G., Poly, G., Simone, R.: Multidimensional limit theorems for homogeneous sums: a survey and a general transfer principle. ESAIM Probab. Stat. **20**, 293–308 (2016)

47. Nourdin, I., Peccati, G., Reinert, G.: Invariance principles for homogeneous sums: Universality of Gaussian Wiener chaos. Ann. Probab. **38**, 1947–1985 (2010)
48. Nourdin, I., Peccati, G., Reinert, G.: Stein's method and stochastic analysis of Rademacher functionals. Electron. J. Probab. **15**, 1703–1742 (2010)
49. Nualart, D., Peccati, G.: Central limit theorems for sequences of multiple stochastic integrals. Ann. Probab. **33**, 177–193 (2005)
50. Paulauskas, V.: A note on the rate of convergence in the CLT for empirical processes. Lith. Math. J. **32**, 312–316 (1992)
51. Peccati, G., Tudor, C.A.: Gaussian limits for vector-valued multiple stochastic integrals. In: Émery, M., Ledoux, M., Yor, M. (eds.), Séminaire de probabilitiés XXXVIII, vol. 1857 of Lecture Notes in Math. Springer, pp. 247–262 (2005)
52. Peccati, G., Zheng, C.: Universal Gaussian fluctuations on the discrete Poisson chaos. Bernoulli **20**, 697–715 (2014)
53. Rotar, V.I.: Limit theorems for multilinear forms and quasipolynomial functions. Theory Probab. Appl. **20**, 512–532 (1975)
54. Rotar', V.I.: Limit theorems for polylinear forms. J. Multivariate Anal. **9**, 511–530 (1979)
55. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)
56. Song, Y., Chen, X., Kato, K.: Approximating high-dimensional infinite-order U-statistics: statistical and computational guarantees. Electron. J. Stat. **13**, 4794–4848 (2019)
57. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, Berlin (1996)
58. Zhai, A.: A high-dimensional CLT in $\mathcal{W}_2$ distance with near optimal convergence rate. Probab. Theory Relat. Fields **170**, 821–845 (2018)
59. Zheng, G.: A Peccati–Tudor type theorem for Rademacher chaoses. ESAIM Probab. Stat. **23**, 874–892 (2019)