Electronic Journal of Probability

Electron. J. Probab. **28** (2023), article no. 83, 1–52.
ISSN: 1083-6489 https://doi.org/10.1214/23-EJP976

# From $p$-Wasserstein bounds to moderate deviations*

Xiao Fang[†] Yuta Koike[‡]

**Abstract**

We use a new method via $p$-Wasserstein bounds to prove Cramér-type moderate deviations in (multivariate) normal approximations. In the classical setting that $W$ is a standardized sum of $n$ independent and identically distributed (i.i.d.) random variables with sub-exponential tails, our method recovers the optimal range of $0 \leqslant x = o(n^{1/6})$ and the near optimal error rate $O(1)(1+x)(\log n + x^2)/\sqrt{n}$ for $P(W > x)/(1 - \Phi(x)) \to 1$, where $\Phi$ is the standard normal distribution function. Our method also works for dependent random variables (vectors) and we give applications to the combinatorial central limit theorem, Wiener chaos, homogeneous sums and local dependence. The key step of our method is to show that the $p$-Wasserstein distance between the distribution of the random variable (vector) of interest and a normal distribution grows like $O(p^\alpha \Delta)$, $1 \leqslant p \leqslant p_0$, for some constants $\alpha, \Delta$ and $p_0$. In the above i.i.d. setting, $\alpha = 1, \Delta = 1/\sqrt{n}, p_0 = n^{1/3}$. For this purpose, we obtain general $p$-Wasserstein bounds in (multivariate) normal approximations using Stein's method.

**Keywords:** central limit theorem; Cramér-type moderate deviations; multivariate normal approximation; $p$-Wasserstein distance; Stein's method.
**MSC2020 subject classifications:** 60F05; 60F10; 62E17.
Submitted to EJP on June 29, 2023, final version accepted on June 16, 2023.

## Contents



*Fang X. was partially supported by Hong Kong RGC GRF 14302418, 14305821, a CUHK direct grant and a CUHK start-up grant. Koike Y. was partly supported by JST CREST Grant Number JPMJCR2115 and JSPS KAKENHI Grant Numbers JP19K13668, JP22H00834, JP22H01139.
[†]The Chinese University of Hong Kong. E-mail: xfang@sta.cuhk.edu.hk
[‡]The University of Tokyo. E-mail: kyuta@ms.u-tokyo.ac.jp

## 1 Introduction

Moderate deviations date back to Cramér (1938) who obtained expansions for tail probabilities for sums of independent random variables about the normal distribution. For independent and identically distributed (i.i.d.) random variables $X_1, \cdots, X_n$ with $\mathbb{E}X_1 = 0$ and $\mathrm{Var}(X_1) = 1$ such that $\mathbb{E}e^{|X_1|/b} \leqslant C < \infty$ for some $b > 0$, it follows from Petrov (1975, Ch.8, Eq.(2.41)) that

$$\left| \frac{P(W > x)}{P(Z > x)} - 1 \right| = O(1)(1 + x^3)/\sqrt{n} \tag{1.1}$$

for $0 \leqslant x \leqslant O(1)n^{1/6}$, where $W = (X_1 + \cdots + X_n)/\sqrt{n}$, $Z \sim N(0,1)$ and $O(1)$ is bounded by a constant that depends on $b$ and $C$. The range $0 \leqslant x \leqslant O(1)n^{1/6}$ and the order of the error term $O(1)(1 + x^3)/\sqrt{n}$ are optimal. von Bahr (1967) obtained a multi-dimensional generalization of the result of Cramér (1938) for sums of independent random vectors.

The classical proof of (1.1) depends on the conjugate method, which relies heavily on the independence assumption. A related method is by controlling the cumulants of the random vector of interest; see Saulis and Statulevičius (1991). In dimension one, Chen, Fang and Shao (2013) developed Stein's method (Stein (1972)) to obtain Cramér-type moderate deviation results for dependent random variables. They needed a boundedness condition, which corresponds to assuming $|X_i| \leqslant b$ for an absolute constant $b$ in the above i.i.d. setting. Recently, Liu and Zhang (2021) relaxed the boundedness condition and obtained results for sums of locally dependent random variables and for the combinatorial central limit theorem (CLT).

In this paper, we use a new method via $p$-Wasserstein bounds to prove Cramér-type moderate deviations. For two probability measures $\mu$ and $\nu$ on $\mathbb{R}^d$, $d \geqslant 1$, their

$p$-Wasserstein distance, $p \geqslant 1$, is defined by

$$\mathcal{W}_p(\mu,\nu) = \left(\inf_{\pi} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(dx,dy)\right)^{1/p}, \tag{1.2}$$

where $|\cdot|$ denotes the Euclidean norm and $\pi$ is a measure on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals $\mu$ and $\nu$. For two random vectors $X,Y\in\mathbb{R}^d$, we also write $\mathcal{W}_p(X,Y)=\mathcal{W}_p(\mathcal{L}(X),\mathcal{L}(Y))$. The key idea of our method, explained in more detail in Section 2, is that for a random variable $W$ of interest and a standard normal variable $Z$, if we can show

$$\mathcal{W}_p(W,Z) \leqslant \frac{Cp}{\sqrt{n}} \tag{1.3}$$

for all $1\leqslant p\leqslant n^{1/3}$ and an absolute constant $C$, then, by a smoothing argument, we can recover the optimal range $0\leqslant x = o(n^{1/6})$ for the relative error $|P(W>x)/P(Z>x)-1|$ to vanish and obtain nearly optimal error rate $O(1)(1+x)(1+\log n + x^2)/\sqrt{n}$ subject to the logarithmic term (cf. (1.1)). This method enables us to prove moderate deviation results for dependent random variables as long as we can prove results similar to (1.3) and we give applications to the combinatorial CLT, Wiener chaos, and homogeneous sums in Section 3. The method also works for multi-dimensional approximations (cf. Sections 4 and 5).

It is well known that classical Cramér-type moderate deviation results can be used to prove strong approximation results. See, for example, Komlós, Major and Tusnády (1975, Eq.(2.6)) and the survey by Mason and Zhou (2012). As far as we know, this is the first time that the reverse direction is explored. It is made possible by recent advances in $p$-Wasserstein bounds. In particular, we adapt the approach (cf. Section 6) of Bonis (2020) to obtain $p$-Wasserstein bounds for general dependent random vectors. See Theorems 2.1 and 7.1 for the results via (generalized) exchangeable pairs and Theorem 5.1 for local dependence.

Here, we introduce some of the notations to be used in the statement of results. More notations will be introduced when they are needed in the proofs. $|\cdot|$ denotes the Euclidean norm, $\|\cdot\|_{H.S.}$ denotes the Hilbert-Schmidt norm and $\|\cdot\|_{op}$ denotes the operator norm. $\otimes$ denotes the tensor product. For a random vector $X$ and $p>0$, we set $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$. For a random matrix $Y$ and $p>0$, we set $\|Y\|_p := (\mathbb{E}\|Y\|_{H.S.}^p)^{1/p}$. For the function $\psi_\alpha : [0,\infty)\to[0,\infty)$, $\alpha>0$, defined as

$$\psi_\alpha(x) := \exp(x^\alpha) - 1,$$

the Orlicz (quasi-)norm of a random vector $X$ is defined as

$$\|X\|_{\psi_\alpha} := \inf\{t>0 : \mathbb{E}\psi_\alpha(|X|/t)\leqslant 1\}. \tag{1.4}$$

Unless otherwise stated, we use $c$ and $C$ to denote positive absolute constants, which may differ in different expressions. For a positive integer $q$, we set $[q] := \{1,\dots,q\}$. For a finite set $S$, we denote by $|S|$ the cardinality of $S$.

## 2 Our approach

### 2.1 $p$-Wasserstein bounds

The first step in our approach is proving a $p$-Wasserstein bound between the distribution of the random vector of interest and a normal distribution. We obtain the following $p$-Wasserstein bound using exchangeable pairs.

**Theorem 2.1.** *Let $(W, W')$ be an exchangeable pair of $d$-dimensional random vectors satisfying the approximate linearity condition*

$$\mathbb{E}[W' - W|\mathcal{G}] = -\Lambda(W + R) \tag{2.1}$$

*for some invertible $d \times d$ matrix $\Lambda$, $d$-dimensional random vector $R$ and $\sigma$-algebra $\mathcal{G}$ containing $\sigma(W)$. Assume $\Lambda = \lambda I_d$ for some $\lambda > 0$ (see Theorem 7.1 for a more general case). Assume that $\mathbb{E}|W|^p < \infty$ for some $p \geqslant 1$ and $\mathbb{E}|D|^4 < \infty$, where $D = W' - W$. Then we have*

$$\mathcal{W}_p(W, Z) \leqslant C \int_0^\infty e^{-t} \left( \|R_t\|_p + \frac{\|E\|_p}{\eta_t(p)} + \min\left\{ \frac{\sqrt{d}}{\eta_t(p)}, \frac{\|\mathbb{E}[D^{\otimes 2}|D|^2 1_{\{|D| \leqslant \eta_t(p)\}}|\mathcal{G}]\|_p}{\lambda \eta_t^3(p)} \right\} \right) dt \tag{2.2}$$

$$\leqslant C \left( \int_0^\infty e^{-t} \|R_t\|_p dt + \sqrt{p}\|E\|_p + p d^{1/4} \sqrt{\frac{\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{\lambda}} \right), \tag{2.3}$$

*where $Z \sim N(0, I_d)$ is a $d$-dimensional standard Gaussian vector, $\eta_t(p) := \sqrt{(e^{2t} - 1)/p}$,*

$$R_t := R + \mathbb{E}[\Lambda^{-1} D 1_{\{|D| > \eta_t(p)\}}|\mathcal{G}], \qquad E := \frac{1}{2}\mathbb{E}[\Lambda^{-1} D \otimes D|\mathcal{G}] - I_d, \tag{2.4}$$

*and $C$ is an absolute constant.*

We defer the proof of Theorem 2.1 to Section 6. The proof heavily relies on the techniques developed in Bonis (2020). However, the concrete error bound and the explicit dependence on $p$ that yields optimal moderate deviation results are new. Such $p$-Wasserstein bounds can also be obtained under other dependency structures, e.g., generalized exchangeable pairs (cf. Theorem 7.1) and local dependence (cf. Theorem 5.1).

Next, we give a corollary of Theorem 2.1 in dimension one.

**Corollary 2.1** (The case $d = 1$)**.** Under the setting of Theorem 2.1, assume $d = 1$. We have

$$\mathcal{W}_p(W, Z) \leqslant C \left( \|R\|_p + \sqrt{p}\|E\|_p + p\sqrt{\lambda^{-1}\|\mathbb{E}[D^4|\mathcal{G}]\|_p} \right). \tag{2.5}$$

*Proof of Corollary* 2.1. The corollary is a direct consequence of Theorem 2.1 except that we bound the additional term from $R_t$ by

$$\begin{aligned}
&C \int_0^\infty e^{-t} \|\mathbb{E}[\lambda^{-1} D 1_{\{|D| > \sqrt{(e^{2t}-1)/p}\}}|\mathcal{G}]\|_p dt \\
\leqslant &C\sqrt{p}\lambda^{-1}\|\mathbb{E}[D^2|\mathcal{G}]\|_p \int_0^\varepsilon \frac{e^{-t}}{\sqrt{e^{2t}-1}} dt + Cp^{3/2}\lambda^{-1}\|\mathbb{E}[D^4|\mathcal{G}]\|_p \int_\varepsilon^\infty \frac{e^{-t}}{(e^{2t}-1)^{3/2}} dt \\
\leqslant &C\sqrt{p}\|E\|_p + C\sqrt{p} \int_0^\varepsilon \frac{e^{-t}}{\sqrt{e^{2t}-1}} dt + Cp^{3/2}\lambda^{-1}\|\mathbb{E}[D^4|\mathcal{G}]\|_p \int_\varepsilon^\infty \frac{e^{-t}}{(e^{2t}-1)^{3/2}} dt,
\end{aligned}$$

which is bounded by the summation of second and third error terms in (2.5) by choosing an appropriate $\varepsilon$ as at the end of the proof of Theorem 2.1. □

### 2.2 From $p$-Wasserstein bounds to moderate deviations in dimension one

The next step in our approach is proving moderate deviation results using $p$-Wasserstein bounds. The following result enables such transition in dimension one. In most of our applications of the following result, $r_0 = \alpha_1 = 1$. See Theorem 4.2 for a multi-dimensional result.

**Theorem 2.2.** *Let $W$ be a one-dimensional random variable and $Z$ a standard normal variable. Suppose that*

$$\mathcal{W}_p(W,Z) \leqslant A \max_{1\leqslant r\leqslant r_0} p^{\alpha_r}\Delta_r \text{ for } 1\leqslant p\leqslant p_0 \tag{2.6}$$

*with some constants $\alpha_1,\dots,\alpha_{r_0}\geqslant 0$, $A>0$, $p_0\geqslant 1$ and $\Delta_1,\dots,\Delta_{r_0}>0$. Suppose also that $\overline{\Delta}:=\max_{1\leqslant r\leqslant r_0}\Delta_r$ satisfies $|\log\overline{\Delta}|\leqslant p_0/2$. Then there exists a positive constant $C$ depending only on $\alpha_1,\dots,\alpha_{r_0}$ and $A$ such that*

$$\left|\frac{P(W>x)}{P(Z>x)}-1\right| \leqslant C(1+x)\left\{\max_{1\leqslant r\leqslant r_0}(|\log\overline{\Delta}|+x^2)^{\alpha_r}\Delta_r+\overline{\Delta}\right\} \tag{2.7}$$

*for all $0\leqslant x\leqslant \sqrt{p_0}\wedge \min_{r=1,\dots,r_0}\Delta_r^{-1/(2\alpha_r+1)}$.*

We remark that because $\mathcal{W}_p(W,Z)$ increases in $p$, to apply Theorem 2.2, we only need to verify the upper bound on $\mathcal{W}_p(W,Z)$ for sufficiently large $p$, for example, for $p\geqslant 2$ in our applications.

*Proof of Theorem 2.2.* In this proof, we use $C$ to denote positive constant, which depends only on $\alpha_1,\dots,\alpha_{r_0}$ and $A$ and may be different in different expressions. First we prove the claim when $\overline{\Delta}<1/e$. Set

$$p=\log(1/\overline{\Delta})+\frac{x^2}{2},\quad \varepsilon=A\max_{1\leqslant r\leqslant r_0}p^{\alpha_r}\Delta_r e.$$

Because $|\log\overline{\Delta}|\leqslant p_0/2$ and $x\leqslant\sqrt{p_0}$ by assumption, we have $p\leqslant p_0$.

Without loss of generality, we may take $W$ and $Z$ so that $\|W-Z\|_p=\mathcal{W}_p(W,Z)$. Then

$$\begin{aligned}P(W>x)&=P(W>x,|W-Z|\leqslant\varepsilon)+P(W>x,|W-Z|>\varepsilon)\\&\leqslant P(Z>x-\varepsilon)+P(|W-Z|>\varepsilon)\\&=P(Z>x)+P(x-\varepsilon<Z\leqslant x)+P(|W-Z|>\varepsilon).\end{aligned}$$

Let $\phi(\cdot)$ denote the standard normal density function. Since

$$P(x-\varepsilon<Z\leqslant x)=\int_{x-\varepsilon}^{x}\phi(z)dz\leqslant\phi((x-\varepsilon)\vee 0)\varepsilon$$

and

$$\begin{aligned}P(|W-Z|>\varepsilon)&\leqslant(\|W-Z\|_p/\varepsilon)^p \qquad \text{(Markov's inequality)}\\&\leqslant(A\max_{1\leqslant r\leqslant r_0}p^{\alpha_r}\Delta_r/\varepsilon)^p \qquad \text{(by (2.6))}\\&=e^{-p}=\overline{\Delta}e^{-x^2/2},\end{aligned}$$

we obtain

$$P(W>x)\leqslant P(Z>x)+\phi((x-\varepsilon)\vee 0)\varepsilon+\overline{\Delta}e^{-x^2/2}.$$

Similarly, we deduce

$$\begin{aligned}P(Z>x)&=P(Z>x+\varepsilon)+P(x<Z\leqslant x+\varepsilon)\\&\leqslant P(W>x)+P(|W-Z|>\varepsilon)+P(x<Z\leqslant x+\varepsilon)\\&\leqslant P(W>x)+\phi(x)\varepsilon+\overline{\Delta}e^{-x^2/2}.\end{aligned}$$

Consequently, we obtain

$$|P(W > x) - P(Z > x)| \leqslant \phi((x-\varepsilon)\vee 0)\varepsilon + \overline{\Delta}e^{-x^2/2}. \tag{2.8}$$

Noting that $p^{\alpha_r} \leqslant C(\{\log(1/\overline{\Delta})\}^{\alpha_r} + x^{2\alpha_r})$ and $x \leqslant \Delta_r^{-1/(2\alpha_r+1)}$ for all $r$, we obtain

$$\begin{aligned}\varepsilon &\leqslant C \max_{1\leqslant r\leqslant r_0} \Delta_r(\{\log(1/\overline{\Delta})\}^{\alpha_r} + x^{2\alpha_r}) \\ &\leqslant C \max_{1\leqslant r\leqslant r_0} \Delta_r(\{\log(1/\Delta_r)\}^{\alpha_r} + \Delta_r^{-2\alpha_r/(2\alpha_r+1)}) \\ &\leqslant C \max_{1\leqslant r\leqslant r_0} \Delta_r^{1-2\alpha_r/(2\alpha_r+1)} = C \max_{1\leqslant r\leqslant r_0} \Delta_r^{1/(2\alpha_r+1)},\end{aligned} \tag{2.9}$$

where the third inequality follows from the inequality $\log t \leqslant \beta t^{1/\beta}$ for all $t > 1$ and $\beta > 0$. In particular, we have $x\varepsilon \leqslant C$ by the assumption on $x$. Now, if $x \geqslant \varepsilon$, we have

$$\phi((x-\varepsilon)\vee 0) \leqslant \phi(x)e^{x\varepsilon} \leqslant C\phi(x).$$

Birnbaum's inequality yields

$$\frac{\phi(x)}{P(Z > x)} \leqslant \frac{2}{\sqrt{4+x^2}-x} = \frac{\sqrt{4+x^2}+x}{2} \leqslant 1 + x. \tag{2.10}$$

Hence, by (2.8),

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C(1+x)(\varepsilon + \overline{\Delta}) \leqslant C(1+x)\{\max_{1\leqslant r\leqslant r_0} (|\log\overline{\Delta}| + x^2)^{\alpha_r}\Delta_r + \overline{\Delta}\},$$

where the second inequality follows by the definition of $\varepsilon$. In the meantime, if $x \leqslant \varepsilon$, we have

$$\frac{1}{P(Z > x)} \leqslant \sqrt{2\pi}(1+\varepsilon)e^{\varepsilon^2/2} \leqslant C,$$

where the first inequality follows by (2.10) and the second by (2.9) (recall that we assume $\overline{\Delta} = \max_{1\leqslant r\leqslant r_0} \Delta_r < 1/e$). Combining this with (2.8) gives

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C(\varepsilon + \overline{\Delta}) \leqslant C(1+x)\{\max_{1\leqslant r\leqslant r_0} (|\log\overline{\Delta}| + x^2)^{\alpha_r}\Delta_r + \overline{\Delta}\}.$$

So we complete the proof of (2.7).

It remains to prove (2.7) when $\overline{\Delta} \geqslant 1/e$. In this case, we have $\Delta_r \geqslant 1/e$ for some $r$, so $x \leqslant \Delta_r^{-1/(2\alpha_r+1)} \leqslant e^{1/(2\alpha_r+1)} \leqslant e$. Thus,

$$\frac{1}{P(Z > x)} \leqslant (1+e)\sqrt{2\pi e^{e^2}}$$

by (2.10). Hence (2.7) holds with $C \geqslant e(1+e)\sqrt{2\pi e^{e^2}}$. □

## 2.3 Sums of independent random variables

Finally, we illustrate our approach in the classical setting of sums of independent random variables.

Let $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, where $\{X_1, \ldots, X_n\}$ are independent with $\mathbb{E}X_i = 0$ for all $i$ and $\mathrm{Var}(W) = 1$. Suppose

$$b := \max_{1\leqslant i\leqslant n} \|X_i\|_{\psi_1} < \infty, \tag{2.11}$$

where $\|\cdot\|_{\psi_1}$ is the Orlicz norm defined in (1.4). This is equivalent to $b$ being the smallest positive constant such that $\mathbb{E}e^{|X_i|/b} \leqslant 2$ for all $i$. Let $Z \sim N(0,1)$. To apply Theorem 2.1, we construct an exchangeable pair (which is standard in Stein's method) as follows. Let $I$ be a uniform random index from $\{1, \ldots, n\}$ and independent of everything else. Let $\{X_1', \ldots, X_n'\}$ be an independent copy of $\{X_1, \ldots, X_n\}$. Let

$$W' = W - \frac{1}{\sqrt{n}} X_I + \frac{1}{\sqrt{n}} X_I' =: W + D.$$

Let $\mathcal{G} = \sigma(X_1, \ldots, X_n)$. It is straightforward to verify that

$$\mathbb{E}(D|\mathcal{G}) = -\frac{W}{n}.$$

Therefore, we can apply Theorem 2.1 with $R = 0$ and $\lambda = 1/n$ to bound $\mathcal{W}_p(W, Z)$.

We have

$$\|R_t\|_p \leqslant \| \sum_{i=1}^n Y_i 1_{\{|Y_i| > \eta_t(p)\}} \|_p, \qquad \|E\|_p \leqslant \| \sum_{i=1}^n (Y_i^2 - \mathbb{E}Y_i^2) \|_p,$$

and

$$\begin{aligned}\lambda^{-1} \| \mathbb{E}[D^4 1_{\{|D| \leqslant \eta_t(p)\}} | \mathcal{G}] \|_p &\leqslant \| \sum_{i=1}^n Y_i^4 1_{\{|Y_i| \leqslant \eta_t(p)\}} \|_p \\ &\leqslant \sum_{i=1}^n \mathbb{E} Y_i^4 + \| \sum_{i=1}^n (Y_i^4 1_{\{|Y_i| \leqslant \eta_t(p)\}} - \mathbb{E}[Y_i^4 1_{\{|Y_i| \leqslant \eta_t(p)\}}]) \|_p,\end{aligned}$$

where $Y_i = (X_i' - X_i)/\sqrt{n}$. We employ the following lemma to bound these quantities. See Kuchibhotla and Chakrabortty (2022, Theorem 3.1 and Remark 3.1) for a related result in dimension one and the literature on such concentration inequalities for sub-Weibull distributions.

**Lemma 2.1.** Let $\xi_1, \ldots, \xi_n$ be independent random vectors in $\mathbb{R}^d$ such that $\max_{i=1,\ldots,n} \|\xi_i\|_{\psi_\alpha} \leqslant M$ for some $M > 0$ and $\alpha \in (0, 1]$. Then, there is a constant $C_\alpha > 0$ depending only on $\alpha$ such that, for any $p \geqslant 2$ and any real numbers $a_1, \ldots, a_n$,

$$\left\| \sum_{i=1}^n a_i (\xi_i - \mathbb{E}\xi_i) \right\|_p \leqslant C_\alpha M \left( \sqrt{p \sum_{i=1}^n a_i^2} + p^{1/\alpha} \max_{1 \leqslant i \leqslant n} |a_i| \right).$$

*Proof.* First, by symmetrization, we have

$$\left\| \sum_{i=1}^n a_i (\xi_i - \mathbb{E}\xi_i) \right\|_p \leqslant 2 \left\| \sum_{i=1}^n a_i \epsilon_i \xi_i \right\|_p,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. Rademacher variables independent of everything else. Next, let $\zeta$ be a symmetric random variable such that $P(|\zeta| > t) = e^{-t^\alpha}$ for all $t \geqslant 0$. Then we have $P(|\epsilon_i \xi_i| > t) \leqslant 2 \exp(-(t/M)^\alpha) = 2P(M|\zeta| > t)$ for all $i = 1, \ldots, n$ and $t > 0$. Thus, by Theorem 3.2.2 in Kwapień and Woyczyński (1992),

$$P\left( \left| \sum_{i=1}^n a_i \epsilon_i \xi_i \right| > t \right) \leqslant 48 P\left( 6M \left| \sum_{i=1}^n a_i \zeta_i \right| > t \right)$$

for any $t > 0$, where $\zeta_1, \ldots, \zeta_n$ are independent copies of $\zeta$. This particularly implies that

$$\left\| \sum_{i=1}^n a_i \epsilon_i \xi_i \right\|_p \leqslant CM \left\| \sum_{i=1}^n a_i \zeta_i \right\|_p.$$

Finally, by Corollary 1.2 in Bogucki (2015),

$$\left\|\sum_{i=1}^n a_i\zeta_i\right\|_p \leqslant L_\alpha \left(\sqrt{p\sum_{i=1}^n a_i^2} + p^{1/\alpha} \max_{i=1,\dots,n} |a_i|\right),$$

where $L_\alpha > 0$ depends only on $\alpha$. All together, we obtain the desired result. □

Now, for any $r \geqslant 1$, from $b := \max_{1\leqslant i\leqslant n} \|X_i\|_{\psi_1}$ and the equivalence of sub-exponential tails and linear growth of $L^r$-norms (cf. Vershynin (2018, Proposition 2.7.1)),

$$\|Y_i 1_{\{|Y_i|>\eta_t(p)\}}\|_r \leqslant \eta_t^{-1}(p)(\mathbb{E}Y_i^{2r})^{1/r} \leqslant Cr^2\eta_t^{-1}(p)b^2/n, \qquad \|Y_i^2\|_r \leqslant Cr^2b^2/n,$$

and

$$\|Y_i^4 1_{\{|Y_i|\leqslant\eta_t(p)\}}\|_r \leqslant \eta_t^2(p)\|Y_i^2\|_r \leqslant Cr^2\eta_t^2(p)b^2/n.$$

Hence, $\|Y_i 1_{\{|Y_i|>\eta_t(p)\}}\|_{\psi_{1/2}} \leqslant C\eta_t^{-1}(p)b^2/n$, $\|Y_i^2\|_{\psi_{1/2}} \leqslant Cb^2/n$ and $\|Y_i^4 1_{\{|Y_i|\leqslant\eta_t(p)\}}\|_{\psi_{1/2}} \leqslant C\eta_t^2(p)b^2/n$. So we obtain by Lemma 2.1, for $p \geqslant 2$,

$$\int_0^\infty e^{-t}\|R_t\|_p dt \leqslant C\frac{\sqrt{np}+p^2}{n}\int_0^\infty \frac{e^{-t}\sqrt{p}}{\sqrt{e^{2t}-1}}b^2 dt \leqslant C(\frac{p}{\sqrt{n}} + \frac{p^{5/2}}{n})b^2,$$

$$\sqrt{p}\|E\|_p \leqslant C(\frac{p}{\sqrt{n}} + \frac{p^{5/2}}{n})b^2,$$

and

$$\begin{aligned}
&\int_0^\infty e^{-t}\min\left\{\frac{1}{\eta_t(p)}, \frac{\|\mathbb{E}[D^4 1_{\{|D|\leqslant\eta_t(p)\}}|\mathcal{G}]\|_p}{\lambda\eta_t^3(p)}\right\}dt \\
&\leqslant \int_0^\infty e^{-t}\min\left\{\frac{\sqrt{p}}{\sqrt{e^{2t}-1}}, \frac{Cp^{3/2}b^4}{n(e^{2t}-1)^{3/2}}\right\}dt + C\int_0^\infty e^{-t}\frac{p/\sqrt{n}+p^{5/2}/n}{\sqrt{e^{2t}-1}}b^2 dt \\
&\leqslant C(\frac{p}{\sqrt{n}} + \frac{p^{5/2}}{n})b^2.
\end{aligned}$$

Here, we evaluate the integrals as in the proof of Theorem 2.1. Consequently, from (2.2),

$$\mathcal{W}_p(W,Z) \leqslant C(\frac{p}{\sqrt{n}} + \frac{p^{5/2}}{n})b^2, \quad \forall\, p \geqslant 2. \tag{2.12}$$

Note that $\mathrm{Var}(W) = 1 \leqslant Cb^2$. Therefore, we can apply Theorem 2.2 with $r_0 = \alpha_1 = 1$, $\Delta_1 = b^2/\sqrt{n}$ and $p_0 = (\sqrt{n}/b^2)^{2/3}$, which implies that:

**Corollary 2.2.** Let $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, where $\{X_1,\dots,X_n\}$ are independent with $\mathbb{E}X_i = 0$ for all $i$, $\mathrm{Var}(W) = 1$ and $b := \max_{1\leqslant i\leqslant n}\|X_i\|_{\psi_1} < \infty$. Then there exist positive absolute constants $c$ and $C$ such that

$$\left|\frac{P(W>x)}{P(Z>x)} - 1\right| \leqslant C\frac{(1+x)(1+|\log(n/b^4)|+x^2)b^2}{\sqrt{n}}$$

for all $0 \leqslant x \leqslant (n/b^4)^{1/6}$ and $\frac{b^2}{\sqrt{n}} \leqslant c$.

**Remark 2.1.** Corollary 2.2 recovers the bound (1.1) when $x \geqslant \sqrt{\log n}$. It seems impossible to avoid the $\log n$ term using our approach because such a term will appear even if we only aim to bound the Kolmogorov distance using $p$-Wasserstein bounds and a smoothing argument.

An inspection of the proof shows that we can replace the range of $x$ by $0 \leqslant x \leqslant c_0(n/b^4)^{1/6}$ with any absolute constant $c_0$ (the constant $C$ will then depend on $c_0$). Because our primary interests are vanishing relative errors and the order of magnitude, we will not worry about such absolute constants and state our results in a form that we find convenient.

# 3 Applications to Cramér-type moderate deviations in dimension one

In this section, we provide more applications in dimension one, including the combinatorial CLT, Wiener chaos and homogeneous sums.

## 3.1 Combinatorial CLT

Let $\mathbb{X} = \{X_{ij}, 1 \leqslant i,j \leqslant n\}$ be an $n \times n$ array of independent random variables where $n \geqslant 2$, $\mathbb{E}X_{ij} = c_{ij}$, $\mathrm{Var}(X_{ij}) = \sigma_{ij}^2 \geqslant 0$. Assume without loss of generality that (cf. Remark 1.3 of Chen and Fang (2015))

$$c_{i\cdot} = c_{\cdot j} = 0,$$

where $c_{i\cdot} = \sum_{j=1}^n c_{ij}/n$, $c_{\cdot j} = \sum_{i=1}^n c_{ij}/n$. Let $\pi$ be a uniform random permutation of $\{1,\dots,n\}$, independent of $\mathbb{X}$, and let

$$S = \sum_{i=1}^n X_{i\pi(i)}. \tag{3.1}$$

It is known that $\mathbb{E}(S) = 0$ and (cf. Theorem 1.1 of Chen and Fang (2015))

$$B_n^2 := \mathrm{Var}(S) = \frac{1}{n-1}\sum_{i,j=1}^n c_{ij}^2 + \frac{1}{n}\sum_{i,j=1}^n \sigma_{ij}^2, \tag{3.2}$$

$$\sup_{x\in\mathbb{R}} |P(W \leqslant x) - P(Z \leqslant x)| \leqslant \frac{C}{n}\sum_{i,j=1}^n \mathbb{E}\big|\frac{X_{ij}}{B_n}\big|^3, \tag{3.3}$$

where

$$W = \frac{S}{B_n}, \tag{3.4}$$

and $Z \sim N(0,1)$. Cramér-type moderate deviation results were obtained by Frolov (2022) and Liu and Zhang (2021). Here, we use our approach to prove a version of such moderate deviation results.

**Theorem 3.1.** *Under the above setting, assume*

$$b := \max_{1\leqslant i,j\leqslant n} \|X_{ij}\|_{\psi_1} < \infty. \tag{3.5}$$

*Then there exist positive absolute constants $c$ and $C$ such that, for*

$$\Delta := \frac{n^{1/2}b^2}{B_n^2} \leqslant c, \quad 0 \leqslant x \leqslant \Delta^{-1/3},$$

*we have*

$$\left|\frac{P(W>x)}{P(Z>x)} - 1\right| \leqslant C(1+x)(1+|\log\Delta| + x^2)\Delta.$$

**Remark 3.1.** Because Frolov (2022)'s result is stated under a different condition and he did not provide a rate of convergence, here we only compare our result with that in Liu and Zhang (2021). In our notation, their bound is $C(1+x^3)\frac{n^{1/2}b^2}{B_n^2}(\frac{n^{1/2}b}{B_n})^5$. From (3.2), we have $B_n^2 \leqslant Cnb^2$ and $B_n^2$ in general can be of smaller order than $nb^2$. Therefore, except for the logarithmic term in the error rate, our bound is in general better.

We prove Theorem 2.1 via the following $p$-Wasserstein bound between $W$ and $Z$.

**Proposition 3.1.** Under the assumptions of Theorem 3.1, there exists a positive absolute constant $C$ such that

$$\mathcal{W}_p(W,Z) \leqslant C(\frac{p\sqrt{n}}{B_n^2} + \frac{p^{5/2}}{B_n^2})b^2 \quad \forall\, p \geqslant 2. \tag{3.6}$$

In the following, we prove Theorem 3.1 using Proposition 3.1. The proof of Proposition 3.1 is deferred to Section 7.2.

*Proof of Theorem* 3.1. We apply Theorem 2.2 with $r_0 = \alpha_1 = 1$ and

$$\Delta_1 := \Delta = \frac{n^{1/2}b^2}{B_n^2},\ p_0 = \Delta_1^{-2/3} = (\frac{B_n^2}{n^{1/2}b^2})^{2/3}.$$

The conditions in Theorem 2.2 are satisfied by choosing $c$ in the statement of Theorem 3.1 to be sufficiently small and using $B_n^2 \leqslant Cnb^2$ from (3.2) to reduce the bound (3.6) to $Cpn^{1/2}b^2/B_n^2$ for $2 \leqslant p \leqslant p_0$. □

### 3.2 Moderate deviations on Wiener chaos

Let $X$ be an isonormal Gaussian process over a real separable Hilbert space $\mathfrak{H}$. Given an integer $q \geqslant 2$, we consider the $q$-th multiple Wiener–Itô integral $W = I_q(f)$ of $f \in \mathfrak{H}^{\odot q}$ with respect to $X$. Here, $\mathfrak{H}^{\odot q}$ denotes the $q$-th symmetric tensor power of $\mathfrak{H}$. Here and below, we use standard concepts and notations in Malliavin calculus. We refer to Nourdin and Peccati (2012) for all unexplained notations.

We assume $\mathrm{Var}(W) = q!\|f\|^2_{\mathfrak{H}^{\otimes q}} = 1$ for simplicity. The celebrated fourth moment theorem states that (cf. Theorem 5.2.6 in Nourdin and Peccati (2012))

$$\sup_{x\in\mathbb{R}} |P(W \leqslant x) - P(Z \leqslant x)| \leqslant \sqrt{\frac{q-1}{3q}(\mathbb{E}W^4 - 3)},$$

where $Z \sim N(0,1)$. Schulte and Thäle (2016) obtained a corresponding Cramér-type moderate deviation result. Here, we use our approach to prove a version of such moderate deviation results.

To state our result, we need to introduce mixed injective norms of elements in $\mathfrak{H}^{\odot q}$ which were originally introduced in Latała (2006) (see also Lehec (2011)). A *partition* of $[q]$ is a collection of nonempty disjoint sets $\{J_1, \dots, J_k\}$ such that $[q] = \bigcup_{l=1}^k J_l$. We denote by $\Pi_q$ the set of partitions of $[q]$. For any $h \in \mathfrak{H}^{\odot q}$ and $\mathcal{J} = \{J_1, \dots, J_k\} \in \Pi_q$, define

$$\|h\|_{\mathcal{J}} := \sup\{\langle h, u_1 \otimes \cdots \otimes u_k\rangle_{\mathfrak{H}^{\otimes q}} : u_l \in \mathfrak{H}^{\otimes |J_l|}, \|u_l\|_{\mathfrak{H}^{\otimes |J_l|}} \leqslant 1, l = 1, \dots, k\}.$$

In the remainder of this section, $C_q$ denotes a positive constant, which depends only on $q$ and may be different in different expressions.

**Theorem 3.2.** *Under the above setting, let*

$$\overline{\Delta} := \max_{r\in[q-1]} \max_{\mathcal{J}\in\Pi_{2q-2r}} \|f\widetilde{\otimes}_r f\|_{\mathcal{J}},$$

*where $f\widetilde{\otimes}_r f$ denotes the symmetrization of $f \otimes_r f$ with $\otimes_r$ the $r$-th contraction operator (cf. Nourdin and Peccati (2012, Eq. (B.3.1) & (B.4.4))). If*

$$0 \leqslant x \leqslant \min_{r\in[q-1]} \min_{\mathcal{J}\in\Pi_{2q-2r}} \|f\widetilde{\otimes}_r f\|_{\mathcal{J}}^{-1/(|\mathcal{J}|+2)}, \tag{3.7}$$

*then*

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C_q(1 + x)\left\{\max_{r\in[q-1]} \max_{\mathcal{J}\in\Pi_{2q-2r}} (|\log \overline{\Delta}| + x^2)^{\frac{1+|\mathcal{J}|}{2}} \|f\widetilde{\otimes}_r f\|_{\mathcal{J}} + \overline{\Delta}\right\}. \tag{3.8}$$

The proof of Theorem 3.2 is deferred to Section 7.3.

**Remark 3.2** (Optimality on the range of $x$)**.** Condition (3.7) is sharp when $q = 2$. To see this, assume that $\mathfrak{H}$ is infinite-dimensional and let $(e_i)_{i=1}^{\infty}$ be an orthonormal basis of $\mathfrak{H}$. Taking $f = \frac{1}{\sqrt{2n}}\sum_{i=1}^{n} e_i^{\otimes 2}$, we obtain $W = \frac{1}{\sqrt{2n}}\sum_{i=1}^{n}(X(e_i)^2 - 1)$ (cf. Theorem 2.7.7 in Nourdin and Peccati (2012)). Since $X(e_i)$ are i.i.d. standard normal variables, $W$ is a normalized sum of i.i.d. random variables with the centered $\chi^2$-distribution with 1 degree of freedom. Meanwhile, since $|\langle \sum_{i=1}^{n} e_i^{\otimes 2}, u_1 \otimes u_2\rangle_{\mathfrak{H}^{\otimes 2}}| \leqslant \|u_1\|_{\mathfrak{H}}\|u_2\|_{\mathfrak{H}}$ for any $u_1, u_2 \in \mathfrak{H}$ by Bessel's inequality and the equality can be attained,

$$\|f\widetilde{\otimes}_1 f\|_{\{1\},\{1\}} = \frac{1}{2n}\left\|\sum_{i=1}^{n} e_i^{\otimes 2}\right\|_{\{1\},\{1\}} = \frac{1}{2n}.$$

Also,

$$\|f\widetilde{\otimes}_1 f\|_{\{1,2\}} = \frac{1}{2n}\left\|\sum_{i=1}^{n} e_i^{\otimes 2}\right\|_{\mathfrak{H}^{\otimes 2}} = \frac{1}{2\sqrt{n}}.$$

Thus, (3.7) is rewritten as $0 \leqslant x \leqslant \min\{(2n)^{1/4}, (4n)^{1/6}\}$. In view of Theorem 2 in Petrov (1975, Chapter VIII), this condition is sharp to obtain a bound like (3.8).

When $q > 2$, it is unclear whether (3.7) is sharp or not. By an analogous argument to the above but using Theorem 2 in Linnik (1961), we can show that $x$ must satisfy $x = O(\underline{\Delta}^{-1/(2q-2)-\varepsilon})$ for any $\varepsilon > 0$, where $\underline{\Delta} := \|f\widetilde{\otimes}_1 f\|_{\{1\},\dots,\{2q-2\}}$. However, (3.7) requires at least $x = O(\underline{\Delta}^{-1/(2q)})$.

Next, we make connections to the fourth moment theorem. For any $\mathcal{J} \in \Pi_{2q-2r}$ with $r \in [q-1]$, we have $|\mathcal{J}| \leqslant 2q - 2r \leqslant 2q - 2$ and

$$\|f\widetilde{\otimes}_r f\|_{\mathcal{J}} \leqslant \|f\widetilde{\otimes}_r f\|_{\mathfrak{H}^{\otimes(2q-2r)}} \leqslant \|f \otimes_r f\|_{\mathfrak{H}^{\otimes(2q-2r)}} \leqslant \|f\|_{\mathfrak{H}^{\otimes q}}^2 = 1/q!,$$

where the first inequality is from $\|h\|_{\mathcal{J}} \leqslant \|h\|_{\mathfrak{H}^{\otimes(2q-2r)}}$ for any $h \in \mathfrak{H}^{\odot(2q-2r)}$, the second inequality is from the definition of symmetrization and the triangle inequality, the third inequality follows by the Cauchy–Schwarz inequality. Therefore, noting that the function $(0,1) \ni \delta \mapsto \delta(y + |\log\delta|)^{(2q-1)/2} \in (0,\infty)$ is increasing for any $y \geqslant (2q-1)/2$, we particularly obtain by Theorem 3.2

$$\begin{aligned}\left|\frac{P(W > x)}{P(Z > x)} - 1\right| &\leqslant C_q(1+x)(1 + |\log\overline{\Delta}| + x^2)^{(2q-1)/2}\overline{\Delta} \\ &\leqslant C_q(1+x)(1 + |\log\Delta| + x^2)^{(2q-1)/2}\Delta\end{aligned} \tag{3.9}$$

for all $0 \leqslant x \leqslant \Delta^{-1/(2q)}$, where

$$\Delta := \max_{r\in[q-1]} \|f \otimes_r f\|_{\mathfrak{H}^{\otimes(2q-2r)}}.$$

From Nourdin and Peccati (2012, Eq. (5.2.6)), we have $\Delta \leqslant C_q\sqrt{\mathbb{E}W^4 - 3}$. Therefore, we obtain a Cramér-type moderate deviation result for the fourth moment theorem:

**Corollary 3.1.** Under the above setting,

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C_q(1+x)(1 + |\log\kappa_4(W)| + x^2)^{(2q-1)/2}\sqrt{\kappa_4(W)}$$

for all $0 \leqslant x \leqslant \kappa_4(W)^{-1/(4q)}$, where $\kappa_4(W) = \mathbb{E}W^4 - 3$ is the fourth cumulant of $W$.

**Remark 3.3** (Comparison with Schulte and Thäle (2016))**.** Using the method of cumulants, Schulte and Thäle (2016) give in their Theorem 5 a Cramér-type moderate deviation result for multiple Wiener-Itô integrals in the following form: Let

$$\alpha(q) := \begin{cases}(q+2)/(3q+2) & \text{if } q \text{ is even,}\\ (q^2-q-1)/(q(3q-5)) & \text{if } q \text{ is odd.}\end{cases}$$

Then, there are constants $c_0, c_1, c_2 > 0$ depending only on $q$ such that, for $\Delta^{-\alpha(q)} \geqslant c_0$ and $0 \leqslant x \leqslant c_1 \Delta^{-\alpha(q)/(q-1)}$,

$$\left| \log \frac{P(W > x)}{P(Z > x)} \right| \leqslant c_2(1 + x^3)\Delta^{\alpha(q)/(q-1)}. \tag{3.10}$$

On the other hand, by the inequality $|\log(1 + y)| \leqslant 2|y|$ for $|y| \leqslant 1/2$, our simplified bound (3.9) implies that there are constants $c'_0, c'_1, c'_2 > 0$ depending only on $q$ such that, for $\Delta \leqslant c'_0$ and $0 \leqslant x \leqslant c'_1 \Delta^{-1/(2q)}$,

$$\left| \log \frac{P(W > x)}{P(Z > x)} \right| \leqslant c'_2(1 + x)(1 + |\log \Delta| + x^2)^{(2q-1)/2}\Delta.$$

We compare this bound with (3.10). Note that $\Delta \leqslant 1$. Then, since we can easily check that $\alpha(q) + 1/(2q) < 1/2$ if and only if $q \geqslant 5$, Theorem 5 in Schulte and Thäle (2016) imposes a weaker condition on $x$ than ours when $q < 5$. However, note that we need $x^3\Delta^{\alpha(q)/(q-1)} = o(1)$ to get a vanishing bound in (3.10). This condition is always stronger than our condition $x^{2q}\Delta = o(1)$ because $\alpha(q) \leqslant 1/2$. Moreover, under the condition $x^3\Delta^{\alpha(q)/(q-1)} = o(1)$, we always have $x^{2q}\Delta = o(x^3\Delta^{\alpha(q)/(q-1)})$ since

$$\frac{\alpha(q)}{3(q-1)} \leqslant \frac{q - 1 - \alpha(q)}{(2q-3)(q-1)}.$$

So our bound always gives a better rate of convergence to 0 than (3.10).

### 3.3 Homogeneous sums

Let $X_1, \dots, X_n$ be independent random variables with mean 0 and variance 1. We consider a multilinear homogeneous sum of these variables, i.e. a random variable of the form

$$W = \sum_{i_1,\dots,i_q=1}^{n} f(i_1, \dots, i_q)X_{i_1} \cdots X_{i_q},$$

where $q \geqslant 2$ and $f : [n]^q \to \mathbb{R}$ is a symmetric function with vanishing diagonals (i.e. $f(i_1, \dots, i_q) = 0$ whenever $i_r = i_s$ for some indices $r \neq s$). $W$ has mean 0 by assumption. For simplicity, we assume that $W$ has variance 1, i.e.

$$\operatorname{Var}[W] = q! \sum_{i_1,\dots,i_q=1}^{n} f(i_1, \dots, i_q)^2 = 1.$$

$W$ is a prominent example of degenerate $U$-statistics of order $q$, and limit theorems for such statistics have been well-studied in the literature. In particular, the prominent work of de Jong (1990) established the following sufficient conditions for the asymptotic normality: $W$ converges in law to $N(0, 1)$ if the following conditions are satisfied:

(i) The fourth cumulant of $W$ converges to 0. That is, $\mathbb{E}W^4$ converges to 3.

(ii) The maximal influence

$$\mathcal{M}(f) := \max_{i \in [n]} \sum_{i_2,\dots,i_q=1}^{n} f(i, i_2, \dots, i_q)^2$$

converges to 0.

Corresponding absolute error bounds were investigated in e.g. Nourdin, Peccati and Reinert (2010); Döbler and Peccati (2017) and Fang and Koike (2022). For example, Corollary 2.1 in Fang and Koike (2022) gives the following optimal 1-Wasserstein bound (throughout this section, $C_q$ denotes a constant, which depends only on $q$ and may be different in different expressions):

$$\mathcal{W}_1(W,Z) \leqslant C_q \sqrt{|\mathbb{E}W^4 - 3| + \left(\max_{i\in[n]} \mathbb{E}X_i^4\right)^q \mathcal{M}(f)},$$

where $Z \sim N(0,1)$. However, to our knowledge, no relative error bound for this type of CLT is available in the literature (but see Remark 3.5). Using our approach, we can obtain such a bound as follows:

**Theorem 3.3.** *Under the above setting, assume that there exits a constant $K \geqslant 1$ such that $\|X_i\|_{\psi_2} \leqslant K$ for all $i \in [n]$. Let*

$$M := \max_{i\in[n]} \mathbb{E}X_i^4, \quad \Delta := K^{2q}\sqrt{|\mathbb{E}W^4 - 3| + M^q \mathcal{M}(f)(1 \vee |\log \mathcal{M}(f)|^{2q-2})},$$

*and assume $\Delta < 1$. Then, for all $0 \leqslant x \leqslant \Delta^{-\frac{1}{2q+1}}$,*

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C_q(1+x)(|\log \Delta| + x^2)^q \Delta. \tag{3.11}$$

Although Theorem 3.3 is the first moderate deviation result corresponding to de Jong (1990)'s CLT for homogeneous sums in the literature, the bound seems suboptimal. In fact, for the case of $q = 2$ and $|X_i| \leqslant K$ a.s., we can obtain the following optimal result. Its proof is a straightforward but very tedious modification of the proof of Theorem 3.3 and we leave it to the appendix. The proof technique would work for general $q$ if we introduce appropriate notations, but computation of mixed injective norms becomes extremely complicated. We do not pursue it further in this paper.

**Theorem 3.4.** *Under the above setting, assume that $q = 2$ and there exists a constant $K \geqslant 1$ such that $|X_i| \leqslant K$ a.s. for all $i \in [n]$. Set $F = (f(i,j))_{1\leqslant i,j\leqslant n}$. Then, there exists a positive absolute constant $C$ such that*

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant CK^4(1+x)(|\log \|F\|_{op}| + x^2)\|F\|_{op} \tag{3.12}$$

*for all $0 \leqslant x \leqslant \|F\|_{op}^{-1/3}$.*

**Remark 3.4** (Optimality of Theorem 3.4)**.** The error bound and the range of $x$ in Theorem 3.4 are optimal. To see this, assume that $n$ is even and $X_i$ are i.i.d. with $\mathbb{E}X_i^3 \neq 0$. Define the function $f$ as

$$f(i,j) = \begin{cases} 1/\sqrt{2n} & \text{if } \{i,j\} = \{2k-1, 2k\} \text{ for some positive integer } k, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$W = \sum_{k=1}^{n/2} \frac{X_{2k-1}X_{2k} + X_{2k}X_{2k-1}}{\sqrt{2n}} = \frac{1}{\sqrt{n/2}} \sum_{k=1}^{n/2} X_{2k-1}X_{2k}.$$

So $W$ is a normalized sum of $n/2$ i.i.d. random variables with mean 0 and variance 1. Since $\mathbb{E}[X_{2k-1}^3 X_{2k}^3] = (\mathbb{E}X_1^3)^2 \neq 0$, we need the condition $x = o(n^{1/6})$ to get a vanishing relative error bound, and in this case the optimal bound is of the form $c(1+x^3)/\sqrt{n}$ for some constant $c > 0$. This result is recovered by Theorem 3.4 when $x \geqslant \sqrt{\log n}$ since $\|F\|_{op} = O(n^{-1/2})$.

**Remark 3.5** (Comparison with Saulis and Statulevičius (1991))**.** Saulis and Statulevičius (1991) give Cramér-type moderate deviation results for polynomial forms of independent random variables in their Theorem 5.1 using the method of cumulants. Their result is in terms of

$$\max_{\substack{r,s\in[q]\\ r+s=q}} \sqrt{\left(\max_{i_1,\dots,i_r\in[n]} \sum_{i_{r+1},\dots,i_q=1}^{n} |f(i_1,\dots,i_q)|\right)\left(\max_{i_1,\dots,i_s\in[n]} \sum_{i_{s+1},\dots,i_q=1}^{n} |f(i_1,\dots,i_q)|\right)}$$

and is not directly comparable with the fourth-moment-influence bound in Theorem 3.3. Therefore, we only compare their result with ours in the setting of Theorem 3.4. Suppose that $X_1,\dots,X_n$ are i.i.d. Then, under the assumptions of Theorem 3.4, Saulis and Statulevičius (1991, Theorem 5.1) leads to a bound of the form $CK^4(1+x^3)\|F\|_{op,\infty}$, where $\|F\|_{op,\infty}$ is the $\ell_\infty$-operator norm of $F$: $\|F\|_{op,\infty} := \max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|f(i,j)|$. Since $\|F\|_{op} \leqslant \|F\|_{op,\infty}$ by Theorem 5.6.9 in Hohn and Johnson (2013), our bound is better except for the logarithmic term in the error rate.

Theorem 3.3 is a straightforward consequence of the following $p$-Wasserstein bound and Theorem 2.2:

**Proposition 3.2.** Under the assumptions of Theorem 3.3, for any $2 \leqslant p \leqslant \mathcal{M}(f)^{-1/2}$,

$$\mathcal{W}_p(W,Z) \leqslant C_q p^q \Delta.$$

The proof of Proposition 3.2 is deferred to Section 7.4.

*Proof of Theorem* 3.3. We first note that $\mathcal{M}(f) \leqslant \sum_{i_1,\dots,i_q=1}^{n} f(i_1,\dots,i_q)^2 = 1/q! \leqslant 1/2$. We apply Theorem 2.2 with $r_0 = 1$, $\alpha_1 = q$, $\Delta_1 = \Delta$ and $p_0 = \mathcal{M}(f)^{-1/2}$. Then, it remains to check $|\log\Delta| \leqslant p_0/2$ and $\sqrt{p_0} \geqslant \Delta^{-1/(2q+1)}$. Since $M \geqslant (\mathbb{E}X_1^2)^2 = 1$, we have $\Delta \geqslant \sqrt{\mathcal{M}(f)}$. This and the assumption $\Delta < 1$ give the desired result. □

## 4 Moderate deviations in multi-dimensions

In this section, we study moderate deviations in multi-dimensions. We first apply Theorem 2.1 to obtain a $p$-Wasserstein bound for multivariate normal approximation of sums of independent random vectors. All the proofs for the results in this section are deferred to Section 7.5.

**Theorem 4.1.** *Let $W = n^{-1/2}\sum_{i=1}^{n} X_i \in \mathbb{R}^d$, where $\{X_1,\dots,X_n\}$ are independent, $\mathbb{E}(X_i) = 0$ for all $i$, and $\mathrm{Var}(W) = I_d$. Suppose $\|X_i\|_{\psi_1} \leqslant b$ for all $1 \leqslant i \leqslant n$. Let $Z \sim N(0, I_d)$. Then, for any $p \geqslant 2$, we have*

$$\mathcal{W}_p(W,Z) \leqslant C\Big(\frac{pd^{1/4}}{\sqrt{n}} + \frac{p^{5/2}}{n}\Big)b^2. \tag{4.1}$$

**Remark 4.1** (Dimension dependence)**.** The dependence on the dimension $d$ of the bound (4.1) is suboptimal. In fact, when we assume $|X_i| \leqslant b$ a.s. instead of $\|X_i\|_{\psi_1} \leqslant b$ and $X_1,\dots,X_n$ are i.i.d., Theorem 1 in Bonis (2020) gives a bound of the form $C_p b\sqrt{d/n}$ with $C_p$ a constant depending only on $p$ (we can use Lemma 6.1 to bound the first term of the right hand side on Eq.(9) therein). In the meantime, since $b \geqslant \sqrt{\mathbb{E}[|X_i|^2]} = \sqrt{d}$, the right hand side of (4.1) is at least of order $O(b\sqrt{d^{3/2}/n})$. As we already remarked after Theorem 2.1, the proof of this theorem (and hence Theorem 4.1) uses the same strategy as in Bonis (2020), and we can indeed derive a bound of order $O(b\sqrt{d/n})$ from Theorem 2.1 under the boundedness assumption. We also remark that the order $O(b\sqrt{d/n})$ is optimal due to Proposition 1.2 in Zhai (2018).

We can use $p$-Wasserstein bounds to obtain moderate deviation results in the multi-dimensional setting. In the following theorem, we provide an analogous result as Theorem 2.2 for $|P(|W|>x)/P(|Z|>x)-1|$. For simplicity, we only state a result corresponding to $r_0=1$ in Theorem 2.2, which suffices for the applications we consider. We remark that our approach can be used to obtain upper bounds on $|P(W\notin A)/P(Z\notin A)-1|$ for more general convex sets $A\subset\mathbb{R}^d$ as long as we have a suitable control on $P(Z\in A^\varepsilon\backslash A^{-\varepsilon})/P(Z\notin A)$ for small $\varepsilon>0$, where $A^\varepsilon\backslash A^{-\varepsilon}$ contains all $x\in\mathbb{R}^d$ within distance $\varepsilon$ away from the boundary of $A$.

**Theorem 4.2.** *Let $W$ be a $d$-dimensional random vector, $d\geqslant 2$, and $Z\sim N(0,I_d)$. Suppose*

$$\mathcal{W}_p(W,Z)\leqslant Ap^\alpha\Delta \text{ for } 1\leqslant p\leqslant p_0$$

*with some constants $\alpha\geqslant 0$, $A>0$, $\Delta>0$, $|\log\Delta|\leqslant p_0/4$ and $\log(\kappa(d))\leqslant p_0/4$ with $\kappa(d):=2^{(d/2)-1}\Gamma(d/2)$. Suppose further that*

$$d(d\log d)^\alpha\Delta\leqslant B_1 \tag{4.2}$$

*and*

$$d\Delta|\log\Delta|^\alpha\leqslant B_2,\ \text{if } 0<\alpha\leqslant 1/2. \tag{4.3}$$

*Then there exists a positive constant $C_{A,\alpha,B_1,B_2}$ depending only on $\alpha$, $A$, $B_1$ and $B_2$ such that*

$$\left|\frac{P(|W|>x)}{P(|Z|>x)}-1\right|\leqslant C_{A,\alpha,B_1,B_2}(1+x)(|\log\Delta|+d\log d+x^2)^\alpha\Delta \tag{4.4}$$

*for all $0\leqslant x\leqslant\min\{\Delta^{-1/(2\alpha+1)},\sqrt{p_0}\}$.*

The following Cramér-type moderate deviation result for sums of independent random vectors is an easy consequence of Theorem 4.1, Theorem 4.2 with $\alpha=1$, $p_0=\Delta^{-2/3}$ and the fact that $d=\mathbb{E}|W|^2\leqslant Cb^2$.

**Theorem 4.3.** *Under the setting of Theorem 4.1 with $d\geqslant 2$, let*

$$\Delta:=\frac{d^{1/4}b^2}{\sqrt{n}}.$$

*Then there exist positive absolute constants $c$ and $C$ such that, for*

$$d^2(\log d)\Delta\leqslant c,\quad 0\leqslant x\leqslant\Delta^{-1/3},$$

*we have*

$$\left|\frac{P(|W|>x)}{P(|Z|>x)}-1\right|\leqslant C(1+x)(d\log d+|\log\Delta|+x^2)\Delta.$$

**Remark 4.2.** The result in Theorem 4.3 recovers the optimal range $0\leqslant x=o(n^{1/6})$ (cf. von Bahr (1967)) for the relative error to vanish. Although it is known that the error rate can be improved because of the symmetry of Euclidean balls, see, for example, von Bahr (1967) and Fang, Liu and Shao (2021), their proofs depend on the conjugate method, which relies heavily on the independence assumption. Our approach works for the dependent case (cf. Theorems 5.2 and 5.3).

## 5 Local dependence

A large class of random vectors that can be approximated by a normal distribution exhibits a local dependence structure. Roughly speaking, we assume that the random vector $W$ is a sum of a large number of random vectors $\{X_i\}_{i=1}^n$ and that each $X_i$ is independent of $\{X_j: j\notin A_i\}$ for a relatively small index set $A_i$. Variations of such local dependence structure and normal approximation results with absolute error bounds can be found in, e.g., Baldi and Rinott (1989), Barbour, Karoński and Ruciński (1989) and Chen and Shao (2004). Moderate deviation results (relative error bounds) under local dependence were recently obtained by Liu and Zhang (2021) in dimension one. See Remark 5.2 for a comparison.

Throughout this section, we assume $n \geqslant 2$.

### 5.1 Bounded case

We first provide a $p$-Wasserstein bound for multivariate normal approximation of sums of locally dependent, bounded random vectors.

**Theorem 5.1.** *Let $W = n^{-1/2}\sum_{i=1}^n X_i \in \mathbb{R}^d$ with $\mathbb{E}(X_i) = 0$ for all $i$ and $\mathrm{Var}(W) = I_d$. We assume that for each $i$, there is a neighborhood $A_i \subset \{1, \dots, n\}$ such that $X_i$ is independent of $\{X_j : j \notin A_i\}$. Assume further that for each $i$ and $j \in A_i$, there exists a second neighborhood $A_{ij}$ such that $\{X_i, X_j\}$ is independent of $\{X_k : k \notin A_{ij}\}$. Let*

$$B_{ij} := \{(k,l) : k \in \{1, \dots, n\}, l \in A_k, k \text{ or } l \in A_{ij}\}.$$

*Suppose*

$$|X_i| \leqslant b_n,\ |X_{ij}| \leqslant b_n',\ |A_i| \leqslant \theta_1,\ |B_{ij}| \leqslant \theta_2,$$

*where $X_{ij}$ denotes the $j$th component of $X_i$ and $|\cdot|$ denotes the cardinality when applied to a set. Then there exist positive absolute constants $c$ and $C$ such that, for*

$$2 \leqslant p \leqslant \min\{\frac{\theta_1}{\theta_2}, \frac{c}{\theta_1^2 b_n^2}\}n \tag{5.1}$$

*we have, with $Z \sim N(0, I_d)$,*

$$\mathcal{W}_p(W, Z) \leqslant Cp\left(\frac{d(\theta_1\theta_2)^{1/2}b_n'^2 + \theta_1^2 b_n^3 \log n}{\sqrt{n}}\right). \tag{5.2}$$

**Remark 5.1.** We will adapt the proof of Theorem 2.1 to prove Theorem 5.1 in Section 7.6. Without exchangeability, we can not use the symmetry trick in (6.8). Therefore, because of the integrability issue of $1/(e^{2t} - 1)$ for $t$ near 0, we get an additional logarithmic term in (5.2) (cf. Section 6.2).

Using Theorem 5.1 together with Theorems 2.2 and 4.2, we obtain the following moderate deviation result for sums of locally dependent, bounded random vectors.

**Theorem 5.2.** *Under the same condition as in Theorem 5.1, for $d = 1$, there exist positive absolute constants $c$ and $C$ such that, if*

$$\Delta_1 := \frac{(\theta_1\theta_2)^{1/2}b_n'^2 + \theta_1^2 b_n^3 \log n}{\sqrt{n}} \leqslant c,$$

*then, for $0 \leqslant x \leqslant \Delta_1^{-1/3}$,*

$$\left|\frac{P(W > x)}{P(Z > x)} - 1\right| \leqslant C(1 + x)(1 + |\log \Delta_1| + x^2)\Delta_1.$$

*For $d \geqslant 2$, let*

$$\Delta_d := \frac{d(\theta_1\theta_2)^{1/2}b_n'^2 + \theta_1^2 b_n^3 \log n}{\sqrt{n}}.$$

*Then, there exist positive absolute constants $c$ and $C$ such that, for $d^2(\log d)\Delta_d \leqslant c$ and $0 \leqslant x \leqslant \Delta_d^{-1/3}$, we have*

$$\left|\frac{P(|W| > x)}{P(|Z| > x)} - 1\right| \leqslant C(1 + x)(|\log \Delta_d| + d\log d + x^2)\Delta_d.$$

*Proof of Theorem* 5.2. Note that $d = \mathbb{E}(W^T W) \leqslant \theta_1 b_n^2$ and $1 \leqslant \theta_1 b_n'^2$. First consider the case $d = 1$. Let $p_0 = \Delta_1^{-2/3}$. If $\Delta_1$ is sufficently small, then $|\log \Delta_1| \leqslant p_0/2$ and moreover, using $d \leqslant \theta_1 b_n^2$ and $1 \leqslant \theta_1 b_n'^2$,

$$p_0 = \Delta_1^{-2/3} \leqslant \min\{(\frac{\theta_1 n}{\theta_2})^{1/3}, (\frac{n}{\theta_1^2 b_n^2})^{1/3}\},$$

which is bounded by the right-hand side of (5.1). Theorem 5.2 then follows from Theorem 2.2 with $r_0 = \alpha_1 = 1$ and Theorem 5.1. The case $d \geqslant 2$ follows by using Theorem 4.2 with $\alpha = 1$ instead of Theorem 2.2. □

## 5.2 Unbounded case

Next, we consider the unbounded case. We will do truncation and use Bernstein's inequality to control the truncation error. For this purpose, we need to assume that the index set $\{1, \dots, n\}$ can be partitioned into $L$ groups $g_1, \dots, g_L$ such that for each group $g_l$, the summands $\{X_i : i \in g_l\}$ are independent. We give two examples below. The next theorem, whose proof is deferred to Section 7.6, provides a moderate deviation result under this setting.

**Theorem 5.3.** *Under the setting of Theorem* 5.2*, replace the boundedness conditions* $|X_i| \leqslant b_n$ *and* $|X_{ij}| \leqslant b_n'$ *by* $\|X_{ij}\|_{\psi_1} \leqslant b$*. Assume in addition the above partition condition with* $L$ *groups. Let*

$$\Delta_d := \frac{dLb\log n + d(\theta_1\theta_2)^{1/2} b^2 \log^2 n + d^{3/2}\theta_1^2 b^3 \log^4 n}{\sqrt{n}}.$$

*For* $d = 1$*, there exist positive absolute constants* $c$ *and* $C$ *such that, if* $\Delta_1 \leqslant c$ *and* $0 \leqslant x \leqslant \Delta_1^{-1/3}$*, then*

$$\left| \frac{P(W > x)}{P(Z > x)} - 1 \right| \leqslant C(1 + x)(1 + |\log \Delta_1| + x^2)\Delta_1. \tag{5.3}$$

*For* $d \geqslant 2$*, there exist positive absolute constants* $c$ *and* $C$ *such that, if* $d^2(\log d)\Delta_d \leqslant c$ *and* $0 \leqslant x \leqslant \Delta_d^{-1/3}$*, then*

$$\left| \frac{P(|W| > x)}{P(|Z| > x)} - 1 \right| \leqslant C(1 + x)(|\log \Delta_d| + d\log d + x^2)\Delta_d. \tag{5.4}$$

**Example 5.1.** In $m$-dependence (cf. Hoeffding and Robbins (1948)), it is assumed that $X_i$ is independent of $\{X_j : |i - j| > m\}$. We obtain the following corollary of Theorem 5.3 for the case $d = 1$.

**Corollary 5.1.** Let $\{X_1, \dots, X_n\}$ be a sequence of $m$-dependent random variables with $m \geqslant 1$, $\mathbb{E}(X_i) = 0$ and $\|X_i\|_{\psi_1} \leqslant b$. Let $W = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. Suppose $\mathrm{Var}(W) = 1$. Let

$$\Delta = \frac{m^2 b^3 \log^4 n}{\sqrt{n}}.$$

Then there exist positive absolute constants $c$ and $C$ such that, for

$$\Delta \leqslant c, \quad 0 \leqslant x \leqslant \Delta^{-1/3},$$

we have

$$\left| \frac{P(W > x)}{P(Z > x)} - 1 \right| \leqslant C(1 + x)(1 + |\log \Delta| + x^2)\Delta.$$

*Proof of Corollary* 5.1. Under $m$-dependence, $\{X_1, \ldots, X_n\}$ can be partitioned into $L = m+1$ groups such that the $X$'s in each group are independent. Moreover, the quantities appearing in the statement of Theorem 5.3 can be taken as

$$\theta_1 \asymp m, \; \theta_2 \asymp m^2.$$

Using $1 \leqslant Cmb^2$, we have, $\Delta_1 \leqslant C\Delta$. The corollary then follows from (5.3). $\square$

**Example 5.2.** In graph dependency structure (cf. Baldi and Rinott (1989)), each index $i \in \{1, \ldots, n\}$ is represented by a node in a simple graph and $\{X_i : i \in A\}$ is assumed to be independent of $\{X_j : j \in B\}$ if $A$ and $B$ are disconnected. In such graph dependency structure, if the maximum degree of the dependency graph is $\deg^*$, then $L$ can be taken as $L = \deg^* + 1$. This is because each time we take out a group of independent summands, we can do it in a way that the max degree is decreased by 1. Therefore, Theorem 5.3 also applies. We omit the straightforward result.

**Remark 5.2.** Liu and Zhang (2021) obtained a moderate deviation result under local dependence in dimension one using a different method. Their result is stated under a more general condition and does not have the additional logarithmic terms. However, the dependence on the neighborhood size and $b$ in their result is worse than ours. For example, under $m$-dependence, the bound using their Theorem 2.1 with $\kappa \asymp m$ and $a_n \asymp \sqrt{n}/(mb)$ is

$$\left| \frac{P(W > x)}{P(Z > x)} - 1 \right| \leqslant C(1 + x^3) \frac{m^9 b^7}{\sqrt{n}},$$

while our bound is (cf. Corollary 5.1), subject to logarithmic terms,

$$\left| \frac{P(W > x)}{P(Z > x)} - 1 \right| \lesssim_{\log} C(1 + x^3) \frac{m^2 b^3}{\sqrt{n}}.$$

Moreover, our approach generalizes easily to multi-dimensions.

## 6 Proof of the $p$-Wasserstein bound

In this section, we prove Theorem 2.1. Without loss of generality, we may assume $Z$ is independent of $\mathcal{G}$ and $W'$.

We introduce some notations. Let $k \in \mathbb{N}$. Given families of real numbers $a = (a_{i_1,\ldots,i_k})_{1 \leqslant i_1,\ldots,i_k \leqslant d}$ and $b = (b_{i_1,\ldots,i_k})_{1 \leqslant i_1,\ldots,i_k \leqslant d}$, we set

$$\langle a, b \rangle := \sum_{i_1,\ldots,i_k=1}^{d} a_{i_1,\ldots,i_k} b_{i_1,\ldots,i_k}, \qquad |a| := \sqrt{\langle a, a \rangle} = \sqrt{\sum_{i_1,\ldots,i_k=1}^{d} a_{i_1,\ldots,i_k}^2}.$$

Note that, if $k = 2$, $\langle a, b \rangle = \langle a, b \rangle_{H.S.}$ and $|a| = \|a\|_{H.S.}$. For $x_1, \ldots, x_k \in \mathbb{R}^d$, we define

$$x_1 \otimes \cdots \otimes x_k := (x_{1,i_1} \cdots x_{k,i_k})_{1 \leqslant i_1,\ldots,i_k \leqslant d}.$$

If $x_1 = \cdots = x_d =: x$, we write $x_1 \otimes \cdots \otimes x_k = x^{\otimes k}$ for short. Also, if a function $f : \mathbb{R}^d \to \mathbb{R}$ is $k$-times differentiable at $w \in \mathbb{R}^d$, we set

$$\nabla^k f(w) := \left( \frac{\partial^k f}{\partial w_{i_1} \cdots \partial w_{i_k}}(w) \right)_{1 \leqslant i_1,\ldots,i_k \leqslant d}. \tag{6.1}$$

Given a family of random variables $X = (X_{i_1,\ldots,i_k})_{1 \leqslant i_1,\ldots,i_k \leqslant d}$ and $p > 0$, we set

$$\|X\|_p := \left( \mathbb{E}|X|^p \right)^{1/p}.$$

We denote by $\phi$ the $d$-dimensional standard normal density. For brevity, we write $\eta_t$ instead of $\eta_t(p)$ throughout this section.

### 6.1 Auxiliary estimates

For every $t > 0$, we set $F_t := e^{-t}W + \sqrt{1-e^{-2t}}Z$. It is straightforward to check that $F_t$ has a smooth density $f_t$ with respect to $N(0, I_d)$. Moreover, $f_t$ is strictly positive by Lemma 3.1 of Johnson and Suhov (2001). Therefore, we can define the *score* of $F_t$ with respect to $N(0, I_d)$ by $\rho_t(w) = \nabla \log f_t(w)$, $w \in \mathbb{R}^d$. We use $C$ to denote positive absolute constants, which may differ in different expressions.

**Proposition 6.1.** Let $p \geqslant 1$ and $t > 0$. Under the assumptions of Theorem 2.1, we have

$$\|\rho_t(F_t)\|_p \leqslant Ce^{-t}\left(\|R_t\|_p + \frac{\|E\|_p}{\eta_t} + \min\left\{\frac{\sqrt{d}}{\eta_t}, \frac{\|\mathbb{E}[D^{\otimes 2}|D|^2 1_{\{|D|\leqslant \eta_t\}}|\mathcal{G}]\|_p}{\lambda\eta_t^3}\right\}\right).$$

We need some lemmas to prove Proposition 6.1.

**Lemma 6.1** (Lemma A.3 of Fang and Koike (2022))**.** Let $Y = (Y_{ij})_{1\leqslant i,j\leqslant d}$ be a $d \times d$ positive semidefinite symmetric random matrix. Let $F$ and $G$ be two random variables such that $|F| \leqslant G$. Suppose that $\mathbb{E}|Y_{ij}G| < \infty$ for all $i, j = 1, \dots, d$. Let $\mathcal{G}$ be an arbitrary $\sigma$-field. Then we have

$$\|\mathbb{E}[YF|\mathcal{G}]\|_{H.S.} \leqslant \|\mathbb{E}[YG|\mathcal{G}]\|_{H.S.}.$$

**Lemma 6.2** (Lemma A.4 of Fang and Koike (2022))**.** Let $Y$ be a random vector in $\mathbb{R}^d$ such that $\mathbb{E}|Y|^k < \infty$ for some integer $k \geqslant 2$. Let $\mathcal{G}$ be an arbitrary $\sigma$-field. Then

$$|\mathbb{E}[Y^{\otimes k}|\mathcal{G}]| \leqslant \|\mathbb{E}[Y^{\otimes 2}|Y|^{k-2}|\mathcal{G}]\|_{H.S.}.$$

**Lemma 6.3.** Let $F$ be a random vector in $\mathbb{R}^m$ whose components are of the form $Q(Z_1, \dots, Z_d)$, where $Q$ is a polynomial of degree $\leqslant k$. Then, for every $p > 0$,

$$\|F\|_p \leqslant \kappa_p^k \|F\|_2,$$

where $\kappa_p := e\sqrt{(p/2-1)\vee 1}$.

*Proof.* Since $|F|^2$ is a polynomial of degree $\leqslant 2k$ in $Z_1, \dots, Z_d$ by assumption, we have by Theorem 5.10 and Remark 5.11 of Janson (1997)

$$\|F\|_p = \||F|^2\|_{p/2}^{1/2} \leqslant (p/2-1)^{k/2}\||F|^2\|_2^{1/2}$$

if $p \geqslant 4$. Since we have $\||F|^2\|_{p/2} \leqslant \||F|^2\|_2$ if $p < 4$, we obtain

$$\|F\|_p \leqslant \{(p/2-1)\vee 1\}^{k/2}\||F|^2\|_2^{1/2}. \tag{6.2}$$

Next, we have by Theorem 5.10 and Remark 5.13 of Janson (1997)

$$\||F|^2\|_2 \leqslant e^{2k}\||F|^2\|_1 = e^{2k}\|F\|_2^2. \tag{6.3}$$

The desired result follows from (6.2)–(6.3). □

Given a bounded measurable function $h : \mathbb{R}^d \to \mathbb{R}$ and $t > 0$, we define the function $T_t h : \mathbb{R}^d \to \mathbb{R}$ by

$$T_t h(w) = \mathbb{E}h(e^{-t}w + \sqrt{1-e^{-2t}}Z), \qquad w \in \mathbb{R}^d.$$

One can easily check that $T_t h$ is infinitely differentiable and

$$\nabla^k T_t h(w) = \frac{(-1)^k}{(e^{2t}-1)^{k/2}}\int_{\mathbb{R}^d} h(e^{-t}w + \sqrt{1-e^{-2t}}z)\nabla^k\phi(z)dz, \qquad k = 1, 2, \dots. \tag{6.4}$$

**Lemma 6.4.** Let $X$ and $X'$ be two $d$-dimensional random vectors such that $|X'-X|$ is bounded, and set $Y := X'-X$. Then, for any integer $l \geqslant 0$, bounded measurable function $h : \mathbb{R}^d \to \mathbb{R}$ and $t > 0$, we have

$$\langle \nabla^l T_t h(X') - \nabla^l T_t h(X), Y^{\otimes l}\rangle = \sum_{k=1}^{\infty} \frac{1}{k!}\langle \nabla^{l+k} T_t h(X), Y^{\otimes (l+k)}\rangle \qquad \text{in } L^\infty(P).$$

*Proof.* By assumption, there is a constant $M > 0$ such that $|Y| \leqslant M$ and $\sup_{x\in\mathbb{R}^d}|h(x)| \leqslant M$. Using (6.4), we deduce

$$\begin{aligned}|\langle \nabla^k T_t h(w), Y^{\otimes k}\rangle| &\leqslant \frac{M}{(e^{2t}-1)^{k/2}} \int_{\mathbb{R}^d} |\langle \nabla^k \phi(z), Y^{\otimes k}\rangle| dz \\ &\leqslant \frac{M}{(e^{2t}-1)^{k/2}} \sqrt{\int_{\mathbb{R}^d} \left(\frac{\langle \nabla^k \phi(z), Y^{\otimes k}\rangle}{\phi(z)}\right)^2 \phi(z) dz}.\end{aligned}$$

Thus, we have by Lemma 4.3 of Fang and Röllin (2015)

$$|\langle \nabla^k T_t h(w), Y^{\otimes k}\rangle| \leqslant \frac{M}{(e^{2t}-1)^{k/2}} \sqrt{k!|Y^{\otimes k}|^2} = \frac{M\sqrt{k!}|Y|^k}{(e^{2t}-1)^{k/2}} \leqslant \frac{M^{k+1}\sqrt{k!}}{(e^{2t}-1)^{k/2}}.$$

Hence, for any integer $K > 0$, we have by Taylor's expansion

$$\begin{aligned}&\left| \langle \nabla^l T_t h(X') - \nabla^l T_t h(X), Y^{\otimes l}\rangle - \sum_{k=1}^{K} \frac{1}{k!}\langle \nabla^{l+k} T_t h(X), Y^{\otimes (l+k)}\rangle \right| \\ &\leqslant \sup_{u\in[0,1]} \frac{1}{(K+1)!} \left|\langle \nabla^{l+K+1} T_t h(X+uY), Y^{\otimes (l+K+1)}\rangle\right| \leqslant \frac{M^{l+K+1}\sqrt{(l+K+1)!}}{(K+1)!(e^{2t}-1)^{(l+K+1)/2}}.\end{aligned}$$

Since the last quantity tends to 0 as $K \to \infty$, we complete the proof. ☐

For every $t > 0$, let

$$D_t := D1_{\{|D|\leqslant \eta_t\}}, \qquad W_t := W + D_t.$$

Note that we have

$$W_t = \begin{cases} W' & \text{if } |D| \leqslant \eta_t, \\ W & \text{if } |D| > \eta_t. \end{cases}$$

One can check that $(W, W_t)$ is an exchangeable pair. In fact, for any $u, v \in \mathbb{R}^d$, we have

$$\begin{aligned}\mathbb{E}[e^{\sqrt{-1}(u\cdot W + v\cdot W_t)}] &= \mathbb{E}[e^{\sqrt{-1}(u\cdot W + v\cdot W')}1_{\{|D|\leqslant\eta_t\}}] + \mathbb{E}[e^{\sqrt{-1}(u\cdot W + v\cdot W)}1_{\{|D|>\eta_t\}}] \\ &= \mathbb{E}[e^{\sqrt{-1}(u\cdot W' + v\cdot W)}1_{\{|D|\leqslant\eta_t\}}] + \mathbb{E}[e^{\sqrt{-1}(u\cdot W + v\cdot W)}1_{\{|D|>\eta_t\}}] \\ &= \mathbb{E}[e^{\sqrt{-1}(u\cdot W_t + v\cdot W)}1_{\{|D|\leqslant\eta_t\}}] + \mathbb{E}[e^{\sqrt{-1}(u\cdot W_t + v\cdot W)}1_{\{|D|>\eta_t\}}] \\ &= \mathbb{E}[e^{\sqrt{-1}(u\cdot W_t + v\cdot W)}],\end{aligned}$$

where the second equality follows from the exchangeability of $(W, W')$. Also, using (2.1) and recalling (2.4), one can easily check

$$\mathbb{E}[W_t - W|\mathcal{G}] = -\Lambda(W + R_t). \tag{6.5}$$

Let us set

$$\tau_t := \mathbb{E}\left[\Lambda^{-1}D_t\left(1 - \frac{1}{2}\frac{\langle\nabla\phi(Z), D_t\rangle}{\phi(Z)\sqrt{e^{2t}-1}} + \frac{1}{2}\sum_{k=3}^{\infty} a_k \frac{(-1)^k\langle\nabla^k\phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)(e^{2t}-1)^{k/2}}\right)|\mathcal{G}\vee\sigma(Z)\right], \tag{6.6}$$

where $a_k := \frac{1}{k!} - \frac{1}{4(k-2)!}$. As in the proof of Lemma 6.4, one can check that the series inside the conditional expectation in (6.6) converges in $L^1(P)$, so $\tau_t$ is well-defined.

**Lemma 6.5.** $\mathbb{E}[\tau_t|F_t] = 0$ for all $t > 0$.

*Proof.* It suffices to prove $\mathbb{E}[\tau_t h(F_t)] = 0$ for any bounded measurable function $h : \mathbb{R}^d \to \mathbb{R}$. We have by exchangeability

$$\mathbb{E}[\Lambda^{-1} D_t \{T_t h(W) + T_t h(W_t)\}] = 0.$$

Applying Lemma 6.4, we obtain

$$\mathbb{E}\left[\Lambda^{-1} D_t \left\{T_t h(W) + \sum_{k=0}^{\infty} \frac{1}{k!} \langle \nabla^k T_t h(W), D_t^{\otimes k}\rangle\right\}\right] = 0. \tag{6.7}$$

Now, we have again by exchangeability

$$\mathbb{E}\left[\Lambda^{-1} D_t \langle \nabla^2 T_t h(W), D_t^{\otimes 2}\rangle\right] = -\mathbb{E}\left[\Lambda^{-1} D_t \langle \nabla^2 T_t h(W_t), D_t^{\otimes 2}\rangle\right]. \tag{6.8}$$

Hence we obtain

$$\begin{aligned}
\mathbb{E}\left[\Lambda^{-1} D_t \langle \nabla^2 T_t h(W), D_t^{\otimes 2}\rangle\right] &= -\frac{1}{2}\mathbb{E}\left[\Lambda^{-1} D_t \langle \nabla^2 T_t h(W_t) - \nabla^2 T_t h(W), D_t^{\otimes 2}\rangle\right] \\
&= -\frac{1}{2}\mathbb{E}\left[\Lambda^{-1} D_t \sum_{k=1}^{\infty} \frac{1}{k!} \langle \nabla^{k+2} T_t h(W), D_t^{\otimes (2+k)}\rangle\right] \\
&= -\frac{1}{2}\mathbb{E}\left[\Lambda^{-1} D_t \sum_{k=3}^{\infty} \frac{1}{(k-2)!} \langle \nabla^{k} T_t h(W), D_t^{\otimes k}\rangle\right].
\end{aligned}$$

Inserting this into (6.7), we deduce

$$\mathbb{E}\left[\Lambda^{-1} D_t \left\{2T_t h(W) + D_t \cdot \nabla T_t h(W) + \sum_{k=3}^{\infty} a_k \langle \nabla^k T_t h(W), D_t^{\otimes k}\rangle\right\}\right] = 0. \tag{6.9}$$

Meanwhile, we have by (6.4)

$$\nabla^k T_t h(w) = \frac{(-1)^k}{(e^{2t}-1)^{k/2}} \mathbb{E} h(e^{-t} w + \sqrt{1-e^{-2t}} Z) \frac{\nabla^k \phi(Z)}{\phi(Z)}.$$

Inserting this into (6.9) and using the definition of $F_t$, we obtain $2\mathbb{E}[\tau_t h(F_t)] = 0$. Hence we complete the proof. □

*Proof of Proposition* 6.1. Recall

$$a_k := \frac{1}{k!} - \frac{1}{4(k-2)!}, \ \kappa_p := e\sqrt{(p/2-1) \vee 1}.$$

We divide the proof into two steps.
**Step 1**. We first prove the following inequality:

$$\|\rho_t(F_t)\|_p \leqslant e^{-t}\left(\|R_t\|_p + \frac{\kappa_p}{\sqrt{e^{2t}-1}}\|E_t\|_p + \frac{1}{2}\sum_{k=3}^{\infty} \frac{|a_k|\kappa_p^k \sqrt{k!}}{(e^{2t}-1)^{k/2}} \left\|\mathbb{E}[(\Lambda^{-1} D_t) \otimes D_t^{\otimes k} | \mathcal{G}]\right\|_p\right), \tag{6.10}$$

where

$$E_t := E - \frac{1}{2}\mathbb{E}[(\Lambda^{-1} D) \otimes D 1_{\{|D| > \eta_t\}} | \mathcal{G}].$$

We have by Lemma 2 of Bonis (2020)

$$\rho_t(F_t) = \mathbb{E}\left[e^{-t} W - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z | F_t\right] = e^{-t}\mathbb{E}\left[W - \frac{1}{\sqrt{e^{2t}-1}} Z | F_t\right]. \tag{6.11}$$

Hence, Lemma 6.5 yields

$$\begin{aligned}\rho_t(F_t) &= e^{-t}\mathbb{E}\left[W - \frac{1}{\sqrt{e^{2t}-1}}Z + \tau_t | F_t\right] \\ &= e^{-t}\mathbb{E}\left[-R_t + \frac{1}{\sqrt{e^{2t}-1}}E_t Z + \frac{1}{2}\sum_{k=3}^{\infty} a_k \mathbb{E}\left[\Lambda^{-1} D_t \frac{(-1)^k \langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)(e^{2t}-1)^{k/2}} | \mathcal{G} \vee \sigma(Z)\right] | F_t\right].\end{aligned}$$

Therefore, we have by the Jensen and Minkowski inequalities

$$\begin{aligned}\|\rho_t(F_t)\|_p \leqslant e^{-t}\Bigg(&\|R_t\|_p + \frac{1}{\sqrt{e^{2t}-1}}\|E_t Z\|_p \\ &+ \frac{1}{2}\sum_{k=3}^{\infty} \frac{|a_k|}{(e^{2t}-1)^{k/2}} \left\| \mathbb{E}\left[\Lambda^{-1} D_t \frac{\langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)} | \mathcal{G} \vee \sigma(Z)\right]\right\|_p\Bigg). \qquad (6.12)\end{aligned}$$

Now, Lemma 6.3 yields

$$\mathbb{E}[|E_t Z|^p | \mathcal{G}] \leqslant \left(\kappa_p^2 \mathbb{E}[|E_t Z|^2 | \mathcal{G}]\right)^{p/2}$$

and

$$\begin{aligned}&\mathbb{E}\left[\left|\mathbb{E}\left[\Lambda^{-1} D_t \frac{\langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)} | \mathcal{G} \vee \sigma(Z)\right]\right|^p | \mathcal{G}\right] \\ &\leqslant \left(\kappa_p^{2k} \mathbb{E}\left[\left|\mathbb{E}\left[\Lambda^{-1} D_t \frac{\langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)} | \mathcal{G} \vee \sigma(Z)\right]\right|^2 | \mathcal{G}\right]\right)^{p/2}.\end{aligned}$$

Note that, conditional on $\mathcal{G}$, $E_t Z \sim N(0, E_t E_t^T)$. Thus we have

$$\mathbb{E}[|E_t Z|^2 | \mathcal{G}] = |E_t|^2.$$

Meanwhile, we have by Lemma 4.3 of Fang and Röllin (2015)

$$\begin{aligned}&\mathbb{E}\left[\left|\mathbb{E}\left[\Lambda^{-1} D_t \frac{\langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)} | \mathcal{G} \vee \sigma(Z)\right]\right|^2 | \mathcal{G}\right] \\ &= \sum_{j=1}^{d} \mathbb{E}\left[\left|\frac{\langle \nabla^k \phi(Z), \mathbb{E}[(\Lambda^{-1} D_t)_j D_t^{\otimes k} | \mathcal{G}]\rangle}{\phi(Z)}\right|^2 | \mathcal{G}\right] \\ &\leqslant k! \sum_{j=1}^{d} \left|\mathbb{E}[(\Lambda^{-1} D_t)_j D_t^{\otimes k} | \mathcal{G}]\right|^2 = k! \left|\mathbb{E}[(\Lambda^{-1} D_t) \otimes D_t^{\otimes k} | \mathcal{G}]\right|^2.\end{aligned}$$

Consequently, we obtain

$$\|E_t Z\|_p \leqslant \kappa_p \|E_t\|_p$$

and

$$\left\| \mathbb{E}\left[\Lambda^{-1} D_t \frac{\langle \nabla^k \phi(Z), D_t^{\otimes k}\rangle}{\phi(Z)} | \mathcal{G} \vee \sigma(Z)\right]\right\|_p \leqslant \kappa_p^k \sqrt{k!} \left\|\mathbb{E}[(\Lambda^{-1} D_t) \otimes D_t^{\otimes k} | \mathcal{G}]\right\|_p.$$

Inserting these estimates into (6.12), we obtain (6.10).

**Step 2**. We have by Lemma 6.1

$$\begin{aligned}|E_t| &\leqslant |E| + (2\lambda)^{-1}\min\{|\mathbb{E}[D^{\otimes 2}|\mathcal{G}]|, \eta_t^{-2}|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]|\}\\ &\leqslant |E| + (2\lambda)^{-1}\min\{2\lambda(|E|+\sqrt{d}), \eta_t^{-2}|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]|\}\\ &\leqslant 2|E| + (2\lambda)^{-1}\min\{2\lambda\sqrt{d}, \eta_t^{-2}|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]|\}.\end{aligned}$$

We also have by Lemmas 6.1 and 6.2

$$\begin{aligned}|\mathbb{E}[D_t^{\otimes(k+1)}|\mathcal{G}]| \leqslant |\mathbb{E}[D_t^{\otimes 2}|D_t|^{k-1}|\mathcal{G}]| &= |\mathbb{E}[D^{\otimes 2}|D|^{k-1}1_{\{|D|\leqslant\eta_t\}}|\mathcal{G}]|\\ &\leqslant \min\{\eta_t^{k-1}|\mathbb{E}[D^{\otimes 2}|\mathcal{G}]|, \eta_t^{k-3}|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]|\}\\ &\leqslant 2\lambda\eta_t^{k-1}|E| + \min\{2\lambda\eta_t^{k-1}\sqrt{d}, \eta_t^{k-3}|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]|\}.\end{aligned}$$

Inserting these estimates into (6.10) and noting $\kappa_p \leqslant e\sqrt{p}$ as well as $\sum_{k=3}^{\infty}|a_k|e^k\sqrt{k!} < \infty$, we obtain the desired result. □

### 6.2 Proof of Theorem 2.1

By Eq.(3.8) of Ledoux, Nourdin and Peccati (2015),

$$\mathcal{W}_p(W,Z) \leqslant \int_0^\infty \|\rho_t(F_t)\|_p dt, \quad p \geqslant 1. \tag{6.13}$$

Strictly speaking, this bound was only proved when $W$ has a bounded $C^\infty$ density $h$ with respect to $N(0,I_d)$ such that $h \geqslant \eta$ for some constant $\eta > 0$ and $|\nabla h|$ is bounded (cf. Eq.(32) of Otto and Villani (2000)). However, this restriction can be removed by a similar argument as in Section 8 of Bonis (2020). For completeness, we give a formal proof in Appendix A.2 of the appendix.

(2.2) follows by combining (6.13) with Proposition 6.1.

Next, take $\varepsilon > 0$ arbitrarily. We have

$$\begin{aligned}&\int_0^\infty e^{-t}\min\left\{\frac{\sqrt{d}}{\eta_t}, \frac{\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{\lambda\eta_t^3}\right\}dt\\ &\leqslant \sqrt{pd}\int_0^\varepsilon \frac{e^{-t}}{\sqrt{e^{2t}-1}}dt + \frac{p^{3/2}\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{\lambda}\int_\varepsilon^\infty \frac{e^{-t}}{(e^{2t}-1)^{3/2}}dt.\end{aligned}$$

Since

$$\int_0^\varepsilon \frac{e^{-t}}{\sqrt{e^{2t}-1}}dt \leqslant \int_0^\varepsilon \frac{1}{\sqrt{2t}}dt = \sqrt{2\varepsilon}$$

and

$$\int_\varepsilon^\infty \frac{e^{-t}}{(e^{2t}-1)^{3/2}}dt \leqslant \int_\varepsilon^\infty \frac{1}{(2t)^{3/2}}dt = \frac{1}{\sqrt{2\varepsilon}},$$

taking

$$\varepsilon = \frac{p\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{2\sqrt{d}\lambda},$$

we obtain

$$\int_0^\infty e^{-t}\min\left\{\frac{\sqrt{d}}{\eta_t}, \frac{\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{\lambda\eta_t^3}\right\}dt \leqslant Cpd^{1/4}\sqrt{\frac{\|\mathbb{E}[D^{\otimes 2}|D|^2|\mathcal{G}]\|_p}{\lambda}}.$$

Also, observe that

$$\int_0^\infty \frac{e^{-t}}{\sqrt{e^{2t}-1}}dt = \frac{1}{2}\int_0^1 \frac{1}{\sqrt{1-x}}dx = 1.$$

Inserting these estimates into (2.2), we obtain (2.3).

# 7 More proofs

## 7.1 Generalized exchangeable pairs

Here we record a $p$-Wasserstein bound for generalized exchangeable pairs. Let $\mathcal{X}$ be a general space and suppose $(X, X')$ is an exchangeable pair of $\mathcal{X}$-valued random variables. Let $W := W(X) \in \mathbb{R}^d$ be the random vector of interest, $W' := W(X')$ and $D := W' - W$. Suppose there exists an antisymmetric function $G := G(X, X') \in \mathbb{R}^d$ (i.e., $G(X, X') = -G(X', X)$ a.s.) such that

$$\mathbb{E}(G|\sigma(X)) = -(W + R). \tag{7.1}$$

Suppose the law of $W$ is approximately $N(0, I_d)$ and we are interested in bounding

$$\mathcal{W}_p(W, Z).$$

The formulation (7.1) with $d = 1$ was first proposed by Chatterjee (2007) for concentration inequalities (see also Zhang (2022) for Kolmogorov bounds). In Corollary 2.11 of Döbler (2023) for 1-Wasserstein bounds, he considered the case $d = 1$, $W = \sum_{l=1}^m W_l$ and $\mathbb{E}[W_l' - W_l|X] = -\lambda_l W_l$. In this case, we can choose $G$ in (7.1) to be $G = \sum_{l=1}^m \frac{W_l' - W_l}{\lambda_l}$. For $d > 1$, the setting of Reinert and Röllin (2009) corresponds to $G = \Lambda^{-1}(W' - W)$.

**Theorem 7.1.** *Under the above setting, assume that $\mathbb{E}|W|^p < \infty$ for some $p \geqslant 1$ and $\mathbb{E}|G||D|^3 < \infty$. Then we have*

$$\mathcal{W}_p(W, Z) \leqslant C\left(\int_0^\infty e^{-t}\|R_t\|_p dt + \sqrt{p}\|E\|_p + p\sqrt{\|\mathbb{E}[|G||D||\sigma(X)]\|_p \|\mathbb{E}[|G||D|^3|\sigma(X)]\|_p}\right),$$

*where $Z \sim N(0, I_d)$ is a $d$-dimensional standard Gaussian vector,*

$$R_t := R + \mathbb{E}[G1_{\{|D| > \sqrt{(e^{2t}-1)/p}\}}|\sigma(X)], \qquad E := \frac{1}{2}\mathbb{E}[G \otimes D|\sigma(X)] - I_d,$$

*and $C$ is an absolute constant.*

*Proof of Theorem* 7.1. The proof is a straightforward modification of that of Theorem 2.1. We use the notation therein. Let

$$G_t := G1_{\{|D| \leqslant \eta_t\}}.$$

We start from the identity

$$\mathbb{E}[G_t\{T_t h(W) + T_t h(W_t)\}] = 0.$$

Following the proof of Proposition 6.1 except that we change $\Lambda^{-1}D_t$ therein by $G_t$ and use $|\mathbb{E}[Y_1 \otimes \cdots \otimes Y_k|\sigma(X)]| \leqslant \mathbb{E}[|Y_1 \otimes \cdots \otimes Y_k||\sigma(X)] = \mathbb{E}[|Y_1|\cdots|Y_k||\sigma(X)]$ instead of Lemmas 6.1 and 6.2, we obtain

$$\begin{aligned}&\|\rho_t(F_t)\|_p\\ \leqslant& Ce^{-t}\left(\|R_t\|_p + \frac{\sqrt{p}}{\sqrt{e^{2t}-1}}\|E\|_p + \min\left\{\frac{\sqrt{p}\|\mathbb{E}[|G||D||\sigma(X)]\|_p}{\sqrt{e^{2t}-1}}, \frac{p^{3/2}\|\mathbb{E}[|G||D|^3|\sigma(X)]\|_p}{(e^{2t}-1)^{3/2}}\right\}\right).\end{aligned}$$

Then, the theorem follows by optimizing the integration as in the proof of Theorem 2.1. □

### 7.2 Proof for combinatorial CLT

*Proof of Proposition* 3.1. In this proof, we use $C$ to denote positive absolute constants, which may differ in different expressions.

**Step 1. The exchangeable pair.** Let $Y_{ij} = X_{ij}/B_n$ and hence, $W = \sum_{i=1}^n Y_{i\pi(i)}$. We construct an exchangeable pair $(W, W')$ by uniformly selecting two different indices $I, J \in \{1, \ldots, n\}$, independent of $\mathbb{X}$ and $\pi$, and let

$$W' = W + D = W - Y_{I\pi(I)} - Y_{J\pi(J)} + Y_{I\pi(J)} + Y_{J\pi(I)}.$$

Let $\mathcal{G} = \sigma(\mathbb{X}, \pi)$. It is known that (cf. Eq. (3.3) of Chen and Fang (2015))

$$\mathbb{E}(W' - W|\mathcal{G}) = -\lambda(W + R), \tag{7.2}$$

where

$$\lambda = \frac{2}{n-1}, \quad R = -\frac{1}{n}\sum_{i,j=1}^n Y_{ij}.$$

For $1 \leqslant i \neq j \leqslant n$, let

$$Y_\pi^{(ij)} := -Y_{i\pi(i)} - Y_{j\pi(j)} + Y_{i\pi(j)} + Y_{j\pi(i)}.$$

For $t > 0$ and $p \geqslant 2$, let $\eta_t(p) = \sqrt{(e^{2t}-1)/p}$ be as in Theorem 2.1. For any given permutation $\pi$, because of the assumption $\|X_{ij}\|_{\psi_1} \leqslant b$, we have, following the same argument as in Section 2.3 for the independent case,

$$\|Y_\pi^{(ij)} 1_{\{|Y_\pi^{(ij)}| > \eta_t(p)\}}\|_{\psi_{1/2}} \leqslant C\eta_t^{-1}(p)\frac{b^2}{B_n^2}, \tag{7.3}$$

$$\|(Y_\pi^{(ij)})^2\|_{\psi_{1/2}} \leqslant \frac{Cb^2}{B_n^2}, \tag{7.4}$$

$$\|(Y_\pi^{(ij)})^4 1_{\{|Y_\pi^{(ij)}| \leqslant \eta_t(p)\}}\|_{\psi_{1/2}} \leqslant C\eta_t^2(p)\frac{b^2}{B_n^2}.$$

We will apply the $p$-Wasserstein bound (2.2), which we recall:

$$\mathcal{W}_p(W, Z) \leqslant C\int_0^\infty e^{-t}\left(\|R_t\|_p + \frac{\|E\|_p}{\eta_t(p)} + \min\left\{\frac{1}{\eta_t(p)}, \frac{\|\mathbb{E}[D^4 1_{\{|D| \leqslant \eta_t(p)\}}|\mathcal{G}]\|_p}{\lambda\eta_t^3(p)}\right\}\right)dt,$$

where

$$R_t := R + \mathbb{E}[\lambda^{-1} D 1_{\{|D| > \eta_t(p)\}}|\mathcal{G}], \qquad E := \frac{1}{2}\mathbb{E}[\lambda^{-1}D^2|\mathcal{G}] - 1.$$

**Step 2. Bounding $R_t$.** For the above exchangeable pair, we have

$$R_t = -\frac{1}{n}\sum_{i,j=1}^n Y_{ij} + \frac{1}{n}\sum_{1 \leqslant i < j \leqslant n} Y_\pi^{(ij)} 1_{\{|Y_\pi^{(ij)}| > \eta_t(p)\}}.$$

Because of centering (i.e., $c_{i\cdot} = c_{\cdot j} = 0$), we have

$$\frac{1}{n}\sum_{i,j=1}^n Y_{ij} = \frac{1}{n}\sum_{i,j=1}^n (Y_{ij} - \mathbb{E}Y_{ij}).$$

From Lemma 2.1 and $\|Y_{ij}\|_{\psi_1} \leqslant b/B_n$, we have

$$\|\frac{1}{n}\sum_{i,j=1}^n Y_{ij}\|_p \leqslant \frac{Cb}{nB_n}(\sqrt{pn^2} + p) \leqslant C(\frac{p\sqrt{n}}{B_n^2} + \frac{p^{5/2}}{B_n^2})b^2,$$

where we used $B_n^2 \leqslant Cnb^2$ from (3.2) in the last inequality.

To deal with the second term in $R_t$, we separate $\sum_{1\leqslant i<j\leqslant n}$ into $O(n)$ sums, each sum is over a collection of $O(n)$ disjoint pairs $(i,j)$. For example, $\{1\leqslant i<j\leqslant n\} = \cup_{l=1}^{n-1}(\mathcal{I}_l^{(1)}\cup\mathcal{I}_l^{(2)})$, where

$$\mathcal{I}_l^{(1)} = \{1\leqslant i<j\leqslant n : j-i=l, i\in\{kl+1,\dots,(k+1)l\}, k\geqslant 0 \text{ an odd integer}\},$$

$$\mathcal{I}_l^{(2)} = \{1\leqslant i<j\leqslant n : j-i=l, i\in\{kl+1,\dots,(k+1)l\}, k\geqslant 0 \text{ an even integer}\}.$$

Consider such a sum

$$\sum_{(i,j)\in\mathcal{I}} Y_\pi^{(ij)} 1_{\{|Y_\pi^{(ij)}|>\eta_t(p)\}}.$$

Conditioning on the unordered pair $\{\pi(i),\pi(j)\}$ for all $(i,j)\in\mathcal{I}$, it is a sum of $O(n)$ independent random variables, each with mean 0 and $\|\cdot\|_{\psi_{1/2}} \leqslant C\eta_t^{-1}(p)b^2/B_n^2$ (cf. (7.3)). From Lemma 2.1, we obtain

$$\|\sum_{(i,j)\in\mathcal{I}} Y_\pi^{(ij)} 1_{\{|Y_\pi^{(ij)}|>\eta_t(p)\}}\|_p \leqslant C\eta_t^{-1}(p)\frac{b^2}{B_n^2}(\sqrt{pn}+p^2).$$

Combining the above bounds, we obtain

$$\int_0^\infty e^{-t}\|R_t\|_p dt \leqslant C(\frac{p\sqrt{n}}{B_n^2} + \frac{p^{5/2}}{B_n^2})b^2.$$

**Step 3. Bounding $E$.** Note that

$$\begin{aligned}
E :=& \frac{1}{2\lambda}\mathbb{E}[D^2|\mathcal{G}] - 1\\
=&\frac{1}{2\lambda}\mathbb{E}[D^2|\mathcal{G}] - \frac{1}{2\lambda}\mathbb{E}[D^2] + \frac{1}{2\lambda}\mathbb{E}[D^2] - 1\\
=&\frac{1}{2n}\sum_{1\leqslant i<j\leqslant n}\left[(Y_\pi^{(ij)})^2 - \mathbb{E}(Y_\pi^{(ij)})^2\right] + \frac{1}{2\lambda}\mathbb{E}[D^2] - 1\\
=:&H_{21} + H_{22}.
\end{aligned}$$

From exchangeability and the linearity condition (7.2), we obtain

$$\begin{aligned}
H_{22} =& \frac{1}{2\lambda}\mathbb{E}(W'-W)^2 - 1 = \frac{1}{2\lambda}(-2\mathbb{E}[(W'-W)W]) - 1\\
=&\mathbb{E}(RW) = -\frac{1}{n}\mathbb{E}\left[\sum_{i,j=1}^n Y_{ij}\sum_{k=1}^n Y_{k\pi(k)}\right] = -\frac{1}{n^2}\mathbb{E}\left[\sum_{i,j=1}^n Y_{ij}\sum_{k,l=1}^n Y_{kl}\right].
\end{aligned}$$

From (3.2), we have

$$|H_{22}| = \frac{1}{n^2}\mathbb{E}(\sum_{i,j=1}^n Y_{ij})^2 = \frac{1}{n^2}\operatorname{Var}(\sum_{i,j=1}^n Y_{ij}) \leqslant \frac{1}{n}.$$

Now we turn to bounding $H_{21}$. Write

$$H_{21} = \frac{1}{2n}\sum_{1\leqslant i<j\leqslant n}\left[(Y_\pi^{(ij)})^2 - \mathbb{E}^\pi(Y_\pi^{(ij)})^2\right] + \frac{1}{2n}\sum_{1\leqslant i<j\leqslant n}\left[\mathbb{E}^\pi(Y_\pi^{(ij)})^2 - \mathbb{E}(Y_\pi^{(ij)})^2\right],$$

where $\mathbb{E}^\pi$ denotes the conditional expectation given the permutation $\pi$. From a similar argument as in bounding $R_t$ and using (7.4) for the first term, we obtain

$$\|\frac{1}{2n}\sum_{1\leqslant i<j\leqslant n}\left[(Y_\pi^{(ij)})^2 - \mathbb{E}^\pi(Y_\pi^{(ij)})^2\right]\|_p \leqslant C(\frac{\sqrt{pn}}{B_n^2} + \frac{p^2}{B_n^2})b^2.$$

Now we turn to bounding the second term of $H_{21}$. Let

$$\xi_{ij} := \frac{\mathbb{E}^{\pi}(Y_{\pi}^{(ij)})^2 - \mathbb{E}(Y_{\pi}^{(ij)})^2}{n^{3/2}b^2/B_n^2},$$

and hence,

$$\frac{1}{2n}\sum_{1\leqslant i<j\leqslant n}\Big[\mathbb{E}^{\pi}(Y_{\pi}^{(ij)})^2 - \mathbb{E}(Y_{\pi}^{(ij)})^2\Big] = \frac{n^{1/2}b^2}{2B_n^2}\sum_{1\leqslant i<j\leqslant n}\xi_{ij}.$$

In the remainder of this step, we show that with $V = \sum_{1\leqslant i<j\leqslant n}\xi_{ij}$ and if $p \geqslant 2$, we have

$$\|V\|_p \leqslant C(\sqrt{p} + \frac{p}{\sqrt{n}}), \tag{7.5}$$

and hence

$$\int_0^\infty e^{-t}\frac{\|E\|_p}{\eta_t(p)}dt \leqslant C(\frac{p\sqrt{n}}{B_n^2} + \frac{p^{5/2}}{B_n^2})b^2,$$

where we used $B_n^2 \leqslant Cnb^2$ again to simplify the upper bound. To prove (7.5), let $h(t) = \mathbb{E}e^{tV}$. We have

$$\begin{aligned} h'(t) &= \sum_{1\leqslant i<j\leqslant n}\mathbb{E}\xi_{ij}e^{tV} = \sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}\mathbb{E}\{\mathbb{E}[\xi_{ij}e^{tV}|\pi(i)=k,\pi(j)=l]\} \\ &= \sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}a_{ijkl}\mathbb{E}\{\mathbb{E}[e^{tV}|\pi(i)=k,\pi(j)=l]\}, \end{aligned} \tag{7.6}$$

where

$$a_{ijkl} := \frac{\mathbb{E}(-Y_{ik}-Y_{jl}+Y_{il}+Y_{jk})^2 - \mathbb{E}(-Y_{i\pi(i)}-Y_{j\pi(j)}+Y_{i\pi(j)}+Y_{j\pi(i)})^2}{n^{3/2}b^2/B_n^2}.$$

Next, it is known that we can define a new permutation $\pi_{ijkl}$ such that it differs from $\pi$ only in absolutely bounded finite number of arguments and (cf. (3.14) of Chen and Fang (2015))

$$\mathcal{L}(\pi_{ijkl}) = \mathcal{L}(\pi|\pi(i)=k,\pi(j)=l). \tag{7.7}$$

Let

$$V_{ijkl} = \sum_{1\leqslant u<v\leqslant n}\frac{1}{n^{3/2}b^2/B_n^2}\Big[\mathbb{E}^{\pi_{ijkl}}(Y_{\pi_{ijkl}}^{(uv)})^2 - \mathbb{E}(Y_{\pi}^{(uv)})^2\Big].$$

From its construction and the bound $|\xi_{ij}| \leqslant C/n^{3/2}$, we have

$$|V_{ijkl} - V| \leqslant Cn\frac{1}{n^{3/2}} = \frac{C}{\sqrt{n}}.$$

Therefore, using the inequality $|e^x - 1| \leqslant |x|e^{|x|}$ for all $x \in \mathbb{R}$, we obtain

$$|e^{tV_{ijkl}} - e^{tV}| = |e^{t(V_{ijkl}-V)} - 1|e^{tV} \leqslant |t|\frac{C}{\sqrt{n}}e^{Ct/\sqrt{n}}e^{tV}. \tag{7.8}$$

Also, by (7.6) and (7.7),

$$h'(t) = \sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}a_{ijkl}\mathbb{E}[e^{tV_{ijkl}}].$$

In addition, observe that

$$\sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}a_{ijkl}\mathbb{E}[e^{tV}]=\mathbb{E}[e^{tV}]\sum_{1\leqslant i<j\leqslant n}\sum_{1\leqslant k\neq l\leqslant n}\mathbb{E}[\xi_{ij}1_{\{\pi(i)=k,\pi(j)=l\}}]$$
$$=\mathbb{E}[e^{tV}]\sum_{1\leqslant i<j\leqslant n}\mathbb{E}[\xi_{ij}]=0.$$

Consequently, for $t\leqslant\sqrt{n}$,

$$\begin{aligned}|h'(t)|&=\left|\sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}a_{ijkl}(\mathbb{E}[e^{tV_{ijkl}}]-\mathbb{E}[e^{tV}])\right|\\&\leqslant C\frac{|t|}{\sqrt{n}}\sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}|a_{ijkl}|\mathbb{E}[e^{tV}]\\&\leqslant C\frac{|t|}{n^2}\sum_{1\leqslant i<j\leqslant n}\frac{1}{n(n-1)}\sum_{1\leqslant k\neq l\leqslant n}\mathbb{E}[e^{tV}]\leqslant C|t|\mathbb{E}[e^{tV}]=C|t|h(t),\end{aligned}$$

where the first inequality follows by (7.8) and the second by $\|Y_{ij}\|_{\psi_1}\leqslant b/B_n$. Hence, by Gronwall's inequality, we have

$$h(t)=Ee^{tV}\leqslant e^{Ct^2}\text{ for }|t|\leqslant\sqrt{n}. \tag{7.9}$$

(7.9) means that $V$ is sub-gamma with variance factor $C$ and scale parameter $1/\sqrt{n}$ in the sense of Boucheron, Lugosi and Massart (2013, Section 2.4). Then, by Theorem 2.3 in Boucheron, Lugosi and Massart (2013) and Stirling's formula,

$$\|V\|_p\leqslant C(\sqrt{p}+p/\sqrt{n}),\quad p\geqslant 2$$

which is (7.5).

**Step 4. Bounding $D^4$.** We have

$$\begin{aligned}\lambda^{-1}\mathbb{E}[D^4 1_{\{|D|\leqslant\eta_t(p)\}}|\mathcal{G}]&=\frac{1}{n}\sum_{1\leqslant i<j\leqslant n}[(Y_\pi^{(ij)})^4 1_{\{|Y_\pi^{(ij)}|\leqslant\eta_t(p)\}}]\\=&\frac{1}{n}\sum_{1\leqslant i<j\leqslant n}\Big[(Y_\pi^{(ij)})^4 1_{\{|Y_\pi^{(ij)}|\leqslant\eta_t(p)\}}-\mathbb{E}^\pi(Y_\pi^{(ij)})^4 1_{\{|Y_\pi^{(ij)}|\leqslant\eta_t(p)\}}\Big]\\&+\frac{1}{n}\sum_{1\leqslant i<j\leqslant n}\mathbb{E}^\pi(Y_\pi^{(ij)})^4 1_{\{|Y_\pi^{(ij)}|\leqslant\eta_t(p)\}}.\end{aligned}$$

Following a similar argument as in the previous two steps, we obtain

$$\begin{aligned}&\int_0^\infty e^{-t}\min\left\{\frac{1}{\eta_t(p)},\frac{\|\mathbb{E}[D^4 1_{\{|D|\leqslant\eta_t(p)\}}|\mathcal{G}]\|_p}{\lambda\eta_t^3(p)}\right\}dt\\\leqslant&C\int_0^\infty e^{-t}\frac{p\sqrt{n}/B_n^2+p^{5/2}/B_n^2}{\sqrt{e^{2t}-1}}b^2dt+\int_0^\infty e^{-t}\min\left\{\frac{\sqrt{p}}{\sqrt{e^{2t}-1}},\frac{Cp^{3/2}nb^4}{B_n^4(e^{2t}-1)^{3/2}}\right\}dt\\\leqslant&C(\frac{p\sqrt{n}}{B_n^2}+\frac{p^{5/2}}{B_n^2})b^2.\end{aligned}$$

Combining all the above bounds proves (3.6). □

### 7.3 Proof for moderate deviations on Wiener chaos

Throughout this subsection, $C_q$ denotes a positive constant, which depends only on $q$ and may be different in different expressions. For the proof, in addition to Theorem 2.2, we use Latała (2006)'s sharp moment estimates for Gaussian homogeneous sums. For later use in Section 7.4, we state the following generalization obtained in Adamczak and Wolff (2015).

**Lemma 7.1** (Adamczak and Wolff (2015), Theorem 1.3)**.** Let $G$ be a standard Gaussian vector in $\mathbb{R}^n$. Then, for every polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most $q$ and every $p \geqslant 2$,

$$C_q^{-1}\sum_{r=1}^{q}\sum_{\mathcal{J}\in\Pi_r} p^{|\mathcal{J}|/2}\|\mathbb{E}\nabla^r Q(G)\|_{\mathcal{J}} \leqslant \|Q(G)-\mathbb{E}Q(G)\|_p \leqslant C_q\sum_{r=1}^{q}\sum_{\mathcal{J}\in\Pi_r} p^{|\mathcal{J}|/2}\|\mathbb{E}\nabla^r Q(G)\|_{\mathcal{J}},$$

where $\nabla^r Q$ is defined by (6.1) and we regard $\mathbb{E}\nabla^r Q(G)$ as an element of $(\mathbb{R}^n)^{\odot r}$.

The next result follows from Lemma 7.1 via a standard approximation argument.

**Lemma 7.2.** For any $h \in \mathfrak{H}^{\odot q}$ and $p \geqslant 2$,

$$\|I_q(h)\|_p \leqslant C_q \sum_{\mathcal{J}\in\Pi_q} p^{|\mathcal{J}|/2}\|h\|_{\mathcal{J}}. \tag{7.10}$$

*Proof.* We prove the claim when $\mathfrak{H}$ is infinite-dimensional; the finite-dimensional case is similar and easier. Let $(e_i)_{i=1}^{\infty}$ be an orthonormal basis of $\mathfrak{H}$. Then $(e_{i_1}\otimes\cdots\otimes e_{i_q})_{i_1,\dots,i_q=1}^{\infty}$ is an orthonormal basis of $\mathfrak{H}^{\otimes q}$. For every $n \in \mathbb{N}$, define

$$h_n := \sum_{i_1,\dots,i_q=1}^{n} a_{i_1,\dots,i_q} e_{i_1}\otimes\cdots\otimes e_{i_q},$$

where $a_{i_1,\dots,i_q} = \langle h, e_{i_1}\otimes\cdots\otimes e_{i_q}\rangle_{\mathfrak{H}^{\otimes q}}$. Then we have $\|h_n - h\|_{\mathfrak{H}^{\otimes q}} \to 0$ as $n \to \infty$. By hypercontractivity (cf. Theorem 2.7.2 of Nourdin and Peccati (2012)), this implies $\|I_q(h_n) - I_q(h)\|_p \to 0$ as $n \to \infty$. Also, it is straightforward to check that $\|h_n - h\|_{\mathcal{J}} \to 0$ as $n \to \infty$ for all $\mathcal{J} \in \Pi_q$. Therefore, it suffices to prove (7.10) with $h$ replaced by $h_n$.

By Theorems 2.7.7 and 2.7.10 in Nourdin and Peccati (2012), we have $I_q(h_n) = Q(X(e_1),\dots,X(e_n))$ for some polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most $q$. Then, for any $j_1,\dots,j_r \in [n]$,

$$\partial_{j_1,\dots,j_r} Q(X(e_1),\dots,X(e_n)) = \langle D^r I_q(h_n), e_{j_1}\otimes\cdots\otimes e_{j_r}\rangle_{\mathfrak{H}^{\otimes r}}.$$

Since $\mathbb{E}D^r I_q(h_n) = 0$ if $r < q$ and $D^q I_q(h_n) = q!h_n$, we obtain

$$\mathbb{E}\nabla^r Q(X(e_1),\dots,X(e_n)) = \begin{cases} 0 & \text{if } r < q, \\ q!A & \text{if } r = q, \end{cases}$$

where $A = (a_{i_1,\dots,i_q})_{1\leqslant i_1,\dots,i_q\leqslant n}$. Regarding $A$ as an element of $(\mathbb{R}^n)^{\odot q}$, we can easily check that $\|A\|_{\mathcal{J}} = \|h_n\|_{\mathcal{J}}$ for all $\mathcal{J} \in \Pi_q$. Thus, the desired result follows from Lemma 7.1. □

*Proof of Theorem 3.2.* According to Theorem 2.2, it suffices to prove

$$\mathcal{W}_p(W,Z) \leqslant C_q \max_{r\in[q-1]} \max_{\mathcal{J}\in\Pi_{2q-2r}} p^{(1+|\mathcal{J}|)/2}\|f\widetilde{\otimes}_r f\|_{\mathcal{J}} \tag{7.11}$$

for all $p \geqslant 2$. By Proposition 3.7 in Nourdin, Peccati and Swan (2014),

$$\tau(w) = \mathbb{E}[\langle -DL^{-1}W, DW\rangle_{\mathfrak{H}} | W = w], \quad w \in \mathbb{R},$$

gives a Stein kernel for $W$ (in the sense that it satisfies Eq.(2.3) in Ledoux, Nourdin and Peccati (2015) with $\nu$ the law of $W$). Hence, using the Stein kernel bound for $p$-Wasserstein distance (cf. Proposition 3.4(ii) in Ledoux, Nourdin and Peccati (2015)), we obtain

$$\mathcal{W}_p(W,Z) \leqslant C\sqrt{p}\|\tau(W)-1\|_p.$$

By Eq.(5.2.2) in Nourdin and Peccati (2012),

$$\tau(W)=\frac{1}{q}\|DW\|_{\mathfrak{H}}^2=1+q\sum_{r=1}^{q-1}(r-1)!\binom{q-1}{r-1}^2 I_{2q-2r}(f\widetilde{\otimes}_r f).$$

Thus, by Minkowski's inequality and Lemma 7.2,

$$\|\tau(W)-1\|_p \leqslant C_q\sum_{r=1}^{q-1}\sum_{\mathcal{J}\in\Pi_{2q-2r}} p^{|\mathcal{J}|/2}\|f\widetilde{\otimes}_r f\|_{\mathcal{J}}.$$

Consequently, we obtain (7.11). □

### 7.4 Proof for homogeneous sums

Throughout this section, $C$ denotes a positive absolute constant and $C_q$ denotes a positive constant depending only on $q$, respectively. Note that their values may be different in different expressions. Also, given a function $g:[n]^q\to\mathbb{R}$, we write

$$\|g\|=\sqrt{\sum_{i_1,\dots,i_q=1}^n g(i_1,\dots,i_q)^2}.$$

We will frequently use the following inequality throughout the proof.

**Lemma 7.3** (Adamczak and Wolff (2015), Theorem 1.4)**.** Let $X=(X_1,\dots,X_n)$ be a random vector with independent components. Suppose that there is a constant $K>0$ such that $\|X_i\|_{\psi_2}\leqslant K$ for all $i=1,\dots,n$. Then, for every polynomial $Q:\mathbb{R}^n\to\mathbb{R}$ of degree at most $q$ and every $p\geqslant 2$,

$$\|Q(X)-\mathbb{E}Q(X)\|_p \leqslant C_q\sum_{r=1}^{q}K^r\sum_{\mathcal{J}\in\Pi_r} p^{|\mathcal{J}|/2}\|\mathbb{E}\nabla^r Q(X)\|_{\mathcal{J}}.$$

*Proof of Proposition* 3.2. First, note that $\mathcal{M}(f)\leqslant\|f\|^2=1/q!\leqslant 1/2$. Hence $|\log\mathcal{M}(f)|\geqslant\log 2$ and $p\mathcal{M}(f)\leqslant p\sqrt{\mathcal{M}(f)}\leqslant 1$.

**Step 1. The exchangeable pair.** Let $X^*=(X_1^*,\dots,X_n^*)$ be an independent copy of $X:=(X_1,\dots,X_n)$. Also, let $I\sim\text{Unif}[n]$ be an index independent of $X$ and $X^*$. Define $X'=(X_1',\dots,X_n')$ by

$$X_i'=\begin{cases}X_i^*, & \text{if } i=I,\\ X_i, & \text{otherwise.}\end{cases}$$

Then we set

$$W'=\sum_{i_1,\dots,i_q=1}^n f(i_1,\dots,i_q)X_{i_1}'\cdots X_{i_q}'.$$

It is easy to check $\mathcal{L}(X,X')=\mathcal{L}(X',X)$; hence, $\mathcal{L}(W,W')=\mathcal{L}(W',W)$. Moreover,

$$\begin{aligned}D:=W'-W&=\sum_{\substack{i_1,\dots,i_q=1\\ \exists r:i_r=I}}^n f(i_1,\dots,i_q)(X_I'-X_I)\prod_{r=1:i_r\neq I}^{q}X_{i_r}\\ &=q(X_I'-X_I)Q_I(X),\end{aligned}$$

where, for every $i = 1, \dots, n$, $Q_i$ is an $n$-variate polynomial defined as

$$Q_i(x_1, \dots, x_n) := \sum_{i_2, \dots, i_q = 1}^{n} f(i, i_2, \dots, i_q) x_{i_2} \cdots x_{i_q}.$$

Hence

$$\mathbb{E}[D|X] = -\frac{q}{n} W.$$

Therefore, by Corollary 2.1

$$\mathcal{W}_p(W, Z) \leqslant C\sqrt{p}\|E\|_p + Cp\sqrt{\frac{n}{q}\|\mathbb{E}[D^4|X]\|_p} =: H_1 + H_2, \tag{7.12}$$

where

$$E = \frac{n}{2q}\mathbb{E}[D^2|X] - 1.$$

**Step 2. Bounding $H_1$.** Observe that

$$\frac{n}{2q}\mathbb{E}[D^2|X] = \frac{q}{2}\sum_{i=1}^{n}(1 + X_i^2)Q_i(X)^2.$$

Define an $n$-variate polynomial $Q$ as

$$Q(x_1, \dots, x_n) = \frac{q}{2}\sum_{i=1}^{n}(1 + x_i^2)Q_i(x_1, \dots, x_n)^2.$$

Observe that $Q$ has total degree $2q$ and degree 2 in $x_i$ for every $i \in [n]$; the latter follows from the fact that $f$ is vanishing on diagonals. Using the latter property, one can easily verify that, with $G \sim N(0, I_n)$, $\mathbb{E}Q(X) = \mathbb{E}Q(G)$ and $\mathbb{E}\nabla^r Q(X) = \mathbb{E}\nabla^r Q(G)$ for all $r = 1, \dots, 2q$. Hence, by Lemmas 7.1 and 7.3,

$$H_1 \leqslant C_q\sqrt{p}K^{2q}\|Q(G) - 1\|_p. \tag{7.13}$$

Let $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$. Without loss of generality, we may assume that $G_i = \mathbf{G}(e_i)$ $(i = 1, \dots, n)$ for some isonormal Gaussian process $\mathbf{G}$ over $\mathfrak{H} = \mathbb{R}^n$. Then, for every $i = 1, \dots, n$, we have

$$Q_i(G) = I_{q-1}(\mathbf{f}_i),$$

where $I_q$ denotes the $q$-th multiple Wiener–Itô integral with respect to $\mathbf{G}$ and

$$\mathbf{f}_i := \sum_{i_2, \dots, i_q = 1}^{n} f(i, i_2, \dots, i_q) e_{i_2} \otimes \cdots \otimes e_{i_q}.$$

Thus we obtain

$$\begin{aligned} Q(G) - 1 &= \frac{q}{2}\sum_{i=1}^{n}(1 + G_i^2)I_{q-1}(\mathbf{f}_i)^2 - 1 \\ &= \frac{q}{2}\sum_{i=1}^{n}(G_i^2 - 1)I_{q-1}(\mathbf{f}_i)^2 + \left\{q\sum_{i=1}^{n}I_{q-1}(\mathbf{f}_i)^2 - 1\right\} =: H_{11} + H_{12}. \end{aligned}$$

To evaluate $H_{11}$, observe that $G_i^2 - 1 = I_2(e_i^{\otimes 2})$ by Theorem 2.7.7 in Nourdin and Peccati (2012). Also, by the product formula for multiple Wiener–Itô integrals (cf. Theorem 2.7.10 in Nourdin and Peccati (2012)),

$$I_{q-1}(\mathbf{f}_i)^2 = \sum_{r=0}^{q-1} r! \binom{q-1}{r}^2 I_{2q-2-2r}(\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i).$$

Using the product formula again and noting that $f(i, i_2, \dots, i_q) = 0$ if $i_r = i$ for some $r$ as well as $e_i \cdot e_j = 0$ if $i \neq j$, we obtain

$$H_{11} = \frac{q}{2} \sum_{r=0}^{q-1} r! \binom{q-1}{r}^2 I_{2q-2r} \left( \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right). \tag{7.14}$$

Let $r \in \{0, 1, \dots, q-1\}$ be fixed. By Lemma 7.2,

$$\left\| I_{2q-2r} \left( \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right) \right\|_p \leqslant C_q \sum_{\mathcal{J} \in \Pi_{2q-2r}} p^{|\mathcal{J}|/2} \left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\mathcal{J}}. \tag{7.15}$$

Observe that

$$\left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\{1\},\dots,\{2q-2r\}} \leqslant \sup_{u \in \mathbb{R}^n : |u| \leqslant 1} \sum_{i=1}^{n} u_i^2 \| \mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i \|_{\mathfrak{H}^{\otimes(2q-2r-2)}}$$

and

$$\left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\mathcal{J}} \leqslant \left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\mathfrak{H}^{\otimes(2q-2r)}} \leqslant \sqrt{\sum_{i=1}^{n} \| \mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i \|^2_{\mathfrak{H}^{\otimes(2q-2r-2)}}}$$

for any $\mathcal{J} \in \Pi_{2q-2r}$. By the Cauchy–Schwarz inequality,

$$\| \mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i \|_{\mathfrak{H}^{\otimes(2q-2r-2)}} \leqslant \sum_{i_2, \dots, i_q = 1}^{n} f(i, i_2, \dots, i_q)^2.$$

Hence we obtain

$$\left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\{1\},\dots,\{2q-2r\}} \leqslant \mathcal{M}(f)$$

and

$$\left\| \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right\|_{\mathcal{J}} \leqslant \sqrt{\mathcal{M}(f) \sum_{i_1, \dots, i_q = 1}^{n} f(i_1, \dots, i_q)^2} = \sqrt{\frac{1}{q!} \mathcal{M}(f)}$$

for any $\mathcal{J} \in \Pi_{2q-2r}$. Inserting these estimates into (7.15), we deduce

$$\left\| I_{2q-2r} \left( \sum_{i=1}^{n} e_i^{\otimes 2} \widetilde{\otimes} (\mathbf{f}_i \widetilde{\otimes}_r \mathbf{f}_i) \right) \right\|_p \leqslant C_q \left( p^{q-r-1/2} \sqrt{\mathcal{M}(f)} + p^{q-r} \mathcal{M}(f) \right).$$

Combining this bound with (7.14) and $p\mathcal{M}(f) \leqslant 1$, we obtain

$$\|H_{11}\|_p \leqslant C_q p^{q-1/2} \sqrt{\mathcal{M}(f)}. \tag{7.16}$$

To evaluate $H_{12}$, observe that $I_{q-1}(\mathbf{f}_i) = q^{-1}\mathbf{D}I_q(\mathbf{f}) \cdot e_i$ for every $i = 1, \ldots, n$, where $\mathbf{D}$ denotes the Malliavin derivative with respect to $\mathbf{G}$ and

$$\mathbf{f} := \sum_{i_1,\ldots,i_q=1}^{n} f(i_1, \ldots, i_q) e_{i_1} \otimes \cdots \otimes e_{i_q}.$$

Hence

$$H_{12} = q^{-1} \sum_{i=1}^{n} (\mathbf{D}I_q(\mathbf{f}) \cdot e_i)^2 - 1 = q^{-1} \|\mathbf{D}I_q(\mathbf{f})\|_{\mathfrak{H}}^2 - 1.$$

Therefore, by the proof of Theorem 3.2,

$$\|H_{12}\|_p \leqslant C_q \sum_{r=1}^{q-1} \sum_{\mathcal{J} \in \Pi_{2q-2r}} p^{|\mathcal{J}|/2} \|\mathbf{f} \widetilde{\otimes}_r \mathbf{f}\|_{\mathcal{J}} \leqslant C_q p^{q-1} \max_{r \in [q-1]} \|f \otimes_r f\|,$$

where, for every $r \in [q]$, the function $f \otimes_r f : [n]^{2q-2r} \to \mathbb{R}$ is defined as

$$f \otimes_r f(i_1, \ldots, i_{2q-2r}) = \sum_{j_1,\ldots,j_r=1}^{n} f(i_1, \ldots, i_{q-r}, j_1, \ldots, j_r) f(i_{q-r+1}, \ldots, i_{2q-2r}, j_1, \ldots, j_r).$$

Combining this with Lemma 2.1 in Koike (2023), we obtain

$$\|H_{12}\|_p \leqslant C_q p^{q-1} \sqrt{|\mathbb{E}W^4 - 3| + M^q \mathcal{M}(f)}. \tag{7.17}$$

By (7.13), (7.16) and (7.17), we conclude

$$H_1 \leqslant C_q p^q K^{2q} \sqrt{|\mathbb{E}W^4 - 3| + M^q \mathcal{M}(f)}. \tag{7.18}$$

**Step 3. Bounding $H_2$.** First, by Lemma 7.3

$$\|Q_i(X)\|_s \leqslant C_q K^{q-1} s^{(q-1)/2} \sqrt{\mathrm{Inf}_i(f)}$$

for any $i \in [n]$ and $s \geqslant 2$, where

$$\mathrm{Inf}_i(f) := \sum_{i_2,\ldots,i_q=1}^{n} f(i, i_2, \ldots, i_q)^2.$$

Hence we have (cf. Lemma A.4 in Koike (2023))

$$P(|Q_i(X)| \geqslant t) \leqslant C_q \exp\left( - \left( \frac{t}{C_q' K^{q-1} \sqrt{\mathrm{Inf}_i(f)}} \right)^{2/(q-1)} \right)$$

for all $t > 0$, where $C_q' > 0$ is a constant depending only on $q$. Let

$$\delta_i := C_q' K^{q-1} \sqrt{\mathrm{Inf}_i(f)} |pq \log \mathcal{M}(f)|^{(q-1)/2}.$$

Then, by Lemma 6.1 in Koike (2023),

$$\mathbb{E}[|Q_i(X)|^s 1_{\{|Q_i(X)| > \delta_i\}}] \leqslant C_q \left( 1 + \frac{2s - 2/(q-1)}{s - 2/(q-1)} \right) \{s(q-1)\}^{s(q-1)/2} \delta_i^s \mathcal{M}(f)^{pq}$$

for any $s > 2/(q-1)$. Since $2/(q-1) \leqslant 2$, we can apply this inequality with $s = 4p$ and then obtain

$$\|Q_i(X)^4 1_{\{|Q_i(X)| > \delta_i\}}\|_p \leqslant C_q p^{2(q-1)} \delta_i^4 \mathcal{M}(f)^q. \tag{7.19}$$

Now we bound $\frac{n}{q}\mathbb{E}[D^4|X]$ as

$$\begin{aligned}\frac{n}{q}\mathbb{E}[D^4|X] &= q^3\sum_{i=1}^n \mathbb{E}[(X_i'-X_i)^4|X]Q_i(X)^4 \\ &\leqslant q^3\sum_{i=1}^n \mathbb{E}[(X_i'-X_i)^4|X]\delta_i^4 + q^3\sum_{i=1}^n \mathbb{E}[(X_i'-X_i)^4|X]Q_i(X)^4 1_{\{|Q_i(X)|>\delta_i\}} \\ &=: H_{21}+H_{22}.\end{aligned} \tag{7.20}$$

We bound $\|H_{21}\|_p$ as

$$\|H_{21}\|_p \leqslant q^3\sum_{i=1}^n \mathbb{E}[(X_i'-X_i)^4]\delta_i^4 + q^3\left\|\sum_{i=1}^n\{\mathbb{E}[(X_i'-X_i)^4|X]-\mathbb{E}(X_i'-X_i)^4\}\delta_i^4\right\|_p.$$

For the first term, we have

$$q^3\sum_{i=1}^n \mathbb{E}[(X_i'-X_i)^4]\delta_i^4 \leqslant C_qK^4\sum_{i=1}^n\delta_i^4 \leqslant C_q p^{2q-2}K^{4q}\mathcal{M}(f)|\log\mathcal{M}(f)|^{2(q-1)}.$$

To bound the second term, note that $\|\mathbb{E}[(X_i'-X_i)^4|X]\|_{\psi_{1/2}} \leqslant CK^4$. Therefore, by Lemma 2.1,

$$\begin{aligned}&\left\|\sum_{i=1}^n\{\mathbb{E}[(X_i'-X_i)^4|X]-\mathbb{E}(X_i'-X_i)^4\}\delta_i^4\right\|_p \\ &\leqslant CK^4\left(\sqrt{p\sum_{i=1}^n\delta_i^8}+p^2\max_{1\leqslant i\leqslant n}\delta_i^4\right) \\ &\leqslant C_qK^{4q}(p^{2q-3/2}\mathcal{M}(f)^{3/2}+p^{2q}\mathcal{M}(f)^2)|\log\mathcal{M}(f)|^{2(q-1)} \\ &\leqslant C_qK^{4q}p^{2q-2}\mathcal{M}(f)|\log\mathcal{M}(f)|^{2(q-1)},\end{aligned}$$

where in the second inequality we used $\sum_{i=1}^n \mathrm{Inf}_i(f)=1/q!$ and the last inequality follows from the condition $p\mathcal{M}(f)\leqslant p\sqrt{\mathcal{M}(f)}\leqslant 1$. All together, we obtain

$$\|H_{21}\|_p \leqslant C_qK^{4q}p^{2q-2}\mathcal{M}(f)|\log\mathcal{M}(f)|^{2(q-1)}. \tag{7.21}$$

In the meantime, noting that $(X_i,X_i')$ and $Q_i(X)$ are independent, we have

$$\|H_{22}\|_p \leqslant q^3\sum_{i=1}^n \|(X_i'-X_i)^4\|_p\|Q_i(X)^4 1_{\{|Q_i(X)|>\delta_i\}}\|_p.$$

Using (7.19) and $p\sqrt{\mathcal{M}(f)}\leqslant 1$, we obtain

$$\|H_{22}\|_p \leqslant C_qK^4p^{2q}\mathcal{M}(f)^q\sum_{i=1}^n\delta_i^4 \leqslant C_qK^{4q}p^{2q-2}\mathcal{M}(f)|\log\mathcal{M}(f)|^{2(q-1)}. \tag{7.22}$$

Combining (7.21) and (7.22) with (7.20) gives

$$H_2 \leqslant Cp\sqrt{H_{21}+H_{22}} \leqslant C_qK^{2q}p^q\sqrt{\mathcal{M}(f)}|\log\mathcal{M}(f)|^{q-1}. \tag{7.23}$$

By (7.12), (7.18) and (7.23), we complete the proof. □

### 7.5 Proof for moderate deviations in multi-dimensions

*Proof of Theorem* 4.1. The proof is almost identical to the arguments leading to (2.12), except that we view $Y_i^{\otimes 2}$ as a $d^2$-vector, use $\|Y_i^{\otimes 2}\|_{H.S.} = |Y_i|^2$ and Lemma 2.1 for independent random vectors in $\mathbb{R}^{d^2}$. The factor $d^{1/4}$ comes from the $\sqrt{d}$ term in (2.2). □

*Proof of Theorem* 4.2. In this proof, we use $C := C_{A,\alpha,B_1,B_2}$ to denote positive constants, which depend only on $\alpha$, $A$, $B_1$ and $B_2$ and may be different in different expressions. Let $f(x) := f(x;d)$ denote the density of the chi-distribution with $d$ degrees of freedom, i.e.,

$$f(x) = \frac{1}{\kappa(d)} x^{d-1} e^{-x^2/2}, \quad \kappa(d) := 2^{(d/2)-1}\Gamma(d/2).$$

Note that $\log(\kappa(d)) \leqslant Cd\log d$. For $d \geqslant 2$ and $x > 0$, we have

$$\int_x^\infty y^{d-1}e^{-y^2/2}dy = x^{d-2}e^{-x^2/2} + \int_x^\infty (d-2)y^{d-3}e^{-y^2/2}dy \geqslant x^{d-2}e^{-x^2/2}.$$

Therefore,

$$\frac{f(x)}{P(|Z|>x)} \leqslant x. \tag{7.24}$$

First we prove the claim when $\Delta < 1/e$. Set

$$p = |\log \Delta| + \log(\kappa(d)) + \frac{x^2}{2}, \quad \varepsilon = Ap^\alpha \Delta e.$$

Because of the condition $|\log \Delta| \leqslant p_0/4$, $\log(\kappa(d)) \leqslant p_0/4$ and $x \leqslant \sqrt{p_0}$, we have $p \leqslant p_0$. From the upper bound on $\mathcal{W}_p(W,Z)$, we can couple $W$ and $Z$ such that $\|W-Z\|_p \leqslant Ap^\alpha\Delta$. We have

$$\begin{aligned} P(|W|>x) &\leqslant P(|Z|>x-\varepsilon) + P(|W-Z|>\varepsilon) \\ &= P(|Z|>x) + P(x-\varepsilon<|Z|\leqslant x) + P(|W-Z|>\varepsilon). \end{aligned}$$

Since

$$P(x-\varepsilon<|Z|\leqslant x) = \int_{(x-\varepsilon)\vee 0}^x f(z)dz$$

and

$$P(|W-Z|>\varepsilon) \leqslant \varepsilon^{-p}\|W-Z\|_p^p \leqslant (Ap^\alpha\Delta/\varepsilon)^p = e^{-p} = \Delta\frac{1}{\kappa(d)}e^{-x^2/2},$$

we obtain

$$P(|W|>x) \leqslant P(|Z|>x) + \int_{(x-\varepsilon)\vee 0}^x f(z)dz + \Delta\frac{1}{\kappa(d)}e^{-x^2/2}.$$

Similarly, we deduce

$$\begin{aligned} P(|Z|>x) &= P(|Z|>x+\varepsilon) + P(x<|Z|\leqslant x+\varepsilon) \\ &\leqslant P(|W|>x) + P(|W-Z|>\varepsilon) + P(x<|Z|\leqslant x+\varepsilon) \\ &\leqslant P(|W|>x) + \int_x^{x+\varepsilon} f(z)dz + \Delta\frac{1}{\kappa(d)}e^{-x^2/2}. \end{aligned}$$

Consequently, we obtain

$$|P(|W|>x) - P(|Z|>x)| \leqslant \int_{(x-\varepsilon)\vee 0}^{x+\varepsilon} f(z)dz + \Delta\frac{1}{\kappa(d)}e^{-x^2/2}.$$

Note that (4.2) implies $d(\log d)\Delta^{2/(2\alpha+1)} \leqslant C$. Therefore, using $x \leqslant \Delta^{-1/(2\alpha+1)}$, we have

$$\begin{aligned} \varepsilon &\leqslant C\Delta(|\log\Delta|^\alpha + \log^\alpha(\kappa(d)) + x^{2\alpha}) \leqslant C\Delta(|\log\Delta|^\alpha + \Delta^{-2\alpha/(2\alpha+1)}) \\ &\leqslant C\Delta^{1-2\alpha/(2\alpha+1)} = C\Delta^{1/(2\alpha+1)}. \end{aligned} \tag{7.25}$$

Note that $\varepsilon \leqslant C$ and for $0 \leqslant x \leqslant \Delta^{-1/(2\alpha+1)}$, we have $x\varepsilon \leqslant C$. Also note that (4.2) implies $d\Delta|\log\Delta|^\alpha \leqslant Cd\Delta^{2/(2\alpha+1)} \leqslant C$ if $\alpha > 1/2$. If $x \geqslant 1$, we have, from (7.24),

$$\begin{aligned} &\frac{\int_{(x-\varepsilon)\vee 0}^{x+\varepsilon} f(z)dz}{P(|Z|>x)} \leqslant C\varepsilon\frac{f(x)}{P(|Z|>x)}e^{x\varepsilon}(\frac{x+\varepsilon}{x})^{d-1} \leqslant Cx\varepsilon e^{d\varepsilon/x} \\ \leqslant& Cx\varepsilon\exp\left\{C\left(d\Delta|\log\Delta|^\alpha + d(d\log d)^\alpha\Delta + d\Delta + d\Delta^{2/(2\alpha+1)}1_{\{\alpha>1/2\}}\right)\right\} \\ \leqslant& Cx\varepsilon, \end{aligned}$$

where we used $1 \leqslant x \leqslant \Delta^{-1/(2\alpha+1)}$, (4.2) and (4.3). Therefore,

$$\left|\frac{P(|W|>x)}{P(|Z|>x)} - 1\right| \leqslant Cx\varepsilon + \Delta \leqslant C(1+x)(|\log\Delta| + d\log d + x^2)^\alpha\Delta.$$

If $x < 1$, the conclusion follows from $1/P(|Z|>x) \leqslant C$ and

$$|P(|W|>x) - P(|Z|>x)| \leqslant C(\varepsilon + \Delta).$$

It remains to prove (4.4) when $\Delta \geqslant 1/e$. In this case, we have $x \leqslant e$ and thus $1/P(|Z|>x)$ is bounded. Hence (4.4) holds with a sufficiently large $C_{A,\alpha,B_1,B_2}$. □

## 7.6 Proof for local dependence

*Proof of Theorem* 5.1. We adapt the proof of Theorem 2.1 and use the notation therein. Let $\mathcal{G} = \sigma(X_1,\dots,X_n)$. Let $I$ be a uniform random index from $\{1,\dots,n\}$ and independent of everything else. Let $D = -n^{-1/2}\sum_{j\in A_I} X_j$. Because $|X_j| \leqslant b_n$ and $|A_I| \leqslant \theta_1$, we have $|D| \leqslant \theta_1 b_n/\sqrt{n}$. Because $X_i$ is independent of $\{X_j : j \notin A_i\}$, we have

$$\mathbb{E}[(-\sqrt{n}X_I)T_t h(W+D)] = 0,$$

and hence,

$$\mathbb{E}\left[(-\sqrt{n}X_I)\left\{T_t h(W) + \langle\nabla T_t h(W), D\rangle + \sum_{k=2}^\infty \frac{1}{k!}\langle\nabla^k T_t h(W), D^{\otimes k}\rangle\right\}\right] = 0.$$

Let

$$\tau_t = \mathbb{E}\left[(-\sqrt{n}X_I)\left(1 - \frac{\langle\nabla\phi(Z), D\rangle}{\phi(Z)\sqrt{e^{2t}-1}} + \sum_{k=2}^\infty \frac{1}{k!}\frac{(-1)^k\langle\nabla^k\phi(Z), D^{\otimes k}\rangle}{\phi(Z)(e^{2t}-1)^{k/2}}\right)|\mathcal{G}\vee\sigma(Z)\right].$$

Following the same argument leading to (6.12), we have

$$\begin{aligned} \|\rho_t(F_t)\|_p \leqslant e^{-t}&\left(\frac{1}{\sqrt{e^{2t}-1}}\|EZ\|_p\right. \\ &\left.+\sum_{k=2}^\infty \frac{1}{k!(e^{2t}-1)^{k/2}}\left\|\mathbb{E}\left[(-\sqrt{n}X_I)\frac{\langle\nabla^k\phi(Z), D^{\otimes k}\rangle}{\phi(Z)}|\mathcal{G}\vee\sigma(Z)\right]\right\|_p\right), \end{aligned}$$

where $E = \mathbb{E}[(-\sqrt{n}X_I)\otimes D|\mathcal{G}] - I_d$. Following the same argument as in the proof of Proposition 6.1, with $\gamma_t = \sqrt{e^{2t}-1}$, the first term is bounded by $\frac{Ce^{-t}}{\gamma_t}\sqrt{p}\|E\|_p$. The second term with $k=2$, is bounded by

$$\frac{Ce^{-t}p\theta_1^2 b_n^3}{\sqrt{n}\gamma_t^2}.$$

The second term with $k \geqslant 3$, if $\gamma_t \geqslant \theta_1 b_n \sqrt{p/n}$, is bounded by

$$\frac{Ce^{-t}p^{3/2}\theta_1^3 b_n^4}{n\gamma_t^3}.$$

Note that

$$d = \mathbb{E}(W^T W) \leqslant \frac{n\theta_1}{n} b_n^2 = \theta_1 b_n^2.$$

Let $t_0$ be such that $\sqrt{e^{2t_0}-1} = \theta_1 b_n \sqrt{p/n}$ and assume it is $\leqslant c$ for a sufficiently small constant $c > 0$ as in the condition (5.1). Then, with $W_0 := e^{-t_0}W + \sqrt{1-e^{-2t_0}}Z$, we have

$$\begin{aligned}
\mathcal{W}_p(W_0, Z) \leqslant & \int_{t_0}^{\infty} \|\rho_t(F_t)\|_p dt \\
\leqslant & C\sqrt{p}\|E\|_p + C\int_{t_0}^{\infty} \frac{e^{-t}}{e^{2t}-1} dt \frac{p\theta_1^2 b_n^3}{\sqrt{n}} + C\int_{t_0}^{\infty} \frac{e^{-t}}{(e^{2t}-1)^{3/2}} dt \frac{p^{3/2}\theta_1^3 b_n^4}{n} \\
\leqslant & C\sqrt{p}\|E\|_p + \frac{Cp\theta_1^2 b_n^3 \log n}{\sqrt{n}}.
\end{aligned}$$

This implies

$$\begin{aligned}
\mathcal{W}_p(W, Z) = & e^{t_0}\mathcal{W}_p(e^{-t_0}W, e^{-t_0}Z) \\
\leqslant & e^{t_0}\mathcal{W}_p(e^{-t_0}W, W_0) + e^{t_0}\mathcal{W}_p(W_0, Z) + e^{t_0}\mathcal{W}_p(Z, e^{-t_0}Z) \\
\leqslant & C\sqrt{p}\|E\|_p + \frac{Cp\theta_1^2 b_n^3 \log n}{\sqrt{n}} + \frac{Cpd^{1/2}\theta_1 b_n}{\sqrt{n}} \\
\leqslant & C\sqrt{p}\|E\|_p + \frac{Cp\theta_1^2 b_n^3 \log n}{\sqrt{n}},
\end{aligned}$$

where we used $\|\sqrt{1-e^{-2t_0}}Z\|_p \leqslant C\theta_1 b_n \sqrt{p/n}\sqrt{dp}$ (cf. Lemma 6.3) in the second inequality and $d \leqslant \theta_1 b_n^2$ in the last inequality. Note that

$$E = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in A_i}(X_i \otimes X_j) - I_d.$$

Denote the $(u,v)$-entry of the $d \times d$ matrix $E$ by $E_{uv}$. Then, for $p \geqslant 2$,

$$\|E\|_p = \left[\mathbb{E}(\sum_{u,v=1}^{d} E_{uv}^2)^{p/2}\right]^{1/p} \leqslant d \max_{u,v} \|E_{uv}\|_p.$$

Write $X_i = (X_{i1}, \dots, X_{id})^T$ and, from $\mathbb{E}(E_{uv}) = 0$,

$$\begin{aligned}
E_{uv} = & \frac{1}{n}\sum_{i=1}^{n} X_{iu}\sum_{j\in A_i} X_{jv} - \delta_{uv} = \frac{2(\theta_1\theta_2)^{1/2}b_n'^2}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j\in A_i}\left[\frac{X_{iu}X_{jv} - \mathbb{E}(X_{iu}X_{jv})}{2(\theta_1\theta_2)^{1/2}b_n'^2\sqrt{n}}\right] \\
=: & \frac{2(\theta_1\theta_2)^{1/2}b_n'^2}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j\in A_i} X_{ij}^{uv} =: \frac{2(\theta_1\theta_2)^{1/2}b_n'^2}{\sqrt{n}} V_{uv}.
\end{aligned}$$

In the remainder of this proof, we show that if $2 \leqslant p \leqslant \theta_1 n/\theta_2$ as in the condition (5.1), then

$$\|V_{uv}\|_p \leqslant C\sqrt{p}, \tag{7.26}$$

and hence conclude (5.2).

Let $V_{uv}^{(ij)} = V_{uv} - \sum_{(k,l)\in B_{ij}} X_{kl}^{uv}$. Then $|V_{uv} - V_{uv}^{(ij)}| \leqslant \sqrt{\theta_2}/\sqrt{\theta_1 n}$. Hence, similarly to the derivation of (7.8), we obtain

$$|e^{tV_{uv}} - e^{tV_{uv}^{(ij)}}| \leqslant |t|\sqrt{\frac{\theta_2}{\theta_1 n}} e^{t\sqrt{\theta_2}/\sqrt{\theta_1 n}} e^{tV_{uv}} \tag{7.27}$$

for any $t \in \mathbb{R}$. Also, with $h(t) = \mathbb{E}e^{tV_{uv}}$, we have

$$h'(t) = \mathbb{E}[V_{uv}e^{tV_{uv}}] = \sum_{i=1}^n \sum_{j\in A_i} \mathbb{E}[X_{ij}^{uv} e^{tV_{uv}}].$$

Further, since

$$V_{uv}^{(ij)} = \sum_{k=1}^n \sum_{l\in A_k} X_{kl}^{uv} - \sum_{k=1}^n \sum_{l\in A_k : k \text{ or } l\in A_{ij}} X_{kl}^{uv} = \sum_{k\notin A_{ij}} \sum_{l\in A_k : l\notin A_{ij}} X_{kl}^{uv},$$

$V_{uv}^{(ij)}$ is independent of $X_{ij}^{uv}$. Hence

$$\sum_{i=1}^n \sum_{j\in A_i} \mathbb{E}[X_{ij}^{uv} e^{tV_{uv}^{(ij)}}] = \sum_{i=1}^n \sum_{j\in A_i} \mathbb{E}[X_{ij}^{uv}]\mathbb{E}[e^{tV_{uv}^{(ij)}}] = 0.$$

Consequently, for $|t| \leqslant \sqrt{\theta_1 n}/\sqrt{\theta_2}$,

$$\begin{aligned} |h'(t)| &= \left| \sum_{i=1}^n \sum_{j\in A_i} \mathbb{E}[X_{ij}^{uv}(e^{tV_{uv}} - e^{tV_{uv}^{(ij)}})] \right| \leqslant C|t|\sqrt{\frac{\theta_2}{\theta_1 n}} \sum_{i=1}^n \sum_{j\in A_i} \mathbb{E}[|X_{ij}^{uv}|e^{tV_{uv}}] \\ &\leqslant C|t|\sqrt{\frac{\theta_2}{\theta_1 n}} \sum_{i=1}^n \sum_{j\in A_i} \frac{\mathbb{E}[e^{tV_{uv}}]}{(\theta_1\theta_2)^{1/2}\sqrt{n}} \leqslant C|t|\mathbb{E}[e^{tV_{uv}}] = C|t|h(t), \end{aligned}$$

where the first inequality follows by (7.27), the second by $|X_{ij}| \leqslant b_n'$, and the third by $|A_i| \leqslant \theta_1$. Hence, by Gronwall's inequality,

$$h(t) \leqslant e^{Ct^2} \text{ for } |t| \leqslant \sqrt{\theta_1 n}/\sqrt{\theta_2}. \tag{7.28}$$

(7.28) means that $V_{uv}$ is sub-gamma with variance factor $C$ and scale parameter $\sqrt{\theta_2}/\sqrt{\theta_1 n}$ in the sense of Boucheron, Lugosi and Massart (2013, Section 2.4). Then, by Theorem 2.3 in Boucheron, Lugosi and Massart (2013) and Stirling's formula,

$$\|V_{uv}\|_p \leqslant C(\sqrt{p} + p\sqrt{\theta_2}/\sqrt{\theta_1 n}) \leqslant C\sqrt{p},$$

where the last inequality follows by (5.1). This proves (7.26). □

*Proof of Theorem* 5.3. We use $C$ to denote positive absolute constants, which may differ in different expressions. We first do truncation. Let $\tilde{X}_{ij} := X_{ij}1_{\{|X_{ij}|\leqslant b\log n\}} - \mathbb{E}X_{ij}1_{\{|X_{ij}|\leqslant b\log n\}}$, $\tilde{X}_i = (\tilde{X}_{i1}, \dots, \tilde{X}_{id})^T$, $W^{(l)} = n^{-1/2}\sum_{i\in g_l} X_i$, $\tilde{W}^{(l)} = n^{-1/2}\sum_{i\in g_l}\tilde{X}_i$ and $\tilde{W} = \sum_{l=1}^L \tilde{W}^{(l)} = n^{-1/2}\sum_{i=1}^n \tilde{X}_i$.

From $\|X_{ij}\|_{\psi_1} \leqslant b$ and Koike (2021, Lemma 5.4), we have, for every positive integer $p$,

$$\begin{aligned} &\mathbb{E}[|n^{-1/2}(X_{ij} - \tilde{X}_{ij})|^p] \leqslant n^{-p/2}2^{p-1}\mathbb{E}[|X_{ij}|^p 1_{\{|X_{ij}|>b\log n\}}] \\ \leqslant& n^{-p/2}2^{p-1}p!2e^{-b\log n/b}(b\log n + b)^p \\ =& \frac{p!}{2}\left(\frac{2b\log n + 2b}{\sqrt{n}}\right)^{p-2} \frac{8(b\log n + b)^2}{n^2}. \end{aligned}$$

Using the independence of the $X$'s within each group $g_l$ and the Bernstein inequality (Boucheron, Lugosi and Massart (2013, Theorem 2.10)), we obtain

$$P(W_j^{(l)} - \tilde{W}_j^{(l)} > \sqrt{2v_0 t} + c_0 t) \vee P(-(W_j^{(l)} - \tilde{W}_j^{(l)}) > \sqrt{2v_0 t} + c_0 t) \leqslant e^{-t}, \quad \forall\, t > 0,$$

where $v_0 = 8(b\log n + b)^2/n$, $c_0 = (2b\log n + 2b)/\sqrt{n}$ and $W_j$ denotes the $j$th component of $W$. Therefore, by Boucheron, Lugosi and Massart (2013, Theorem 2.3) we obtain, for $p \geqslant 1$,

$$\|W_j^{(l)} - \tilde{W}_j^{(l)}\|_p \leqslant C(\sqrt{pv_0} + pc_0) \leqslant \frac{Cpb\log n}{\sqrt{n}},$$

$$\|W_j - \tilde{W}_j\|_p \leqslant \sum_{l=1}^{L} \|W_j^{(l)} - \tilde{W}_j^{(l)}\|_p \leqslant \frac{CpLb\log n}{\sqrt{n}},$$

and

$$\|W - \tilde{W}\|_p \leqslant \sum_{j=1}^{d} \|W_j - \tilde{W}_j\|_p \leqslant \frac{CpdLb\log n}{\sqrt{n}}. \tag{7.29}$$

Using the triangle inequality, we have

$$\mathcal{W}_p(W, Z) \leqslant \mathcal{W}_p(W, \tilde{W}) + \mathcal{W}_p(\tilde{W}, \tilde{Z}) + \mathcal{W}_p(\tilde{Z}, Z), \tag{7.30}$$

where $\tilde{Z} \sim N(0, \mathrm{Var}(\tilde{W}))$. Note that

$$\begin{aligned}
&|\mathbb{E}[\tilde{W}_j \tilde{W}_k] - \mathbb{E}[W_j W_k]| \\
=&\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{i'\in A_i}\left\{\mathbb{E}[X_{ij}1_{\{|X_{ij}|\leqslant b\log n\}}X_{i'k}1_{\{|X_{i'k}|\leqslant b\log n\}}] - \mathbb{E}[X_{ij}X_{i'k}]\right\}\right| \\
=&\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{i'\in A_i}\left\{-\mathbb{E}[X_{ij}1_{\{|X_{ij}|\leqslant b\log n\}}X_{i'k}1_{\{|X_{i'k}|> b\log n\}}] - \mathbb{E}[X_{ij}1_{\{|X_{ij}|> b\log n\}}X_{i'k}]\right\}\right| \\
\leqslant& 2b\log n \max_i \sum_{i'\in A_i} \mathbb{E}[|X_{i'k}|1_{\{|X_{i'k}|\geqslant b\log n\}}] \\
&+ \max_i \sum_{i'\in A_i} \sqrt{\mathbb{E}[X_{ij}^2 1_{\{|X_{ij}|> b\log n\}}]\mathbb{E}[X_{i'k}^2 1_{\{|X_{i'k}|> b\log n\}}]} \\
\leqslant& \frac{C\theta_1 b^2\log^2 n}{n},
\end{aligned}$$

where we used Koike (2021, Lemma 5.4) in the last inequality. This implies, from the $p$-Wasserstein bound via Stein kernels by Ledoux, Nourdin and Peccati (2015, Proposition 3.4(ii)),

$$\mathcal{W}_p(\tilde{Z}, Z) \leqslant \frac{Cp^{1/2} d\theta_1 b^2 \log^2 n}{n} \text{ for all } p \geqslant 2. \tag{7.31}$$

Further, given a $d \times d$ matrix $A$, we write $\|A\|_{op,\infty} := \max_{j=1,\ldots,d}\sum_{k=1}^{d}|A_{jk}|$. Then, for $D := I_d - \mathrm{Var}(\tilde{W}) = \mathrm{Var}(W) - \mathrm{Var}(\tilde{W})$, we have

$$\|D\|_{op,\infty} = \max_{j=1,\ldots,d}\sum_{k=1}^{d}|\mathbb{E}[W_j W_k] - \mathbb{E}[\tilde{W}_j\tilde{W}_k]| \leqslant \frac{Cd\theta_1 b^2\log^2 n}{n}.$$

Thus, assuming $d^{1/2}\theta_1^{1/2} b\log n/\sqrt{n}$ to be sufficiently small as in the condition of Theorem 5.3, we have

$$\|\mathrm{Var}(\tilde{W})^{-1/2}\|_{op,\infty} = \|(I_d - D)^{-1/2}\|_{op,\infty} = \|\sum_{r=0}^{\infty}\frac{(2r)!}{4^r(r!)^2}D^r\|_{op,\infty} \leqslant \sum_{r=0}^{\infty}\frac{(2r)!}{4^r(r!)^2}\|D\|_{op,\infty}^r \leqslant C.$$

Therefore, we can apply Theorem 5.1 to $\mathrm{Var}(\tilde{W})^{-1/2}\tilde{X}_i$'s with $b_n' = Cb\log n$ and $b_n = Cd^{1/2}b\log n$. In addition, we have

$$\|\mathrm{Var}(\tilde{W})^{1/2}\|_{op} = \|\mathrm{Var}(\tilde{W})\|_{op}^{1/2} \leqslant \|\mathrm{Var}(\tilde{W})\|_{op,\infty}^{1/2} \leqslant (\|I_d\|_{op,\infty} + \|D\|_{op,\infty})^{1/2} \leqslant C,$$

where the first inequality follows by Theorem 5.6.9 in Hohn and Johnson (2013). These arguments imply that there exist positive absolute constants $c$ and $C$ such that, if

$$2 \leqslant p \leqslant \min\{\frac{\theta_1}{\theta_2}, \frac{c}{\theta_1^2 d b^2 \log^2 n}\} n, \tag{7.32}$$

then

$$\mathcal{W}_p(\tilde{W}, \tilde{Z}) \leqslant Cp\left(\frac{d(\theta_1\theta_2)^{1/2}b^2\log^2 n + \theta_1^2 d^{3/2} b^3 \log^4 n}{\sqrt{n}}\right).$$

Combining this with (7.29)–(7.31) and noting $1 \leqslant C\theta_1 b$, we have

$$\mathcal{W}_p(W, Z) \leqslant Cp\Delta_d \qquad \text{if (7.32) holds.}$$

The upper bounds (5.3) and (5.4) then follow from Theorem 2.2 and Theorem 4.2 respectively by a similar argument as in the proof of Theorem 5.2. □

# A Appendix

## A.1 Proof of Theorem 3.4

Theorem 3.4 is a straightforward consequence of the following $p$-Wasserstein bound and Theorem 2.2:

**Proposition A.1.** Under the assumptions of Theorem 3.4, for any $2 \leqslant p \leqslant 2\|F\|_{op}^{-2/3}$,

$$\mathcal{W}_p(W, Z) \leqslant CpK^4\|F\|_{op},$$

where $C$ is a positive absolute constant.

*Proof of Theorem* 3.4. We first note that $\|F\|_{op} \leqslant \|F\|_{H.S.} = 1/\sqrt{2}$. We apply Theorem 2.2 with $r_0 = \alpha_1 = 1$, $\Delta_1 = \|F\|_{op}$ and $p_0 = 2\|F\|_{op}^{-2/3}$. Then it remains to check $\log\|F\|_{op}^{-1} \leqslant \|F\|_{op}^{-2/3}$. This follows from the fact that $\log x \leqslant x^{2/3}$ for all $x > 0$. □

*Proof of Proposition* A.1. We construct an exchangeable pair $(W, W')$ in the same way as in the proof of Proposition 3.2. So we obtain the bound (7.12). We derive refined bounds for $H_1$ and $H_2$ using the assumption $q = 2$ and the boundedness of $X_i$. In the proof, a symmetric function $g : [n]^r \to \mathbb{R}$ is also regarded as an element of $(\mathbb{R}^n)^{\odot r}$. In particular, given a partition $\mathcal{J} \in \Pi_r$, we define the mixed injective norm $\|g\|_{\mathcal{J}}$ as in Section 3.2. Note that, if two partitions $\mathcal{J}_1, \mathcal{J}_2 \in \Pi_r$ are such that any element of $\mathcal{J}_1$ is contained in an element of $\mathcal{J}_2$, then $\|g\|_{\mathcal{J}_1} \leqslant \|g\|_{\mathcal{J}_2}$ by definition. Note also that $\|F\|_{op} \leqslant \|F\|_{H.S.} = 1/\sqrt{2} < 1$. Also, we will freely use tensor notations introduced in Section 6.

**Step 1. Bounding $H_1$.** We decompose $E$ as

$$E = \sum_{i=1}^n (X_i^2 - 1)Q_i(X)^2 + \left\{2\sum_{i=1}^n Q_i(X)^2 - 1\right\} =: E_1 + E_2. \tag{A.1}$$

Define an $n$-variate polynomial $\widetilde{Q}$ as

$$\widetilde{Q}(x_1, \dots, x_n) = \sum_{i=1}^n (x_i^2 - 1)Q_i(x_1, \dots, x_n)^2 = \sum_{i=1}^n (x_i^2 - 1)\left(\sum_{i'=1}^n f(i, i')x_{i'}\right)^2.$$

By Lemma 7.3,

$$\|E_1\|_p \leqslant C \sum_{r=1}^{4} K^r \sum_{\mathcal{J}\in\Pi_r} p^{|\mathcal{J}|/2} \|\mathbb{E}\nabla^r \widetilde{Q}(X)\|_{\mathcal{J}}. \tag{A.2}$$

We bound summands of $\sum_{r=1}^4$ in the following way.

Case 1: $r = 1$. Since $\mathbb{E}\nabla\widetilde{Q}(X) = 0$, we have

$$K \sum_{\mathcal{J}\in\Pi_1} p^{|\mathcal{J}|/2} \left\|\mathbb{E}\nabla\widetilde{Q}(X)\right\|_{\mathcal{J}} = 0.$$

Case 2: $r = 2$. For $j, k \in \{1, \dots, n\}$,

$$\mathbb{E}\partial_{j,k}\widetilde{Q}(X) = 2\sum_{i=1}^{n} f(i,j)^2 1_{\{j=k\}}.$$

Hence, using $\|f\| = 1/\sqrt{2}$ by standardization and $\sqrt{\mathcal{M}(f)} \leqslant \|F\|_{op}$ (we will use these two facts implicitly in the remainder of the proof),

$$\|\mathbb{E}\nabla^2\widetilde{Q}(X)\|_{\{1,2\}} = 2\sqrt{\sum_{j=1}^{n}\left(\sum_{i=1}^{n} f(i,j)^2\right)^2} \leqslant 2\sqrt{\mathcal{M}(f)}\|f\| \leqslant \sqrt{2}\|F\|_{op}$$

and

$$\|\mathbb{E}\nabla^2\widetilde{Q}(X)\|_{\{1\},\{2\}} = \|\mathbb{E}\nabla^2\widetilde{Q}(X)\|_{op} = 2\mathcal{M}(f) \leqslant 2\|F\|_{op}^2.$$

Therefore,

$$K^2 \sum_{\mathcal{J}\in\Pi_2} p^{|\mathcal{J}|/2} \left\|\mathbb{E}\nabla^2\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant CK^2(\sqrt{p}\|F\|_{op} + p\|F\|_{op}^2).$$

Case 3: $r = 3$. Since $\mathbb{E}\nabla^3\widetilde{Q}(X) = 0$,

$$K^3 \sum_{\mathcal{J}\in\Pi_3} p^{|\mathcal{J}|/2} \left\|\mathbb{E}\nabla^3\widetilde{Q}(X)\right\|_{\mathcal{J}} = 0.$$

Case 4: $r = 4$. Define a function $f_1 : [n]^4 \to \mathbb{R}$ as $f_1(j,k,l,m) = 1_{\{j=k\}} f(j,l) f(j,m)$ for $j,k,l,m \in [n]$. Then, for $j,k,l,m \in [n]$,

$$\mathbb{E}\partial_{j,k,l,m}\widetilde{Q}(X) = 4!\widetilde{f}_1(j,k,l,m),$$

where $\widetilde{f}_1$ is the symmetrization of $f_1$.

(i) Case $|\mathcal{J}| = 1$. In this case, we have

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant C\sqrt{\sum_{j,l,m=1}^{n} f(j,l)^2 f(j,m)^2} \leqslant C\|F\|_{op}\|F\|_{H.S.} \leqslant C\|F\|_{op}.$$

(ii) Case $|\mathcal{J}| = 2$. Observe that

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant \left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2\},\{3,4\}} \vee \left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2,3\},\{4\}}.$$

Since $f$ is symmetric, we have

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2\},\{3,4\}} \leqslant C \sup_{U,V\in(\mathbb{R}^n)^{\otimes 2}:|U|\vee|V|\leqslant 1}\left|\sum_{j,l,m=1}^n U_{jj}V_{lm}f(j,l)f(j,m)\right| + C \sup_{U,V\in(\mathbb{R}^n)^{\otimes 2}:|U|\vee|V|\leqslant 1}\left|\sum_{j,k,m=1}^n U_{jk}V_{jm}f(j,k)f(j,m)\right|$$

and

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2,3\},\{4\}} \leqslant C \sup_{U\in(\mathbb{R}^n)^{\otimes 3},v\in\mathbb{R}^n:|U|\vee|v|\leqslant 1}\left|\sum_{j,l,m=1}^n U_{jjl}v_m f(j,l)f(j,m)\right| + C \sup_{U\in(\mathbb{R}^n)^{\otimes 3},v\in\mathbb{R}^n:|U|\vee|v|\leqslant 1}\left|\sum_{j,k,l=1}^n U_{jkl}v_j f(j,l)f(j,k)\right|.$$

For any $U,V\in(\mathbb{R}^n)^{\otimes 2}$,

$$\left|\sum_{j,l,m=1}^n U_{jj}V_{lm}f(j,l)f(j,m)\right| = \left|\sum_{j=1}^n U_{jj}(FVF)_{jj}\right| \leqslant \|U\|_{H.S.}\|FVF\|_{H.S.} \leqslant \|F\|_{op}^2\|U\|_{H.S.}\|V\|_{H.S.}$$

and

$$\left|\sum_{j,k,m=1}^n U_{jk}V_{jm}f(j,k)f(j,m)\right| = \left|\sum_{j=1}^n (UF)_{jj}(VF)_{jj}\right| \leqslant \|UF\|_{H.S.}\|VF\|_{H.S.} \leqslant \|F\|_{op}^2\|U\|_{H.S.}\|V\|_{H.S.}.$$

Hence

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2\},\{3,4\}} \leqslant C\|F\|_{op}^2. \tag{A.3}$$

In the meantime, for any $U\in(\mathbb{R}^n)^{\otimes 3}$ and $v\in\mathbb{R}^n$,

$$\left|\sum_{j,l,m=1}^n U_{jjl}v_m f(j,l)f(j,m)\right| = \left|\sum_{j=1}^n\left(\sum_{l=1}^n U_{jjl}f(j,l)\right)(Fv)_j\right| \leqslant \sqrt{\sum_{j=1}^n\left(\sum_{l=1}^n U_{jjl}f(j,l)\right)^2}\,|Fv| \leqslant \|F\|_{op}^2|U||v|$$

and, with $U_j=(U_{jkl})_{1\leqslant k,l\leqslant n}$,

$$\left|\sum_{j,k,l=1}^n U_{jkl}v_j f(j,l)f(j,k)\right| = \left|\sum_{j=1}^n (FU_jF)_{jj}v_j\right| \leqslant \|F\|_{op}^2\sum_{j=1}^n\|U_j\|_{H.S.}|v_j| \leqslant \|F\|_{op}^2|U||v|.$$

Hence

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2,3\},\{4\}} \leqslant C\|F\|_{op}^2.$$

Consequently,

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant C\|F\|_{op}^2.$$

(iii) Case $|\mathcal{J}| = 3$. In this case, we have

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} = \left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\{1,2\},\{3\},\{4\}}.$$

Hence, by (A.3),

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant C\|F\|_{op}^2.$$

(iv) Case $|\mathcal{J}| = 4$. In this case we have $\mathcal{J} = \{\{1\},\{2\},\{3\},\{4\}\}$. Therefore, by (A.3),

$$\left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant C\|F\|_{op}^2.$$

All together, we obtain

$$K^4 \sum_{\mathcal{J}\in\Pi_4} p^{|\mathcal{J}|/2} \left\|\mathbb{E}\nabla^4\widetilde{Q}(X)\right\|_{\mathcal{J}} \leqslant CK^4(\sqrt{p}\|F\|_{op} + p^2\|F\|_{op}^2).$$

Combining these bounds with (A.2) gives

$$\|E_1\|_p \leqslant CK^4(\sqrt{p}\|F\|_{op} + p^2\|F\|_{op}^2). \tag{A.4}$$

In the meantime, by a similar argument to the proof of Proposition 3.2 (cf. (7.13) and the bound on $\|H_{12}\|_p$ therein),

$$\|E_2\|_p \leqslant CK^2 \max_{\mathcal{J}\in\Pi_2} p^{|\mathcal{J}|/2}\|f\widetilde{\otimes}_1 f\|_{\mathcal{J}}.$$

Observe that

$$\left(f\widetilde{\otimes}_1 f(j,k)\right)_{1\leqslant j,k\leqslant n} = \left(\sum_{i=1}^n f(i,j)f(i,k)\right)_{1\leqslant j,k\leqslant n} = F^2.$$

Hence we have

$$\|f\widetilde{\otimes}_1 f\|_{\{1,2\}} = \|F^2\|_{H.S.} \leqslant \|F\|_{op}\|F\|_{H.S.} = \|F\|_{op}/\sqrt{2}$$

and

$$\|f\widetilde{\otimes}_1 f\|_{\{1\},\{2\}} = \|F\|_{op}^2.$$

Consequently,

$$\|E_2\|_p \leqslant CK^4 \max\{\sqrt{p}\|F\|_{op}, p\|F\|_{op}^2\} \leqslant CK^4\sqrt{p}\|F\|_{op}. \tag{A.5}$$

Combining (A.1), (A.4) and (A.5) gives

$$H_1 \leqslant C\sqrt{p}(\|E_1\|_p + \|E_2\|_p) \leqslant CK^4\left(p\|F\|_{op} + p^{5/2}\|F\|_{op}^2\right) \leqslant CpK^4\|F\|_{op}, \tag{A.6}$$

where the last inequality follows by the condition $p\|F\|_{op}^{2/3} \leqslant 2$.

**Step 2. Bounding $H_2$.** Since $|X_i' - X_i| \leqslant 2K$ a.s.,

$$\frac{n}{2}\mathbb{E}[D^4|X] \leqslant 8(2K)^4\sum_{i=1}^n Q_i(X)^4. \tag{A.7}$$

By Lemma 7.3,

$$\sum_{i=1}^n \mathbb{E}Q_i(X)^4 \leqslant CK^4 \sum_{i=1}^n \left(\sum_{j=1}^n f(i,j)^2\right)^2 \leqslant CK^4 \mathcal{M}(f)\|f\|^2 \leqslant CK^4\|F\|_{op}^2 \tag{A.8}$$

and

$$\left\|\sum_{i=1}^n Q_i(X)^4 - \sum_{i=1}^n \mathbb{E}Q_i(X)^4\right\|_p \leqslant C\sum_{r=1}^4 K^r \sum_{\mathcal{J}\in\Pi_r} p^{|\mathcal{J}|/2}\left\|\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X)\right\|_{\mathcal{J}}. \tag{A.9}$$

We bound summands of $\sum_{r=1}^4$ in the following way.

Case 1: $r=1$. For $j\in\{1,\dots,n\}$,

$$\mathbb{E}\partial_j Q_i^4(X) = 4f(i,j)\mathbb{E}\left(\sum_{i'=1}^n f(i,i')X_{i'}\right)^3.$$

Therefore, with

$$v := \left(\mathbb{E}\left(\sum_{i'=1}^n f(1,i')X_{i'}\right)^3, \dots, \mathbb{E}\left(\sum_{i'=1}^n f(n,i')X_{i'}\right)^3\right)^T,$$

we have

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla Q_i^4(X)\right\|_{\{1\}} = 4|Fv| \leqslant 4\|F\|_{op}|v|.$$

By Lemma 7.3,

$$|v|^2 = \sum_{i=1}^n \left|\mathbb{E}\left(\sum_{i'=1}^n f(i,i')X_{i'}\right)^3\right|^2 \leqslant CK^6\sum_{i=1}^n\left(\sum_{i'=1}^n f(i,i')^2\right)^3 \leqslant CK^6\|F\|_{op}^4.$$

Hence

$$K\sum_{\mathcal{J}\in\Pi_1} p^{|\mathcal{J}|/2}\left\|\sum_{i=1}^n \mathbb{E}\nabla Q_i(X)\right\|_{\mathcal{J}} \leqslant CK^4\sqrt{p}\|F\|_{op}^3.$$

Case 2: $r=2$. For $j,k\in\{1,\dots,n\}$,

$$\mathbb{E}\partial_{jk}Q_i^4(X) = 12f(i,j)f(i,k)\sum_{i'=1}^n f(i,i')^2.$$

Hence

$$\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X) = 12F\mathrm{diag}\left(\sum_{i'=1}^n f(1,i')^2,\dots,\sum_{i'=1}^n f(n,i')^2\right)F.$$

Therefore,

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X)\right\|_{\{1\},\{2\}} = \left\|\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X)\right\|_{op} \leqslant 12\|F\|_{op}^2\mathcal{M}(f) \leqslant 12\|F\|_{op}^4$$

and

$$\begin{aligned}\left\|\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X)\right\|_{\{1,2\}} = \left\|\sum_{i=1}^n \mathbb{E}\nabla^r Q_i^4(X)\right\|_{H.S.} &\leqslant 12\|F\|_{op}^2\sqrt{\sum_{i=1}^n\left(\sum_{i'=1}^n f(i,i')^2\right)^2} \\ &\leqslant 6\sqrt{2}\|F\|_{op}^3.\end{aligned}$$

Hence

$$K^2 \sum_{\mathcal{J}\in\Pi_2} p^{|\mathcal{J}|/2} \left\| \sum_{i=1}^n \mathbb{E}\nabla^2 Q_i^4(X) \right\|_{\mathcal{J}} \leqslant CK^2(\sqrt{p}\|F\|_{op}^3 + p\|F\|_{op}^4).$$

Case 3: $r=3$. Since $\mathbb{E}\partial_{jkl}Q_i^4(X) = 0$ for all $j,k,l \in \{1,\dots,n\}$,

$$K^3 \sum_{\mathcal{J}\in\Pi_3} p^{|\mathcal{J}|/2} \left\| \sum_{i=1}^n \mathbb{E}\nabla^3 Q_i^4(X) \right\|_{\mathcal{J}} = 0.$$

Case 4: $r=4$. For $j,k,l,m \in \{1,\dots,n\}$,

$$\mathbb{E}\partial_{jklm}Q_i^4(X) = 24 f(i,j)f(i,k)f(i,l)f(i,m).$$

(i) Case $|\mathcal{J}|=1$. In this case, we have

$$\begin{aligned}
\left\| \sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X) \right\|_{\mathcal{J}} &= 24\sqrt{\sum_{j,k,l,m=1}^n \left( \sum_{i=1}^n f(i,j)f(i,k)f(i,l)f(i,m) \right)^2} \\
&= 24\sqrt{\sum_{i,i'=1}^n |(F^2)_{ii'}|^4} \leqslant 24\|F^2\|_{op}\|F^2\|_{H.S.} \\
&\leqslant 24\|F\|_{op}^3\|F\|_{H.S.} = 12\sqrt{2}\|F\|_{op}^3.
\end{aligned}$$

(ii) Case $|\mathcal{J}|=2$. Observe that

$$\left\| \sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X) \right\|_{\mathcal{J}} \leqslant \left\| \sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X) \right\|_{\{1,2\},\{3,4\}} \vee \left\| \sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X) \right\|_{\{1,2,3\},\{4\}}.$$

For any $U,V \in (\mathbb{R}^n)^{\otimes 2}$,

$$\begin{aligned}
\sum_{i=1}^n \langle \mathbb{E}\nabla^4 Q_i^4(X), U\otimes V\rangle &= 24 \sum_{i,j,k,l,m=1}^n f(i,j)f(i,k)f(i,l)f(i,m)U_{jk}V_{lm} \\
&= 24\sum_{i=1}^n (FUF)_{ii}(FVF)_{ii} \leqslant 24\|FUF\|_{H.S.}\|FVF\|_{H.S.} \\
&\leqslant 24\|F\|_{op}^4\|U\|_{H.S.}\|V\|_{H.S.}.
\end{aligned}$$

Hence

$$\left\| \sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X) \right\|_{\{1,2\},\{3,4\}} \leqslant 24\|F\|_{op}^4. \tag{A.10}$$

In the meantime, for any $U \in (\mathbb{R}^n)^{\otimes 3}$ and $v\in\mathbb{R}^n$,

$$\begin{aligned}
\sum_{i=1}^n \langle \mathbb{E}\nabla^4 Q_i^4(X), U\otimes v\rangle &= 24 \sum_{i,j,k,l,m=1}^n f(i,j)f(i,k)f(i,l)f(i,m)U_{jkl}v_m \\
&= 24\sum_{i,j=1}^n f(i,j)(FU_jF)_{ii}(Fv)_i,
\end{aligned}$$

where $U_j = (U_{jkl})_{1\leqslant k,l\leqslant n}$. Thus, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\sum_{i=1}^n \langle \mathbb{E}\nabla^4 Q_i^4(X), U\otimes v\rangle &\leqslant 24\sqrt{\sum_{i,j=1}^n f(i,j)^2|(Fv)_i|^2\sum_{i,j=1}^n (FU_jF)_{ii}^2} \\ &\leqslant 24\|F\|_{op}^2\sqrt{\sum_{i=1}^n|(Fv)_i|^2\sum_{j=1}^n\|U_j\|_{H.S.}^2} \\ &\leqslant 24\|F\|_{op}^3|v||U|.\end{aligned}$$

Hence

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\{1,2,3\},\{4\}} \leqslant 24\|F\|_{op}^3.$$

Consequently,

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\mathcal{J}} \leqslant 24\|F\|_{op}^3.$$

(iii) Case $|\mathcal{J}| = 3$. In this case, we have

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\mathcal{J}} = \left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\{1,2\},\{3\},\{4\}}.$$

Therefore, by (A.10),

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\mathcal{J}} \leqslant 24\|F\|_{op}^4.$$

(iv) Case $|\mathcal{J}| = 4$. In this case we have $\mathcal{J} = \{\{1\},\{2\},\{3\},\{4\}\}$. Therefore, by (A.10),

$$\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\mathcal{J}} \leqslant 24\|F\|_{op}^4.$$

All together, we obtain

$$K^4\sum_{\mathcal{J}\in\Pi_4} p^{|\mathcal{J}|/2}\left\|\sum_{i=1}^n \mathbb{E}\nabla^4 Q_i^4(X)\right\|_{\mathcal{J}} \leqslant CK^4(p\|F\|_{op}^3 + p^2\|F\|_{op}^4).$$

Combining these bounds with (A.9) and the condition $p \leqslant 2\|F\|_{op}^{-1}$, we obtain

$$\left\|\sum_{i=1}^n Q_i(X)^4 - \sum_{i=1}^n \mathbb{E}Q_i(X)^4\right\| \leqslant CK^4\|F\|_{op}^2. \tag{A.11}$$

By (A.7), (A.8) and (A.11), we conclude

$$H_2 \leqslant CK^4 p\|F\|_{op}. \tag{A.12}$$

Combining (7.12), (A.6) and (A.12), we complete the proof. □

### A.2 Removing the extra assumptions in derivation of (6.13)

In the literature, the bound (6.13) was formally established only when $W$ has a bounded $C^\infty$ density $h$ with respect to $N(0, I_d)$ such that $h \geqslant \eta$ for some constant $\eta > 0$ and $|\nabla h|$ is bounded. In this appendix, we show this assumption can be replaced with $\mathbb{E}|W|^p < \infty$. Our argument is largely the same as in Section 8 of Bonis (2020). Below we assume $W$ and $Z$ are independent without loss of generality.

**Step 1**. In this step, we prove (6.13) when $W$ has a compactly supported $C^\infty$ density $f$. Let $U$ be a uniform random variable on $[0, 1]$ independent of $W$ and $Z$. Also, let $Z' \sim N(0, I_d)$ be independent of everything else. Take $\eta \in (0, 1)$ arbitrarily, and define $I^\eta := 1_{\{U \leqslant \eta\}}$ and $W^\eta := I^\eta Z' + (1 - I^\eta)W$. Then, for any bounded measurable function $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}g(W^\eta) = \eta\mathbb{E}g(Z') + (1 - \eta)\mathbb{E}g(W) = \eta \int_{\mathbb{R}^d} g(x)\phi(x)dx + (1 - \eta) \int_{\mathbb{R}^d} g(x)f(x)dx.$$

Hence $\eta + (1 - \eta)f/\phi$ is a density of $W^\eta$ with respect to $N(0, I_d)$. In this case we already have

$$\mathcal{W}_p(W^\eta, Z) \leqslant \int_0^\infty \|\rho_t^\eta(F_t^\eta)\|_p dt, \tag{A.13}$$

where $F_t^\eta := e^{-t}W^\eta + \sqrt{1 - e^{-2t}}Z$ and $\rho_t^\eta$ is the score of $F_t^\eta$ with respect to $N(0, I_d)$. By the triangle inequality for the $p$-Wasserstein distance, we have

$$\begin{aligned}|\mathcal{W}_p(W, Z) - \mathcal{W}_p(W^\eta, Z)| &\leqslant \mathcal{W}_p(W, W^\eta) \leqslant \|W - W^\eta\|_p \\ &= (\mathbb{E}I^\eta|W - Z'|^p)^{1/p} = \eta^{1/p}\|W - Z'\|_p.\end{aligned}$$

Hence $|\mathcal{W}_p(W, Z) - \mathcal{W}_p(W^\eta, Z)| \to 0$ as $\eta \downarrow 0$.

Meanwhile, by Lemma 2 in Bonis (2020),

$$\rho_t^\eta(F_t^\eta) = \mathbb{E}\left[e^{-t}W^\eta - \frac{e^{-2t}}{\sqrt{1 - e^{-2t}}}Z|F_t^\eta\right]. \tag{A.14}$$

In particular,

$$\|\rho_t^\eta(F_t^\eta)\|_p \leqslant e^{-t}(\|Z'\|_p + \|W\|_p) + \frac{e^{-2t}}{\sqrt{1 - e^{-2t}}}\|Z\|_p.$$

Hence, by the reverse Fatou lemma,

$$\limsup_{\eta\downarrow 0} \int_0^\infty \|\rho_t^\eta(F_t^\eta)\|_p dt \leqslant \int_0^\infty \limsup_{\eta\downarrow 0} \|\rho_t^\eta(F_t^\eta)\|_p dt.$$

Therefore, we complete the proof once we show that $\|\rho_t^\eta(F_t^\eta)\|_p \to \|\rho_t(F_t)\|_p$ as $\eta \downarrow 0$ for any fixed $t > 0$. The latter follows once we verify the following two statements:

(i) $\rho_t^\eta(F_t^\eta) \to \rho_t(F_t)$ as $\eta \downarrow 0$ a.s.

(ii) $\{|\rho_t^\eta(F_t^\eta)|^p : \eta \in (0, 1)\}$ is uniformly integrable.

*Proof of (i).* For any bounded measurable function $g : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}g(e^{-t}W^\eta + \sqrt{1 - e^{-2t}}Z) &= \eta\mathbb{E}g(e^{-t}Z' + \sqrt{1 - e^{-2t}}Z) + (1 - \eta)\mathbb{E}g(F_t) \\ &= \eta \int_{\mathbb{R}^d} g(x)\phi(x)dx + (1 - \eta) \int_{\mathbb{R}^d} g(x)f_t(x)\phi(x)dx,\end{aligned}$$

where $f_t$ is the density of $F_t$ with respect to $N(0, I_d)$. Hence $\eta + (1-\eta) f_t$ is the smooth density of $F_t^\eta$ with respect to $N(0, I_d)$, and thus

$$\rho_t^\eta(F_t^\eta) = (1-\eta)\nabla f_t(F_t^\eta)/(\eta + (1-\eta) f_t(F_t^\eta)).$$

Since $f_t$ is smooth and $F_t^\eta \to F_t$ as $\eta \downarrow 0$ a.s., we have $\rho_t^\eta(F_t^\eta) \to \nabla f_t(F_t)/f_t(F_t) = \rho_t(F_t)$ as $\eta \downarrow 0$ a.s. $\square$

*Proof of (ii).* Let

$$G_t := e^{-t}(|W| + |Z'|) + \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}|Z|.$$

Then we have $|\rho_t^\eta(F_t^\eta)|^p \leqslant \mathbb{E}[G_t^p | F_t^\eta]$ for any $\eta \in (0,1)$ by (A.14) and Jensen's inequality. Hence, for any $K > 0$,

$$\mathbb{E}[|\rho_t^\eta(F_t^\eta)|^p; |\rho_t(F_t^\eta)|^p > K] \leqslant \mathbb{E}[\mathbb{E}[G_t^p|F_t^\eta]; \mathbb{E}[G_t^p|F_t^\eta] > K].$$

$\square$

Since $\mathbb{E}G_t^p < \infty$, $\{\mathbb{E}[G_t^p|F_t^\eta] : \eta \in (0,1)\}$ is uniformly integrable by Theorem 13.4 in Williams (1991). Hence $\{|\rho_t^\eta(F_t^\eta)|^p : \eta \in (0,1)\}$ is uniformly integrable as well.

**Step 2**. In this step, we prove (6.13) when $W$ is bounded. Let $N$ be a random variable independent of $W$ and $Z$ and such that $N$ has a $C^\infty$ density $\psi$ and takes values in the unit ball in $\mathbb{R}^d$. Take $\varepsilon > 0$ arbitrarily and define $W^\varepsilon := W + \varepsilon N$. Then, for any bounded measurable function $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}g(W^\varepsilon) = \int_{\mathbb{R}^d} \mathbb{E}[g(W + \varepsilon x)]\psi(x)dx = \varepsilon^{-d}\int_{\mathbb{R}^d} g(y)\mathbb{E}[\psi((y-W)/\varepsilon)]dy.$$

Hence $f(y) = \varepsilon^{-d}\mathbb{E}[\psi((y-W)/\varepsilon)]$ is a density of $W^\varepsilon$. Since $\psi$ is $C^\infty$ and compactly supported, $f$ is $C^\infty$. Also, since $W$ is bounded, $f$ is compactly supported. Thus, by Step 1,

$$\mathcal{W}_p(W^\varepsilon, Z) \leqslant \int_0^\infty \|\rho_t^\varepsilon(F_t^\varepsilon)\|_p dt, \tag{A.15}$$

where $F_t^\varepsilon := e^{-t}W^\varepsilon + \sqrt{1-e^{-2t}}Z$ and $\rho_t^\varepsilon$ is the score of $F_t^\varepsilon$ with respect to $N(0, I_d)$. By the triangle inequality for the $p$-Wasserstein distance, we have

$$|\mathcal{W}_p(W, Z) - \mathcal{W}_p(W^\varepsilon, Z)| \leqslant \mathcal{W}_p(W, W^\varepsilon) \leqslant \|W - W^\varepsilon\|_p = \varepsilon\|N\|_p.$$

Meanwhile, by Lemma 2 in Bonis (2020),

$$\begin{aligned}
\rho_t^\varepsilon(F_t^\varepsilon) &= \mathbb{E}\left[e^{-t}W^\varepsilon - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}Z | F_t^\varepsilon\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{-t}W^\varepsilon - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}Z | F_t, N\right] | F_t^\varepsilon\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{-t}W - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}Z | F_t, N\right] + \mathbb{E}\left[e^{-t}\varepsilon N | F_t, N\right] | F_t^\varepsilon\right] \\
&= \mathbb{E}[\rho_t(F_t)|F_t^\varepsilon] + \varepsilon\mathbb{E}\left[e^{-t}N|F_t^\varepsilon\right],
\end{aligned}$$

where we used the independence between $(W, Z)$ and $N$ in the last line. Hence

$$\int_0^\infty \|\rho_t^\varepsilon(F_t^\varepsilon)\|_p dt \leqslant \int_0^\infty \|\rho_t(F_t)\|_p dt + \varepsilon\|N\|_p.$$

Consequently, letting $\varepsilon \downarrow 0$ in (A.15), we obtain (6.13).

**Step 3**. In this step, we prove (6.13) when $\mathbb{E}|W|^p < \infty$. Take $R > 0$ arbitrarily and define $W^R := W1_{\{|W| \leqslant R\}}$. Since $W^R$ is bounded, we have by Step 2

$$\mathcal{W}_p(W^R, Z) \leqslant \int_0^\infty \|\rho_t^R(F_t^R)\|_p dt, \tag{A.16}$$

where $F_t^R := e^{-t}W^R + \sqrt{1-e^{-2t}}Z$ and $\rho_t^R$ is the score of $F_t^R$ with respect to $N(0, I_d)$. By the triangle inequality for the $p$-Wasserstein distance, we have

$$|\mathcal{W}_p(W, Z) - \mathcal{W}_p(W^R, Z)| \leqslant \mathcal{W}_p(W, W^R) \leqslant \|W - W^R\|_p = (\mathbb{E}[|W|^p 1_{\{|W>R|\}}])^{1/p}.$$

Since $\mathbb{E}|W|^p < \infty$, we obtain $|\mathcal{W}_p(W, Z) - \mathcal{W}_p(W^R, Z)| \to 0$ as $R \to \infty$ by the dominated convergence theorem. Meanwhile, by Lemma 2 in Bonis (2020),

$$\rho_t^R(F_t^R) = \mathbb{E}\left[e^{-t}W^R - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z | F_t^R\right] \tag{A.17}$$

and

$$\rho_t(F_t) = \mathbb{E}\left[e^{-t}W - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z | F_t\right]. \tag{A.18}$$

In particular,

$$\|\rho_t^R(F_t^R)\|_p \leqslant e^{-t}\|W\|_p + \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\|Z\|_p.$$

Hence, by the reverse Fatou lemma,

$$\limsup_{R\to\infty} \int_0^\infty \|\rho_t^R(F_t^R)\|_p dt \leqslant \int_0^\infty \limsup_{R\to\infty} \|\rho_t^R(F_t^R)\|_p dt.$$

Therefore, we complete the proof once we show that $\|\rho_t^R(F_t^R)\|_p \to \|\rho_t(F_t)\|_p$ as $R \to \infty$ for any fixed $t > 0$. The latter follows once we verify the following two statements:

(i) $\rho_t^R(F_t^R) \to \rho_t(F_t)$ as $R \to \infty$ a.s.

(ii) $\{|\rho_t^R(F_t^R)|^p : R > 0\}$ is uniformly integrable.

*Proof of (i).* For any $u \in \mathbb{R}^d$,

$$|\mathbb{E}[W^R e^{\sqrt{-1}u\cdot F_t^R}]| = |\mathbb{E}[W^R e^{\sqrt{-1}u\cdot e^{-t}W^R}]\mathbb{E}[e^{\sqrt{-1}u\cdot\sqrt{1-e^{-2t}}Z}]| \leqslant \mathbb{E}|W|e^{-(1-e^{-2t})u^2/2} \tag{A.19}$$

and

$$|\mathbb{E}[Z e^{\sqrt{-1}u\cdot F_t^R}]| = |\mathbb{E}[e^{\sqrt{-1}u\cdot e^{-t}W^R}]\mathbb{E}[Z e^{\sqrt{-1}u\cdot\sqrt{1-e^{-2t}}Z}]| \leqslant |u|e^{-(1-e^{-2t})u^2/2}. \tag{A.20}$$

Hence, we can define a function $g_R : \mathbb{R}^d \to \mathbb{C}$ as

$$g_R(x) = \frac{1}{f_R(x)(2\pi)^d} \int_{\mathbb{R}^d} e^{-\sqrt{-1}u\cdot x}\mathbb{E}\left[\left(e^{-t}W^R - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}Z\right) e^{\sqrt{-1}u\cdot F_t^R}\right] du,\ x \in \mathbb{R}^d,$$

where $f_R(x) = (1-e^{-2t})^{-d/2}\mathbb{E}[\phi((x - e^{-t}W^R)/\sqrt{1-e^{-2t}})]$ is the density of $F_t^R$. Similarly, we can define a function $g : \mathbb{R}^d \to \mathbb{C}$ as

$$g(x) = \frac{1}{f(x)(2\pi)^d} \int_{\mathbb{R}^d} e^{-\sqrt{-1}u\cdot x}\mathbb{E}\left[\left(e^{-t}W - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}Z\right) e^{\sqrt{-1}u\cdot F_t}\right] du,\ x \in \mathbb{R}^d,$$

where $f(x) = (1-e^{-2t})^{-d/2}\mathbb{E}[\phi((x - e^{-t}W)/\sqrt{1-e^{-2t}})]$ is the density of $F_t$. By Theorem 2 in Yeh (1974) and (A.17)–(A.18), we have $g_R(F_t^R) = \rho_t^R(F_t^R)$ a.s. and $g(F_t) = \rho_t(F_t)$ a.s. Moreover, by (A.19), (A.20) and the dominated convergence theorem, $g_R(x) \to g(x)$ as $R \to \infty$ for any $x \in \mathbb{R}^d$. Hence $\rho_t^R(F_t^R) \to \rho_t(F_t)$ as $R \to \infty$ a.s. □

*Proof of (ii).* Let

$$G_t := e^{-t}|W| + \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}|Z|.$$

Then we have $|\rho_t^R(F_t^R)|^p \leqslant \mathbb{E}[G_t^p|F_t^R]$ for any $R > 0$ by (A.17) and Jensen's inequality. Hence, for any $K > 0$,

$$\mathbb{E}[|\rho_t^R(F_t^R)|^p; |\rho_t(F_t^R)|^p > K] \leqslant \mathbb{E}[\mathbb{E}[G_t^p|F_t^R]; \mathbb{E}[G_t^p|F_t^R] > K].$$

Since $\mathbb{E}G_t^p < \infty$, $\{\mathbb{E}[G_t^p|F_t^R] : R > 0\}$ is uniformly integrable by Theorem 13.4 in Williams (1991). Hence $\{|\rho_t^R(F_t^R)|^p : R > 0\}$ is uniformly integrable as well. □

## References


R. Adamczak and P. Wolff (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Relat. Fields* **162**, 531–586. MR3383337

P. Baldi and Y. Rinott (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17**, 1646–1650. MR1048950

A.D. Barbour, M. Karoński and A. Ruciński (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**, no. 2, 125–145. MR1047781

R. Bogucki (2015). Suprema of canonical Weibull processes. *Statist. Probab. Lett.* **107**, 253–263. MR3412784

T. Bonis (2020). Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem. *Probab. Theory Related Fields* **178**, 827–860. MR4168389

S. Boucheron, G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Clarendon Press, Oxford. MR3185193

S. Chatterjee (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138**, 305–321. MR2288072

L.H.Y. Chen and X. Fang (2015). On the error bound in a conbinatorial central limit theorem. *Bernoulli* **21**, 335–359. MR3322321

L.H.Y. Chen, X. Fang and Q.M. Shao (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41**, 262–293 MR3059199

L.H.Y. Chen and Q.M. Shao (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, no. 3A, 1985–2028. MR2073183

H. Cramér (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736**, 5–23.

P. de Jong (1990). A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.* **34**, 275–289. MR1073110

C. Döbler (2023). Normal approximation via non-linear exchangeable pairs. *Alea.* **20**, 167–224. MR4554229

C. Döbler and G. Peccati (2017). Quantitative de Jong theorems in any dimension. *Electron. J. Probab.* **22**, 1–35. MR3613695

X. Fang and Y. Koike (2022). New error bounds in multivariate normal approximations via exchangeable pairs with applications to Wishart matrices and fourth moment theorems. *Ann. Appl. Probab.* **32**, 602–631. MR4386537

X. Fang, S.H. Liu and Q.M. Shao (2021). Cramér-type moderate deviation for quadratic forms with a fast rate. *Preprint.* Available at https://arxiv.org/abs/2111.00679 MR4580924

X. Fang and A. Röllin (2015). Rates of convergence for multivariate normal approximation with applications to dense graphs and doubly indexed permutation statistics. *Bernoulli* **21**, 2157–2189. MR3378463

A.N. Frolov (2022). On large deviations for combinatorial sums. *J. Statist. Plann. Inference* **217**, 24–32. MR4292372

W. Hoeffding and H. Robbins (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780. MR0026771

R. A. Hohn and C. R. Johnson (2013). Matrix Analysis (2nd edition). *Cambridge University Press.* MR2978290

S. Janson (1997). *Gaussian Hilbert Spaces.* Cambridge University Press. MR1474726

O. Johnson and Y. Suhov (2001). Entropy and random vectors. *J. Stat. Phys.* **104**, 145–165. MR1851387

Y. Koike (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4**, 643–696. MR4273258

Y. Koike (2023). High-dimensional central limit theorems for homogeneous sums. *J. Theoret. Probab.* **36**, 1–45. MR4585109

J. Komlós, P. Major and G. Tusnády (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32**, 111–131. MR0375412

A. K. Kuchibhotla and A. Chakrabortty (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Inf. Inference* **11**, 1389–1456. MR4526326

S. Kwapień and W. A. Woyczyński (1992). *Random series and stochastic integrals: Single and multiple.* Birkhäuser. MR1167198

R. Latała (2006). Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34**, 2315–2331. MR2294983

M. Ledoux, I. Nourdin and G. Peccati (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25**, 256–306. MR3320893

J. Lehec (2011). Moments of the Gaussian chaos. In: *Séminaire de Probabilités XLIII*, 327–340, Springer. MR2790379

Yu. V. Linnik (1961). On the probability of large deviations for the sums of independent variables. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory*, 289–306, University of California Press. MR0137142

S.H. Liu and Z.S. Zhang (2021). Cramér-type moderate deviations under local dependence. *Preprint.* Available at https://arxiv.org/abs/2112.10946

D.M. Mason and H.H. Zhou (2012). Quantile coupling inequalities and their applications. *Probab. Surv.* **9**, 439–479. MR3007210

I. Nourdin and G. Peccati (2012). *Normal approximations with Malliavin calculus: From Stein's method to universality.* Cambridge University Press. MR2962301

I. Nourdin, G. Peccati and G. Reinert (2010). Invariance principles for homogeneous sums: Universality of Gaussian Wiener chaos. *Ann. Probab.* **38**, 1947–1985. MR2722791

I. Nourdin, G. Peccati and Y. Swan (2014). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, 3170–3207. MR3158721

F. Otto and C. Villani (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173**, 361–400. MR1760620

V.V. Petrov (1975). *Sums of independent random variables.* Springer-Verlag. MR0388499

G. Reinert and A. Röllin (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37**, 2150–2173. MR2573554

L. Saulis and V.A. Statulevičius (1991). *Limit theorems for large deviations.* Kluwer Academic Publishers Group, Dordrecht. MR1171883

M. Schulte and C. Thäle (2016). Cumulants on Wiener chaos: Moderate deviations and the fourth moment theorem. *J. Funct. Anal.* **270**, 2223–2248. MR3460239

C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **2**, 583–602, Univ. California Press. Berkeley, Calif. MR0402873

R. Vershynin (2018). *High-dimensional probability. An introduction with applications in data science.* Cambridge Series in Statistical and Probabilistic Mathematics, **47**. Cambridge University Press, Cambridge. MR3837109

B. von Bahr (1967). Multi-dimensional integral limit theorems for large deviations. *Ark. Mat.*, **7**, 89–99. MR0224129

D. Williams (1991). *Probability with martingales.* Cambridge University Press. MR1155402

J. J.-C. Yeh (1974). Inversion of conditional expectations. *Pacific J. Math.* **52**, 631–640. MR0365644

A. Zhai. (2018). A high-dimensional CLT in $\mathcal{W}_2$ distance with near optimal convergence rate. *Probab. Theory Related Fields* **170**, no. 3-4, 821–845. MR3773801

Z.S. Zhang (2022). Berry–Esseen bounds for generalized $U$-statistics. *Electro. J. Probab.* **27**, 1–36. MR4492985

**Acknowledgments.** We thank the anonymous referee for his/her valuable suggestions which led to many improvements.