The Annals of Applied Probability
2021, Vol. 31, No. 4, 1660–1686
https://doi.org/10.1214/20-AAP1629

HIGH-DIMENSIONAL CENTRAL LIMIT THEOREMS BY STEIN'S METHOD

BY XIAO FANG[1] AND YUTA KOIKE[2]

[1]*Department of Statistics, Chinese University of Hong Kong, xfang@sta.cuhk.edu.hk*
[2]*Graduate School of Mathematical Sciences, University of Tokyo, kyuta@ms.u-tokyo.ac.jp*

We obtain explicit error bounds for the d-dimensional normal approximation on hyperrectangles for a random vector that has a Stein kernel, or admits an exchangeable pair coupling, or is a nonlinear statistic of independent random variables or a sum of n locally dependent random vectors. We assume the approximating normal distribution has a nonsingular covariance matrix. The error bounds vanish even when the dimension d is much larger than the sample size n. We prove our main results using the approach of Götze (1991) in Stein's method, together with modifications of an estimate of Anderson, Hall and Titterington (1998) and a smoothing inequality of Bhattacharya and Rao (1976). For sums of n independent and identically distributed isotropic random vectors having a log-concave density, we obtain an error bound that is optimal up to a $\log n$ factor. We also discuss an application to multiple Wiener–Itô integrals.

1. Introduction and main results. Motivated by modern statistical applications in large-scale data, there has been a recent wave of interest in proving high-dimensional central limit theorems. Starting from the pioneering work by Chernozhukov, Chetverikov and Kato (2013), who established a Gaussian approximation for maxima of sums of centered independent random vectors, many articles have been devoted to the development of this subject: For example, see Chernozhukov, Chetverikov and Kato (2017a), Chernozhukov et al. (2019) for generalization to normal approximation on hyperrectangles and improvements of the error bound, Chen (2018), Chen and Kato (2019), Song, Chen and Kato (2019) for extensions to U-statistics, Chernozhukov, Chetverikov and Kato (2019), Zhang and Cheng (2018), Zhang and Wu (2017) for sums of dependent random vectors and Belloni et al. (2018) for a general survey and statistical applications. In particular, for $W = n^{-1/2}\sum_{i=1}^n X_i$ where $\{X_1, \dots, X_n\}$ are centered independent random vectors in $\mathbb{R}^d$ and satisfy certain regularity conditions, Chernozhukov et al. (2019) proved that

$$\sup_{h=1_A : A\in\mathcal{R}} \big|Eh(W) - Eh(Z)\big| \le C_0\bigg(\frac{\log^5(dn)}{n}\bigg)^{1/4}, \tag{1.1}$$

where $\mathcal{R} := \{\Pi_{j=1}^d (a_j, b_j), -\infty \le a_j \le b_j \le \infty\}$, Z is a centered Gaussian vector with the same covariance matrix as W and C_0 is a positive constant that is independent of d and n.

The distance between two probability measures on $\mathbb{R}^d$ considered in (1.1) is stronger than the multivariate Kolmogorov distance. The error bound vanishes if $\log d = o(n^{1/5})$, which allows d to be much larger than n. The result in (1.1) is useful in many statistical applications in high-dimensional inference such as construction of simultaneous confidence intervals and strong control of the family-wise error rate in multiple testing; see Belloni et al. (2018) for details. In the literature, people have also considered bounding other (stronger) distances in

Received January 2020; revised September 2020.
MSC2020 subject classifications. 60F05, 62E17.
Key words and phrases. Central limit theorem, exchangeable pairs, high dimensions, local dependence, multiple Wiener–Itô integrals, nonlinear statistic, Stein kernel, Stein's method.

multivariate normal approximations. However, they typically require d to be sublinear in n. We discuss some of the recent results in Section 1.1 below.

To date, the proofs for results such as (1.1) in the literature all involve smoothing the maximum function $\max_{1\le j\le d} x_j$ by $\frac{1}{\beta}\log\sum_{j=1}^d e^{\beta x_j}$ for a large β (cf. Theorem 1.3 of Chatterjee (2005)). In this paper, we use a new method to prove high-dimensional normal approximations on hyperrectangles. We assume the approximating normal distribution has a nonsingular covariance matrix. Our method combines the approach of Götze (1991) in Stein's method with modifications of an estimate of Anderson, Hall and Titterington (1998) and a smoothing inequality of Bhattacharya and Rao (1976). We improve the bound in (1.1) to $C_0(\frac{\log^4(dn)}{n})^{1/3}$ when the smallest eigenvalue of $\mathrm{Cov}(W)$ is bounded away from 0 by a constant independent of d and n (cf. Corollary 1.3 below). We further improve the bound to $C_0(\frac{\log^3 d}{n})^{1/2}\log n$, which is optimal up to the $\log n$ factor, for sums of independent and identically distributed (i.i.d.) isotropic random vectors with log-concave distributions (cf. Corollary 1.1 below). Moreover, our method works for general dependent random vectors and we state our main results for W that has a Stein kernel, or admits an exchangeable pair coupling, or is a nonlinear statistic of independent random variables or a sum of locally dependent random vectors. We prove our main results in Section 2. We also discuss an application to multiple Wiener–Itô integrals. Some details are deferred to an Appendix.

Throughout the paper, we always assume $d \ge 3$ so that $\log d > 1$. Also, W denotes a random vector in $\mathbb{R}^d$ with $EW = 0$. We use $Z \sim N(0, \Sigma)$ to denote a d-dimensional Gaussian variable with covariance matrix $\Sigma = (\Sigma_{jk})_{1\le j,k\le d}$ and denote

$$\begin{aligned}
\overline{\sigma}^2 &:= \overline{\sigma}^2(\Sigma) = \max_{1\le j\le d} \Sigma_{jj},\\
\underline{\sigma}^2 &:= \underline{\sigma}^2(\Sigma) = \min_{1\le j\le d} \Sigma_{jj},\\
\sigma_*^2 &:= \sigma_*^2(\Sigma) = \text{smallest eigenvalue of } \Sigma.
\end{aligned} \tag{1.2}$$

Note that in the isotropic case $\Sigma = I_d$, $\overline{\sigma}^2 = \underline{\sigma}^2 = \sigma_*^2 = 1$. We use C to denote positive absolute constants, which may differ in different expressions. We use $\partial_j f$, $\partial_{jk} f$, etc. to denote partial derivatives. For an $\mathbb{R}^d$-vector w, we use w_j, $1 \le j \le d$ to denote its components and write $\|w\|_\infty = \max_{1\le j\le d} |w_j|$.

We first consider random vectors that have a Stein kernel, which was defined in Ledoux, Nourdin and Peccati (2015) and used implicitly in, for example, Chatterjee (2009) and Nourdin and Peccati (2009) (see also Lecture VI of Stein (1986)).

DEFINITION 1.1 (Stein kernel). A $d \times d$ matrix-valued measurable function $\tau^W = (\tau_{ij}^W)_{1\le i,j\le d}$ on $\mathbb{R}^d$ is called a *Stein kernel* for (the law of) W if $E|\tau_{ij}^W(W)| < \infty$ for any $i, j \in \{1, \dots, d\}$ and

$$\sum_{j=1}^d E\big[\partial_j f(W) W_j\big] = \sum_{i,j=1}^d E\big[\partial_{ij} f(W)\tau_{ij}^W(W)\big]$$

for any C^∞ function $f : \mathbb{R}^d \to \mathbb{R}$ with bounded partial derivatives of all orders.

If W has a Stein kernel, then in applying Stein's method, we only need to deal with the second derivatives of the solution to the Stein equation. In this case, we obtain the following simple bound.

THEOREM 1.1 (Error bound using Stein kernels). *Suppose that W has a Stein kernel $\tau^W = (\tau^W_{jk})_{1\le j,k\le d}$. Let $Z \sim N(0,\Sigma)$. Then we have*

$$\sup_{h=1_A:A\in\mathcal{R}} \big|Eh(W) - Eh(Z)\big| \le C\frac{\Delta_W}{\sigma_*^2}(\log d)\bigg(\bigg|\log\bigg(\frac{\underline{\sigma}\Delta_W}{\overline{\sigma}\sigma_*^2}\bigg)\bigg| \vee 1\bigg), \tag{1.3}$$

where the σ's are defined in (1.2) and

$$\Delta_W := E\Big[\max_{1\le j,k\le d}\big|\Sigma_{jk} - \tau^W_{jk}(W)\big|\Big].$$

REMARK 1.1. In practice, we typically choose $\Sigma = \mathrm{Cov}(W)$ (so that $E\tau^W_{jk}(W) = \Sigma_{jk}$), although it is not required in the above theorem. Moreover, since

$$\sup_{h=1_A:A\in\mathcal{R}} \big|Eh(W) - Eh(Z)\big| = \sup_{h=1_A:A\in\mathcal{R}} \big|Eh(MW) - Eh(MZ)\big|$$

for any diagonal matrix M, we have the freedom to do component-wise scaling for W so that the right-hand side of (1.3) is minimized. This minimization problem seems nontrivial, except that one should obvious shrink each component of W as much as possible for a given value of σ_*. This remark applies to all the general bounds below (cf. Theorems 1.2–1.4). For simplicity, in applications below (cf. Corollaries 1.1–1.3), we do the most natural component-wise scaling for W so that $\mathrm{Var}(W_j) = 1$, $1 \le j \le d$ and choose $\Sigma = \mathrm{Cov}(W)$. As a result, $\overline{\sigma} = \underline{\sigma} = 1$ and only $1/\sigma_*$ appears in the upper bound. This factor can be removed if σ_* is bounded away from 0 by an absolute constant. We call it the *strongly nonsingular* case. One example is the isotropic case where $\Sigma = I_d$.

REMARK 1.2. Chernozhukov et al. ((2019), Theorem 5.1) proved[1] that if W has a Stein kernel $\tau^W = (\tau^W_{jk})_{1\le j,k\le d}$ and $Z \sim N(0,\Sigma)$ with the diagonal entries $\Sigma_{jj} \ge c$ for all $j = 1,\dots,d$ and some constant $c > 0$, then

$$\sup_{h=1_A:A\in\mathcal{R}} \big|P(W\in A) - P(Z\in A)\big| \le C'\Delta_W^{1/2}\log d, \tag{1.4}$$

where C' depends only on c. They also showed that the bound (1.4) is asymptotically sharp (personal communication). Theorem 1.1 shows that under the additional assumption that Σ is nonsingular and the ratio of the largest and the smallest diagonal entries of Σ, $\frac{\overline{\sigma}}{\underline{\sigma}}$, is bounded, the bound (1.4) can be improved to

$$C_*\Delta_W \log d\big(|\log\Delta_W| \vee 1\big),$$

where C_* depends only on σ_*. Since Σ is singular in the example attaining the upper bound in (1.4) asymptotically, this improvement comes from the nonsingularity assumption on Σ.

As an illustration, we apply Theorem 1.1 to sums of i.i.d. variables with log-concave densities. Recall that a probability measure μ on $\mathbb{R}^d$ has a log-concave density if it is supported on (the closure of) an open convex set $\Omega \subset \mathbb{R}^d$ and, on Ω, it has a density of the form e^{-V} with $V : \Omega \to \mathbb{R}$ a convex (hence continuous) function; see Saumard and Wellner (2014) for more details. Note that the support of μ must be full dimensional because μ has a density. In this situation, under some regularity assumptions, Fathi (2019) provides a way to construct Stein kernels having some nice properties. This enables us to obtain the following near optimal error bound.

[1](1.4) is deduced from their result together with the proof of Chernozhukov, Chetverikov and Kato ((2017a), Corollary 5.1) and the Stein kernel for $(W^\top, -W^\top)^\top$.

COROLLARY 1.1. *Let μ be a probability measure on $\mathbb{R}^d$ with a log-concave density. Suppose μ has mean 0 and a covariance matrix Σ with diagonal entries all equal to 1 and smallest eigenvalue $\sigma_*^2 > 0$. Let $W = n^{-1/2}\sum_{i=1}^n X_i \in \mathbb{R}^d$ with $n \geq 3$, where $\{X_1, \ldots, X_n\}$ are i.i.d. with law μ. Let $Z \sim N(0, \Sigma)$. Then*

$$\sup_{h=1_A : A\in\mathcal{R}} \bigl|Eh(W) - Eh(Z)\bigr| \leq \frac{C}{\sigma_*^2}\sqrt{\frac{\log^3 d}{n}}\log n.$$

As we see in the following proposition, if σ_* is bounded away from 0 by an absolute constant, $\sqrt{\frac{\log^3 d}{n}}$ is generally the optimal convergence rate in this situation, so the above corollary gives a nearly optimal rate.

PROPOSITION 1.1. *Let $X = (X_{ij})_{i,j=1}^\infty$ be an array of i.i.d. random variables such that $Ee^{c|X_{ij}|} < \infty$ for some $c > 0$, $EX_{ij} = 0$, $EX_{ij}^2 = 1$ and $\gamma := EX_{ij}^3 \neq 0$. Let $W = n^{-1/2}\sum_{i=1}^n X_i$ with $X_i := (X_{i1}, \ldots, X_{\mathrm{id}})^\top$. Suppose that d depends on n so that $(\log^3 d)/n \to 0$ and $d(\log^3 d)/n \to \infty$ as $n \to \infty$. Also, let $Z \sim N(0, I_d)$. Then we have*

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\log^3 d}} \sup_{x\in\mathbb{R}} \Bigl|P\Bigl(\max_{1\leq j\leq d} W_j \leq x\Bigr) - P\Bigl(\max_{1\leq j\leq d} Z_j \leq x\Bigr)\Bigr| > 0.$$

Proposition 1.1 is proved in Section A.1. Note that it is possible to find an example which simultaneously satisfies the assumptions in both Corollary 1.1 and Proposition 1.1. In fact, in the setting of Proposition 1.1, if the law of X_{ij} has a log-concave density, the assumptions of Corollary 1.1 are satisfied due to the independence across the coordinates of X_i. For example, this is the case when X_{ij} follows a normalized exponential distribution.

Theorem 1.1 is also interesting in the context of the so-called Malliavin–Stein method (see Nourdin and Peccati (2012) for an exposition of this topic). For simplicity, we focus on the case that the coordinates of W are multiple Wiener–Itô integrals with common orders. We refer to Nourdin and Peccati (2012) for unexplained concepts appearing in the following corollary (and its proof).

COROLLARY 1.2. *Let X be an isonormal Gaussian process over a real separable Hilbert space $\mathfrak{H}$. Let $q \in \mathbb{N}$ and denote by $I_q(f)$ the qth multiple Wiener–Itô integral of $f \in \mathfrak{H}^{\odot q}$ with respect to X, where $\mathfrak{H}^{\odot q}$ denotes the qth symmetric tensor power of $\mathfrak{H}$. For every $j = 1, \ldots, d$, suppose $W_j = I_q(f_j)$ for some $f_j \in \mathfrak{H}^{\odot q}$. Suppose also $\mathrm{Cov}(W) = \Sigma$ with diagonal entries all equal to 1 and smallest eigenvalue $\sigma_*^2 > 0$. Let $Z \sim N(0, \Sigma)$. Then we have*

$$\sup_{h=1_A : A\in\mathcal{R}} \bigl|Eh(W) - Eh(Z)\bigr| \leq C_q \frac{\overline{\Delta}_W}{\sigma_*^2}\bigl(\log^q d\bigr)\bigl(|\log \overline{\Delta}_W| \vee 1\bigr), \tag{1.5}$$

where $C_q > 0$ is a constant depending only on q and

$$\overline{\Delta}_W := \max_{1\leq j\leq d} \sqrt{EW_j^4 - 3\bigl(EW_j^2\bigr)^2}.$$

Corollary 1.2 is comparable with Corollary 1.3 in Nourdin, Peccati and Yang (2020), where an analogous bound to (1.5) is obtained when $\mathcal{R}$ is replaced by the set of all measurable convex subsets of $\mathbb{R}^d$ (see also Kim and Park (2015) for related results). Meanwhile, considering the restricted class $\mathcal{R}$, we improve the dimension dependence of the bounds: Typically,

the bound of Corollary 1.3 in Nourdin, Peccati and Yang (2020) depends on the dimension through $d^{41/24+1}$, while our bound depends through $\log^q d$.

We also remark that Nourdin, Peccati and Yang (2020) succeeded in removing the logarithmic factor from their bound by an additional recursion argument. However, it does not seem straightforward to apply their argument to our situation.

Stein kernels do not exist for discrete random vectors. Next, we apply other commonly used approaches in Stein's method to exploit the dependence structure of a random vector and obtain error bounds in the normal approximation. First, we consider the exchangeable pair approach developed in Stein (1986) in one-dimensional normal approximations and Chatterjee and Meckes (2008) and Reinert and Röllin (2009) for multivariate normal approximations.

THEOREM 1.2 (Error bound using exchangeable pairs). *Suppose we can construct another random vector W' on the same probability space such that (W, W') and (W', W) have the same distribution (exchangeable), and moreover,*

$$E(W' - W|W) = -\Lambda(W + R) \tag{1.6}$$

for some invertible $d \times d$ matrix Λ (linearity condition). Let $D = W' - W$ and suppose $E\|D\|_\infty^4 < \infty$. Also, let $Z \sim N(0, \Sigma)$. Then, if $\eta \geq 0$ and $t > 0$ satisfy $\eta/\sqrt{t} \leq \sigma_*/\sqrt{\log d}$, we have

$$\begin{aligned}
&\sup_{h=1_A : A \in \mathcal{R}} \big|Eh(W) - Eh(Z)\big| \\
&\quad \leq C\bigg\{\frac{1}{\sigma_*} E\Big(\max_{1\leq j\leq d} |R_j|\Big)\sqrt{\log d} + \frac{1}{\sigma_*^2}\Delta_1\big(|\log t| \vee 1\big)\log d \\
&\qquad + \frac{1}{\sigma_*^4}\big(\Delta_2 + \Delta_3(\eta)\big)\frac{(\log d)^2}{t} + \frac{\overline{\sigma}}{\underline{\sigma}}\sqrt{t}\log d\bigg\},
\end{aligned}$$

where the σ's are defined in (1.2),

$$\begin{aligned}
\Delta_1 &:= E\bigg[\max_{1\leq j,k\leq d}\Big|\Sigma_{jk} - \frac{1}{2}E\big[\big(\Lambda^{-1}D\big)_j D_k|W\big]\Big|\bigg], \\
\Delta_2 &:= E\Big[\max_{1\leq j,k,l,m\leq d} E\big[\big|\big(\Lambda^{-1}D\big)_j D_k D_l D_m\big||W\big]\Big], \\
\Delta_3(\eta) &:= E\Big[\max_{1\leq j,k,l,m\leq d}\big|\big(\Lambda^{-1}D\big)_j D_k D_l D_m\big|; \|D\|_\infty > \eta\Big].
\end{aligned}$$

REMARK 1.3. The exchangeability and the linearity condition in the statement of Theorem 1.2 may be motivated by considering a bivariate normal vector (W, W') with correlation ρ and $E(W' - W|W) = -(1-\rho)W$. If $W = \sum_{i=1}^n \xi_i$ is a sums of independent random vectors and $W' = W - \xi_I + \xi_I'$, where I is an independent random index uniformly chosen from $\{1, \ldots, n\}$ and $\{\xi_i' : 1 \leq i \leq n\}$ is an independent copy of $\{\xi_i : 1 \leq i \leq n\}$, then it can be verified that (1.6) is satisfied with $\Lambda = \frac{1}{n} I_d$ and $R = 0$. The exchangeable pair approach was proved to be useful for dependent random vectors as well; See Reinert and Röllin (2009) and the references therein for many applications.

REMARK 1.4. We can take $\eta = 0$ and $t = (\frac{\sigma}{\overline{\sigma}\sigma_*^4}\Delta_3(0)\log d)^{2/3}$ in Theorem 1.2 to obtain a simpler bound

$$\begin{aligned}
&C\bigg\{\frac{1}{\sigma_*} E\Big(\max_{1\leq j\leq d} |R_j|\Big)\sqrt{\log d} + \frac{1}{\sigma_*^2}\Delta_1\bigg(\bigg|\log\Big(\frac{\sigma}{\overline{\sigma}\sigma_*^4}\Delta_3(0)\Big)\bigg| \vee 1\bigg)\log d \\
&\quad + \bigg(\frac{\overline{\sigma}^2}{\underline{\sigma}^2\sigma_*^4}\Delta_3(0)\log^4 d\bigg)^{1/3}\bigg\}.
\end{aligned}$$

We can simplify the bound in Theorem 1.3 below similarly. However, these simplified bounds result in a worse bound $C(B_n^4(\log^6 d)/\sigma_*^4 n)^{1/3}$ for Corollary 1.3. We introduce the parameter η in the same spirit as in the Chernozhukov–Chetverikov–Kato theory: It plays a similar role to the parameter γ in Chernozhukov, Chetverikov and Kato (2013) and serves for better control of maximal moments appearing in the bound.

We note that Meckes (2006) introduced an infinitesimal version of the exchangeable pairs approach. Her method can be used to find the Stein kernel for certain random vectors with a continuous symmetry; hence, we can apply Theorem 1.1 to obtain a near optimal rate of convergence. In general, however, the convergence rate obtained using Theorem 1.2 is slower, as demonstrated below in Corollary 1.3.

Next, we consider nonlinear statistics along the lines of Chatterjee (2008a), Chen and Röllin (2010) and Dung (2019).

THEOREM 1.3 (Error bound for nonlinear statistics). *Let $X = (X_1, \dots, X_n)$ be a sequence of independent random variables taking values in a measurable space $\mathcal{X}$. Let $F : \mathcal{X}^n \to \mathbb{R}^d$ be a measurable function, and let $W = F(X)$. Let $X' = (X_1', \dots, X_n')$ be an independent copy of X. For each $A \subset \{1, \dots, n\}$, define $X^A = (X_1^A, \dots, X_n^A)$ where*

$$X_i^A = \begin{cases} X_i' & \text{if } i \in A, \\ X_i & \text{if } i \notin A. \end{cases}$$

Let $W^A = F(X^A)$. Suppose $E(W) = 0$ and $E\|W\|_\infty^4 < \infty$. Also, let $Z \sim N(0, \Sigma)$. Then, if $\eta \geq 0$ and $t > 0$ satisfy $\eta/\sqrt{t} \leq \sigma_*/\sqrt{\log d}$, we have

$$\begin{aligned} &\sup_{h=1_A : A \in \mathcal{R}} \big|Eh(W) - Eh(Z)\big| \\ &\quad \leq C\bigg(\frac{1}{\sigma_*^2}\delta_1\big(|\log t| \vee 1\big)\log d + \frac{1}{\sigma_*^4}\big(\delta_2 + \delta_3(\eta)\big)\frac{1}{t}(\log d)^2 + \frac{\overline{\sigma}}{\underline{\sigma}}\sqrt{t}\log d\bigg), \end{aligned}$$

where the σ's are defined in (1.2),

$$\delta_1 := E\bigg[\max_{1 \leq j,k \leq d}\bigg|\Sigma_{jk} - \frac{1}{2}\sum_{i=1}^n \big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j \big(W^{\{i\}} - W\big)_k\bigg|\bigg],$$

$$\delta_2 := E\bigg[\max_{1 \leq j \leq d}\sum_{i=1}^n \big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j^4\bigg] + E\bigg[\max_{1 \leq j \leq d}\sum_{i=1}^n \big(W^{\{i\}} - W\big)_j^4\bigg],$$

$$\begin{aligned} \delta_3(\eta) := &\sum_{i=1}^n E\Big[\max_{1 \leq j \leq d}\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j^4; \big\|W^{\{i\}} - W\big\|_\infty > \eta\Big] \\ &+ \sum_{i=1}^n E\Big[\max_{1 \leq j \leq d}\big(W^{\{i\}} - W\big)_j^4; \big\|W^{\{i\}} - W\big\|_\infty > \eta\Big] \end{aligned}$$

and

$$\{1 : i\} := \begin{cases} \{1, 2, \dots, i\} & \text{if } i \geq 1, \\ \varnothing & \text{if } i = 0. \end{cases}$$

Using either Theorem 1.2 or 1.3 with a truncation argument, we can improve Chernozhukov et al. ((2019), Theorem 2.1) in the strongly nonsingular case.

COROLLARY 1.3. *Let $W = n^{-1/2}\sum_{i=1}^n X_i \in \mathbb{R}^d$, where $\{X_1, \dots, X_n\}$ are centered independent variables with $\mathrm{Cov}(W) = \Sigma$ with diagonal entries all equal to 1 and smallest eigenvalue $\sigma_*^2 > 0$. Let $Z \sim N(0, \Sigma)$. Suppose that there is a constant $B_n \geq 1$ such that $\max_{i,j} E\exp(X_{ij}^2/B_n^2) \leq 2$ and $\max_j n^{-1}\sum_{i=1}^n EX_{ij}^4 \leq B_n^2$, where X_{ij} denotes the jth component of the vector X_i. Then we have*

$$\sup_{h=1_A : A\in\mathcal{R}} \bigl|Eh(W) - Eh(Z)\bigr| \leq C\biggl(\frac{B_n^2 \log^4(dn)}{\sigma_*^4 n}\biggr)^{1/3}. \tag{1.7}$$

Finally, we consider sums of random vectors with a local dependence structure. Unlike in Theorems 1.2 and 1.3, there is no longer an underlying symmetry that we can exploit. In the end, we obtain a third-moment error bound with a slower convergence rate.

THEOREM 1.4 (Error bound for sums of locally dependent variables). *Let $W = \sum_{i=1}^n X_i$ with $EX_i = 0$ and $\mathrm{Cov}(W) = \Sigma$. Let $Z \sim N(0, \Sigma)$. Assume that for each $i \in \{1, \dots, n\}$, there exists $A_i \subset \{1, \dots, n\}$ such that X_i is independent of $\{X_{i'} : i' \notin A_i\}$. Moreover, assume that for each $i \in \{1, \dots, n\}$ and $i' \in A_i$, there exists $A_{ii'} \subset \{1, \dots, n\}$ such that $\{X_i, X_{i'}\}$ is independent of $\{X_{i''} : i'' \notin A_i\}$. Then we have*

$$\begin{aligned}&\sup_{h=1_A : A\in\mathcal{R}} \bigl|Eh(W) - Eh(Z)\bigr| \\ &\leq C\biggl(\frac{\overline{\sigma}}{\underline{\sigma}\sigma_*^3}\sum_{i=1}^n \sum_{i'\in A_i}\sum_{i''\in A_{ii'}} E\Bigl[\max_{1\leq j,k,l\leq d}\bigl(|X_{ij}X_{i'k}X_{i''l}| \\ &\qquad + |X_{ij}X_{i'k}|E|X_{i''l}|\bigr)\Bigr](\log d)^{5/2}\biggr)^{1/2},\end{aligned} \tag{1.8}$$

where the σ's are defined in (1.2).

Theorem 1.4 may be improved using a truncation as in Theorems 1.2 and 1.3. We leave it as it is for simplicity.

1.1. *Literature on multivariate normal approximations.* There is a large literate on multivariate normal approximations. Here we discuss some of the recent results providing error bounds on various distributional distances with the best-known dependence on dimension.

Let $W = n^{-1/2}\sum_{i=1}^n X_i \in \mathbb{R}^d$, where $\{X_1, \dots, X_n\}$ are centered independent variables with $\mathrm{Cov}(W) = \Sigma$. Bentkus (2005) proved that, with $Z \sim N(0, \Sigma)$,

$$\sup_{h=1_A : A\in\mathcal{C}} \bigl|Eh(W) - Eh(Z)\bigr| \leq \frac{Cd^{1/4}}{n^{3/2}}\sum_{i=1}^n E\bigl[|\Sigma^{-1/2}X_i|^3\bigr], \tag{1.9}$$

where $\mathcal{C}$ is the collection of all (measurable) convex sets of $\mathbb{R}^d$ and $|\cdot|$ denotes the Euclidean norm when applied to a vector. The bound (1.9) is optimal up to the $d^{1/4}$ factor (Nagaev (1976)). See Raič (2019) for a bound with explicit constant. In the typical case where $E[|\Sigma^{-1/2}X_i|^3]$ is of the order $O(d^{3/2})$, the error bound in (1.9) is of the order $O(\frac{d^{7/2}}{n})^{1/2}$. Chernozhukov, Chetverikov and Kato (2013) and subsequent works showed that by restricting to the class of hyperrectangles, one may obtain much better dependence on d.

If we restrict to the class of Euclidean balls and assume $\Sigma = I_d$, then we can obtain a bound as in (1.9) but without the factor $d^{1/4}$. This follows from Raič ((2019), Theorem 1.3 and Example 1.2) and Sazonov ((1972), Remark 2.1), for example. For Euclidean balls centered at 0, it is known that (cf. Götze and Zaitsev (2014)) the dependence on n may be improved

from $1/\sqrt{n}$ to $1/n$ for $d \geq 5$, which is in general the smallest possible dimension for such an improvement. See also Esseen (1945), Bentkus and Götze (1997), Götze and Ulyanov (2003), Bogatyrev, Götze and Ulyanov (2006) and Prokhorov and Ulyanov (2013) for earlier and related results. We do not know any results with explicit dependence on d and such improved dependence on n.

Another class of distributional distances of interest is the L^p transportation distance for a number $p \geq 1$, also known as the Kantorovich distance or the p-Wasserstein distance. For two probability measures μ and ν on $\mathbb{R}^d$, it is defined to be

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\gamma \in \Lambda(\mu,\nu)} \int |x - y|^p \, d\gamma(x, y) \right)^{1/p},$$

where $\Lambda(\mu, \nu)$ is the space of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with μ and ν as marginals. If X and Y are random variables with distributions μ and ν, respectively, we will also write

$$\mathcal{W}_p(X, Y) = \mathcal{W}_p(\mu, \nu).$$

Let $W = n^{-1/2} \sum_{i=1}^n X_i \in \mathbb{R}^d$, where $\{X_1, \ldots, X_n\}$ are centered i.i.d. variables with $\mathrm{Cov}(W) = \Sigma$, and let $Z \sim N(0, \Sigma)$. Suppose $|X_i| \leq \beta$ almost surely for some $\beta > 0$. Eldan, Mikulincer and Zhai (2020) proved that

$$\mathcal{W}_2(W, Z) \leq \frac{\beta \sqrt{d} \sqrt{32 + 2 \log_2(n)}}{\sqrt{n}}. \tag{1.10}$$

The bound in (1.10) is optimal up to the $\log_2(n)$ factor (Zhai (2018)). See Courtade, Fathi and Pananjady (2019) and Eldan, Mikulincer and Zhai (2020) for results for the log-concave case. Following the proof of Proposition 1.4 of Zhai (2018), these bounds on the L^2 transportation distance can be used to deduce a bound on $\sup_{h=1_A : A \in \mathcal{R}} |Eh(W) - Eh(Z)|$. For example, we can obtain the following proposition. We defer its proof to the end of the Appendix.

PROPOSITION 1.2. *Let T be any $\mathbb{R}^d$-valued random variable. Let $Z \sim N(0, \Sigma)$ where $\Sigma_{jj} \geq 1, \forall 1 \leq j \leq d$. Then,*

$$\sup_{h=1_A : A \in \mathcal{R}} \left| Eh(T) - Eh(Z) \right| \leq C (\log d)^{1/3} \mathcal{W}_2(T, Z)^{2/3}.$$

Applying Proposition 1.2 to (1.10), we have the following corollary.

COROLLARY 1.4. *Let $W = n^{-1/2} \sum_{i=1}^n X_i \in \mathbb{R}^d$, where $\{X_1, \ldots, X_n\}$ are centered i.i.d. variables with $\mathrm{Cov}(W) = \Sigma$. Suppose $\Sigma_{jj} \geq 1, \forall 1 \leq j \leq d$. Suppose further that $|X_i| \leq \beta$ almost surely for some $\beta > 0$. Let $Z \sim N(0, \Sigma)$. Then,*

$$\sup_{h=1_A : A \in \mathcal{R}} \left| Eh(W) - Eh(Z) \right| \leq C (\log d)^{1/3} \frac{d^{1/3} \beta^{2/3}}{n^{1/3}} (1 + \log n)^{1/3}.$$

Since the $\mathcal{W}_2$ error bound in (1.10) scales like $\sqrt{d}$, we see that such a deduced bound again requires d to be sublinear in n. Allowing to go well beyond this restriction is a key feature of this paper.

If we assume in addition that the mixed third moments of X_1 are all equal to zero, then it is possible to improve the dependence on n from $1/\sqrt{n}$ to $1/n$. See, for example, Bobkov, Chistyakov and Götze (2013) for such improved rate in total variation in dimension one and Fathi (2021) for results on the 2-Wasserstein distance in multi-dimensions.

2. Proofs.

2.1. *Lemmas.* We first state four lemmas that are needed in the proofs of the main results. Set $R(0;\epsilon) := \{x \in \mathbb{R}^d : \|x\|_\infty \le \epsilon\}$ for $\epsilon > 0$. Throughout this section, we denote by ϕ the density function of the standard d-dimensional normal distribution.

LEMMA 2.1 (Gaussian anti-concentration inequality). *Let Y be a centered Gaussian vector in $\mathbb{R}^d$ such that $\min_{1\le j\le d} EY_j^2 \ge \underline{\sigma}^2$ for some $\underline{\sigma} > 0$. Then, for any $y \in \mathbb{R}^d$ and $\varepsilon > 0$,*

$$P(Y \le y+\varepsilon) - P(Y \le y) \le \frac{\varepsilon}{\underline{\sigma}}(\sqrt{2\log d} + 2),$$

where $\{Y \le y\} := \{Y_j \le y_j : 1 \le j \le d\}$.

A proof of Lemma 2.1 is found in Chernozhukov, Chetverikov and Kato (2017b).

LEMMA 2.2 (Modification of (2.10) of Anderson, Hall and Titterington (1998)). *Let $K \ge 0$ be a constant and set $\eta = \eta_d := K/\sqrt{\log d}$. Then, for all $r \in \mathbb{N}$, there is a constant $C_{K,r} > 0$ depending only on K and r such that*

$$\sup_{A\in\mathcal{R}} \sum_{j_1,\dots,j_r=1}^{d} \sup_{y\in R(0;\eta)} \left| \int_A \partial_{j_1,\dots,j_r} \phi(z+y)\, dz \right| \le C_{K,r}(\log d)^{r/2}.$$

The special version of Lemma 2.2 with $\eta = 0$ is found in the proof of (2.10) of Anderson, Hall and Titterington (1998). Introduction of the parameter η is motivated by a standard argument used in the Chernozhukov–Chetverikov–Kato theory to efficiently control maximal moments appearing in normal approximation error bounds; see Equation (24) in Chernozhukov, Chetverikov and Kato (2013), for example.

To clarify the structure of the proof, we first give a proof of the case with $r = 1$ and $\eta = 0$ here. The proof of the general case follows the same strategy and will be given in Section A.2, where we need a few technical lemmas and more complicated notation.

PROOF OF LEMMA 2.2 WITH $r = 1$ AND $\eta = 0$. We denote by ϕ_1 and Φ_1 the density and distribution function of the standard normal distribution, respectively. We set $\bar{\phi}_1(u) := \phi_1(u)/\Phi_1(u)$.

For any $A = \prod_{j=1}^d (a_j, b_j) \in \mathcal{R}$, we have, by considering $x_j := |a_j| \wedge |b_j|$ in the first inequality,

$$\begin{aligned}
\sum_{j=1}^{d} \left| \int_A \partial_j \phi(z)\, dz \right| &= \sum_{j=1}^{d} |\phi_1(b_j) - \phi_1(a_j)| \prod_{k:k\ne j} (\Phi_1(b_k) - \Phi_1(a_k)) \\
&\le \sup_{x\in[0,\infty)^d} \sum_{j=1}^{d} \phi_1(x_j) \prod_{k:k\ne j} \Phi_1(x_k) \qquad (2.1) \\
&= \sup_{x\in[0,\infty)^d} \sum_{j=1}^{d} \bar{\phi}_1(x_j) \prod_{k=1}^{d} \Phi_1(x_k).
\end{aligned}$$

Therefore, it suffices to prove $\sup_{x\in[0,\infty)^d} f(x) = O(\sqrt{\log d})$, where

$$f(x) = F(x)G(x) \quad \text{with } F(x) = \sum_{j=1}^{d} \bar{\phi}_1(x_j) \quad \text{and} \quad G(x) = \prod_{k=1}^{d} \Phi_1(x_k).$$

The remaining proof proceeds as follows. We first show that f has a maximizer x^* satisfying $x_1^* = \cdots = x_d^* =: u^*$. From this, we will see $\sup_{x\in[0,\infty)^d} f(x) = O(u^*)$ and $u^* = O(\sqrt{\log d})$ as $d \to \infty$. This completes the proof.

Noting that $\bar{\phi}_1'(u) = -(u + \bar{\phi}_1(u))\bar{\phi}_1(u)$, we have

$$\begin{aligned}\partial_l f(x) &= \bar{\phi}_1'(x_l)G(x) + F(x)\bar{\phi}_1(x_l)G(x)\\ &= \bigl\{-\bigl(x_l + \bar{\phi}_1(x_l)\bigr) + F(x)\bigr\}\bar{\phi}_1(x_l)G(x).\end{aligned}$$

In particular, $\partial_l f(x) < 0$ if $x \in [0,\infty)^d$ and $x_l \geq d$ because $F(x) \leq d\sqrt{2/\pi} < d$ for all $x \in [0,\infty)^d$. This means $f(x) \leq f(x_1 \wedge d, \ldots, x_d \wedge d)$ for all $x \in [0,\infty)^d$, and thus we obtain $\sup_{x\in[0,\infty)^d} f(x) = \sup_{x\in[0,d)^d} f(x)$. Note also that $\partial_l f(x) > 0$ if $x_l = 0$. As a result, f has a maximizer x^* satisfying $x^* \in (0,d)^d$ and $\nabla f(x^*) = 0$. The latter equation yields

$$x_1^* + \bar{\phi}_1\bigl(x_1^*\bigr) = \cdots = x_d^* + \bar{\phi}_1\bigl(x_d^*\bigr) = F\bigl(x^*\bigr). \tag{2.2}$$

Now, it is easy to see that the function $[0,\infty) \ni u \mapsto u + \bar{\phi}_1(u) \in (0,\infty)$ is strictly increasing (this is in fact a special case of Lemma A.2). Consequently, we obtain $x_1^* = \cdots = x_d^* =: u^*$.

Now, from (2.2) we have $u^* = (d-1)\bar{\phi}_1(u^*)$. So we obtain

$$\sup_{x\in[0,\infty)^d} f(x) = f\bigl(x^*\bigr) = \frac{d}{d-1}u^*\Phi_1\bigl(u^*\bigr)^d \leq \frac{d}{d-1}u^*.$$

Therefore, we complete the proof once we prove $u^* = O(\sqrt{\log d})$.

Setting $g_2(u) := u - (d-1)\bar{\phi}_1(u)$, we have $g_2(u^*) = 0$ and $g_2(\sqrt{2\log d}) \to \infty$ as $d \to \infty$. Since g_2 is increasing on $[0,\infty)$, we conclude $u^* = O(\sqrt{\log d})$ as $d \to \infty$. □

From Lemma 2.2, we can obtain the following lemma. For any bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$ and $\sigma > 0$, we define the function $\mathcal{N}_\sigma f : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathcal{N}_\sigma f(x) = \int_{\mathbb{R}^d} f(x + \sigma z)\phi(z)\,dz = \int_{\mathbb{R}^d} f(\sigma z)\phi(z - x/\sigma)\,dz. \tag{2.3}$$

Note that $\mathcal{N}_\sigma f$ is infinitely differentiable.

LEMMA 2.3. *Let $K \geq 0$ be a constant and set $\eta = \eta_d := K/\sqrt{\log d}$. Then, for all $r \in \mathbb{N}$, there is a constant $C_{K,r} > 0$ depending only on K and r such that*

$$\sup_{h=1_A : A\in\mathcal{R}} \sup_{x\in\mathbb{R}^d} \sum_{j_1,\ldots,j_r=1}^{d} \sup_{y\in R(0;\sigma\eta)} \bigl|\partial_{j_1,\ldots,j_r}\mathcal{N}_\sigma h(x+y)\bigr| \leq C_{K,r}\sigma^{-r}(\log d)^{r/2}$$

for any $h = 1_A$, $A \in \mathcal{R}$ and $\sigma > 0$.

PROOF. Fix $h = 1_A$, $A \in \mathcal{R}$, arbitrarily. For any $x, y \in \mathbb{R}^d$ and $j_1, \ldots, j_r \in \{1, \ldots, d\}$, we have

$$\begin{aligned}\partial_{j_1,\ldots,j_r}\mathcal{N}_\sigma h(x+y) &= (-1)^r\sigma^{-r}\int_{\mathbb{R}^d} h(\sigma z)\partial_{j_1,\ldots,j_r}\phi\bigl(z - (x+y)/\sigma\bigr)\,dz\\ &= (-1)^r\sigma^{-r}\int_{\mathbb{R}^d} h(x+\sigma z)\partial_{j_1,\ldots,j_r}\phi(z - y/\sigma)\,dz\\ &= (-1)^r\sigma^{-r}\int_{\sigma^{-1}(A-x)} \partial_{j_1,\ldots,j_r}\phi(z - y/\sigma)\,dz,\end{aligned}$$

where $\sigma^{-1}(A-x) := \{\sigma^{-1}(z-x) : z \in A\} \in \mathcal{R}$. Hence we obtain

$$\sum_{j_1,\dots,j_r=1}^{d} \sup_{y\in R(0;\sigma\eta)} \bigl|\partial_{j_1,\dots,j_r}\mathcal{N}_\sigma h(x+y)\bigr|$$

$$\le \sigma^{-r} \sup_{A\in\mathcal{R}} \sum_{j_1,\dots,j_r=1}^{d} \sup_{y\in R(0;\eta)} \left|\int_A \partial_{j_1,\dots,j_r}\phi(z+y)\,dz\right|.$$

Now, the desired result follows from Lemma 2.2. □

Next, we state a smoothing lemma. The test function $h = 1_A$ we deal with in bounding $\sup_{h=1_A:A\in\mathcal{R}} |Eh(W) - Eh(Z)|$ is not continuous. It is a common strategy to smooth it first, then quantify the error introduced by such smoothing, finally balance the smoothing error with the smooth test function bound. We follow Bhattacharya and Rao (1976) to smooth h by convoluting it with a Gaussian distribution K with a small variance.

LEMMA 2.4 (Modification of Lemma 11.4 of Bhattacharya and Rao (1976)). *Let μ, ν, K be probability measures on $\mathbb{R}^d$. Let $\epsilon > 0$ be a constant such that*

$$\alpha := K\bigl(R(0;\epsilon)\bigr) > 1/2.$$

Let $h : \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function. Then we have

$$\left|\int h\,d(\mu-\nu)\right| \le (2\alpha-1)^{-1}\bigl[\gamma^*(h;\epsilon) + \tau^*(h;2\epsilon)\bigr],$$

where

$$\gamma^*(h;\epsilon) = \sup_{y\in\mathbb{R}^d} \gamma(h_y;\epsilon), \qquad \tau^*(h;2\epsilon) = \sup_{y\in\mathbb{R}^d} \tau(h_y;2\epsilon), \qquad h_y(x) = h(x+y),$$

$$\gamma(h;\epsilon) = \max\left\{\int M_h(\cdot;\epsilon)d(\mu-\nu) * K, -\int m_h(\cdot;\epsilon)d(\mu-\nu) * K\right\},$$

$$\tau(\cdot;2\epsilon) = \int \bigl[M_h(\cdot;2\epsilon) - m_h(\cdot;2\epsilon)\bigr]\,d\nu,$$

$$M_h(x;\epsilon) = \sup_{y:\|y-x\|_\infty\le\epsilon} h(y), \qquad m_h(x;\epsilon) = \inf_{y:\|y-x\|_\infty\le\epsilon} h(y)$$

and $*$ denotes the convolution of two probability measures.

Lemma 2.4 can be shown in a completely parallel way to that of Lemma 11.4 in Bhattacharya and Rao (1976) by changing the ϵ-balls therein to ϵ-rectangles, so we omit its proof.

2.2. *Basic estimates.* In Theorems 1.1–1.4, we aim to bound

$$\delta := \sup_{h=1_A:A\in\mathcal{R}} \bigl|Eh(W) - Eh(Z)\bigr|, \qquad Z \sim N(0,\Sigma).$$

In this subsection, we collect some basic estimates used in all of their proofs. Fix $A \in \mathcal{R}$. Let

$$h = 1_A, \qquad \tilde{h} = 1_A - P(Z \in A).$$

For $s > 0$, let

$$T_s\tilde{h}(x) = E\tilde{h}\bigl(e^{-s}x + \sqrt{1-e^{-2s}}Z\bigr).$$

Note that $ET_s\tilde{h}(Z) = E\tilde{h}(Z) = 0$. Let Q and G be the laws of W and Z, respectively. For a probability distribution μ on $\mathbb{R}^d$ and $\sigma \geq 0$, we denote by μ_σ the law of the random vector σY with $Y \sim \mu$. For $t > 0$ to be chosen, we are going to apply Lemma 2.4 with

$$\mu = Q_{e^{-t}}, \qquad \nu = G_{e^{-t}}, \qquad K = G_{\sqrt{1-e^{-2t}}}, \qquad h = 1_A,$$

and ϵ be such that

$$G_{\sqrt{1-e^{-2t}}}\{\|z\|_\infty \leq \epsilon\} = 7/8.$$

We first bound $\tau^*(h; 2\epsilon)$ in Lemma 2.4. Recall the definition of σ's from (1.2). Markov's inequality and Lemma 2.1 of Chatterjee (2008b) yield

$$\epsilon \leq C\sqrt{1-e^{-2t}}E\|Z\|_\infty \leq C\overline{\sigma}\sqrt{1-e^{-2t}}\sqrt{\log d}.$$

Thus, applying the Gaussian anti-concentration inequality in Lemma 2.1 with $Y = (e^{-t}Z^\top, -e^{-t}Z^\top)^\top$, we obtain

$$\tau^*(h; 2\epsilon) \leq Ce^t\frac{\log d}{\underline{\sigma}}\epsilon \leq \frac{C\overline{\sigma}}{\underline{\sigma}}e^t\sqrt{t}\log d, \tag{2.4}$$

where we used the elementary inequality $1 - e^{-x} \leq x$ for all $x \geq 0$.

Now we turn to bounding $\gamma^*(h; \epsilon)$ in Lemma 2.4. Note that $M_h(\cdot; \epsilon)$ and $m_h(\cdot; \epsilon)$ are again indicator functions of rectangles. Note also that the class $\mathcal{R}$ is invariant under translation and scalar multiplication. Therefore, it suffices to obtain a uniform upper bound for the absolute value of

$$\int h d(\mu - \nu) * K = \int \tilde{h}\, d\mu * K = ET_t\tilde{h}(W) \tag{2.5}$$

over all $h = 1_A$, $A \in \mathcal{R}$. In fact, we have by Lemma 2.4 and (2.4)–(2.5)

$$\delta \leq C\Big(\sup_{h=1_A:A\in\mathcal{R}}|ET_t\tilde{h}(W)| + \frac{\overline{\sigma}}{\underline{\sigma}}e^t\sqrt{t}\log d\Big). \tag{2.6}$$

We use (various versions of) Stein's method to bound $ET_t\tilde{h}(W)$. Similar to (1.14) and (3.1) of Bhattacharya and Holmes (2010), one can verify that

$$\psi_t(x) = -\int_t^\infty T_s\tilde{h}(x)\, ds$$

is a solution to the Stein equation

$$\langle\Sigma, \operatorname{Hess}\psi_t(w)\rangle_{H.S.} - w \cdot \nabla\psi_t(w) = T_t\tilde{h}(w).$$

Thus we have

$$ET_t\tilde{h}(W) = E\big[\langle\Sigma, \operatorname{Hess}\psi_t(W)\rangle_{H.S.} - W \cdot \nabla\psi_t(W)\big]. \tag{2.7}$$

Set $\tilde{\Sigma} := \Sigma - \sigma_*^2 I_d$. Note that $\tilde{\Sigma}$ is positive semidefinite because σ_*^2 is the smallest eigenvalue of Σ. Let us take independent random vectors $\tilde{Z}$ and Z' such that $\tilde{Z} \sim N(0, \tilde{\Sigma})$, $Z' \sim N(0, I_d)$ and they are independent of everything else. Then, since $\tilde{Z} + \sigma_* Z' \sim N(0, \Sigma)$, we can rewrite $T_s\tilde{h}(x)$ as

$$\begin{aligned} T_s\tilde{h}(x) &= E\tilde{h}\big(e^{-s}x + \sqrt{1-e^{-2s}}\tilde{Z} + \sigma_*\sqrt{1-e^{-2s}}Z'\big) \\ &= E\mathcal{N}_{\sigma_*\sqrt{1-e^{-2s}}}\tilde{h}\big(e^{-s}x + \sqrt{1-e^{-2s}}\tilde{Z}\big), \end{aligned} \tag{2.8}$$

where $\mathcal{N}_{\sigma_*\sqrt{1-e^{-2s}}}\tilde{h}$ is defined by (2.3). Therefore, applying Lemma 2.3 with $\eta = 0$ and noting (3.14)–(3.15) of Bhattacharya and Holmes (2010), we obtain

$$\sum_{j=1}^{d} \bigl|\partial_j \psi_t(x)\bigr| \le C\sigma_*^{-1}\sqrt{\log d}, \tag{2.9}$$

$$\sum_{j,k=1}^{d} \bigl|\partial_{jk} \psi_t(x)\bigr| \le C\sigma_*^{-2}\bigl(|\log t| \vee 1\bigr)(\log d), \tag{2.10}$$

$$\sum_{j,k,l=1}^{d} \bigl|\partial_{jkl} \psi_t(x)\bigr| \le C\sigma_*^{-3}\frac{1}{\sqrt{t}}(\log d)^{3/2}. \tag{2.11}$$

2.3. *Proof of Theorem 1.1.* Without loss of generality, we may assume $\frac{\Delta_W}{\sigma_*^2} < 1$; otherwise, the bound (1.3) is trivial. Since W has a Stein kernel τ^W, we obtain by (2.7)

$$ET_t\tilde{h}(W) = E\biggl[\sum_{j,k=1}^{d} \partial_{jk}\psi_t(W)\bigl(\Sigma_{jk} - \tau_{jk}^W(W)\bigr)\biggr].$$

Therefore, we deduce by (2.10)

$$\bigl|ET_t\tilde{h}(W)\bigr| \le \frac{C}{\sigma_*^2}(\log d)\Delta_W\bigl(|\log t| \vee 1\bigr).$$

Consequently, we have by (2.6)

$$\delta \le C\biggl\{\frac{\overline{\sigma}}{\underline{\sigma}}e^t\sqrt{t}\log d + \frac{1}{\sigma_*^2}(\log d)\Delta_W\bigl(|\log t| \vee 1\bigr)\biggr\}.$$

Setting $\sqrt{t} = \frac{\underline{\sigma}\Delta_W}{\overline{\sigma}\sigma_*^2}$ and noting $\frac{\Delta_W}{\sigma_*^2} < 1$, we obtain the desired result.

2.4. *Proof of Corollary 1.1.* Without loss of generality, we may assume

$$\frac{1}{\sigma_*^2}\sqrt{\frac{\log^3 d}{n}} \le 1. \tag{2.12}$$

As in the proof of Fathi ((2019), Theorem 3.3), we first prove the result when μ is compactly supported and its density is bounded away from zero on its support. Then, by Theorem 2.3 and Proposition 3.2 in Fathi (2019), μ has a Stein kernel $\tau = (\tau_{jk})_{1\le j,k\le d}$ such that $\tau(x)$ is positive definite for all $x \in \mathbb{R}^d$ and $\max_{1\le j\le d} E[|\tau_{jj}(X_1)|^p] \le 8^p p^{2p}$ for all $p \ge 1$ (here we used the assumption that $\mathrm{Var}(W_j) = 1, \forall 1 \le j \le d$). Note that we indeed have $\max_{1\le j,k\le d} E[|\tau_{jk}(X_1)|^p] \le 8^p p^{2p}$ for all $p \ge 1$ due to positive definiteness. In particular, Lemma A.7 in Koike (2019a) (with $q = 4$) yields

$$\max_{1\le j,k\le d} E\exp\biggl(\sqrt{\frac{|\tau_{jk}(X_1)|}{C}}\biggr) \le 2. \tag{2.13}$$

Now we define the function $\tau^W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ by

$$\tau^W(x) = E\biggl[\frac{1}{n}\sum_{i=1}^{n}\tau(X_i)\Big|W = x\biggr], \quad x \in \mathbb{R}^d.$$

It is well known that τ^W is a Stein kernel for W (cf. page 271 of Ledoux, Nourdin and Peccati (2015)). Jensen's inequality yields

$$E\Big[\max_{1\le j,k\le d}|\Sigma_{jk}-\tau_{jk}^W(W)|\Big]\le E\bigg[\max_{1\le j,k\le d}\bigg|\frac{1}{n}\sum_{i=1}^n\big(\tau_{jk}(X_i)-\Sigma_{jk}\big)\bigg|\bigg]. \tag{2.14}$$

We will use Theorem 3.1 and Remark A.1 in Kuchibhotla and Chakrabortty (2018) to bound the right-hand side of (2.14). We need the following definitions.

DEFINITION 2.1 (Orlicz norms). Let $g:[0,\infty)\to[0,\infty)$ be a nondecreasing function with $g(0)=0$. The "g-Orlicz norm" of a random variable X is given by

$$\|X\|_g:=\inf\big\{\eta>0:E\big[g\big(|X|/\eta\big)\big]\le 1\big\}.$$

DEFINITION 2.2 (Sub-Weibull variables). A random variable X is said to be sub-Weibull of order $\alpha>0$, denotes as sub-Weibull (α), if

$$\|X\|_{\psi_\alpha}<\infty,\quad \text{where } \psi_\alpha(x)=\exp\big(x^\alpha\big)-1 \text{ for } x\ge 0.$$

DEFINITION 2.3 (Generalized Bernstein–Orlicz norm). Fix $\alpha>0$ and $L\ge 0$. Define the function $\Psi_{\alpha,L}(\cdot)$ based on the inverse function

$$\Psi_{\alpha,L}^{-1}(t):=\sqrt{\log(1+t)}+L\big(\log(1+t)\big)^{1/\alpha}\quad \text{for all } t\ge 0.$$

The generalized Bernstein–Orlicz (GBO) norm of a random variable X is then given by $\|X\|_{\Psi_{\alpha,L}}$ as in Definition 2.1.

Applying Theorem 3.1 of Kuchibhotla and Chakrabortty (2018) to the sequence of independent mean zero sub-Weibull $(\frac{1}{2})$ random variable (cf. (2.13)) $\{\tau_{jk}(X_i)-\Sigma_{jk}:i=1,\dots,n\}$, we have

$$\bigg\|\frac{1}{n}\sum_{i=1}^n\big(\tau_{jk}(X_i)-\Sigma_{jk}\big)\bigg\|_{\Psi_{\frac{1}{2},L_n}}\le\frac{C}{\sqrt{n}},$$

for some $L_n=C/\sqrt{n}$. Combining with Remark A.1 of Kuchibhotla and Chakrabortty (2018), we have,

$$\begin{aligned}
&E\bigg[\max_{1\le j,k\le d}\bigg|\frac{1}{n}\sum_{i=1}^n\big(\tau_{jk}(X_i)-\Sigma_{jk}\big)\bigg|\bigg]\\
&\quad\le C\max_{1\le j,k\le d}\bigg\|\frac{1}{n}\sum_{i=1}^n\big(\tau_{jk}(X_i)-\Sigma_{jk}\big)\bigg\|_{\Psi_{\frac{1}{2},L_n}}\big(\sqrt{\log d}+L_n\log^2 d\big)\\
&\quad\le C\frac{1}{\sqrt{n}}\bigg(\sqrt{\log d}+\frac{1}{\sqrt{n}}\log^2 d\bigg).
\end{aligned}\tag{2.15}$$

From (2.14) and (2.15), we have

$$\begin{aligned}
E\Big[\max_{1\le j,k\le d}|\Sigma_{jk}-\tau_{jk}^W(W)|\Big]&\le C\frac{1}{\sqrt{n}}\bigg(\sqrt{\log d}+\frac{1}{\sqrt{n}}\log^2 d\bigg)\\
&=C\sqrt{\frac{\log d}{n}}\bigg(1+\frac{\log^{3/2} d}{\sqrt{n}}\bigg)\le C\sqrt{\frac{\log d}{n}},
\end{aligned}$$

where the last inequality follows from (2.12). Therefore, an application of Theorem 1.1, together with the fact that $x(|\log x| \vee 1)$ is an increasing function for $x \geq 0$ and the assumption (2.12), yields the desired result.

Next we prove the result when μ is supported on the whole space $\mathbb{R}^d$. We take a sequence of convex bodies F_ℓ that converge to $\mathbb{R}^d$. Define the probability measure ν_ℓ on $\mathbb{R}^d$ by $\nu_\ell(A) = \mu(A \cap F_\ell)/\mu(F_\ell)$ for any Borel set $A \subset \mathbb{R}^d$ (note that $\mu(F_\ell) \to 1$ as $\ell \to \infty$ by construction, so $\mu(F_\ell) > 0$ for sufficiently large ℓ). Then, let μ_ℓ be the law of the variable $M_\ell^{-1/2}(Y_\ell - EY_\ell)$, where Y_ℓ is a random vector with law ν_ℓ and M_ℓ is the diagonal matrix with the diagonal entries equal to those of $\mathrm{Cov}(Y_\ell)$ (note that $\mathrm{Cov}(Y_\ell) \to \Sigma$ as $\ell \to \infty$ by construction, so $M_\ell^{-1/2}$ exists for sufficiently large ℓ). Note that $M_\ell \to I_d$. Also, the density of μ_ℓ is bounded away from zero on its support because μ is supported on $\mathbb{R}^d$ and has a continuous density. Hence, letting $W_\ell = n^{-1/2}\sum_{i=1}^n X_i^{(\ell)} \in \mathbb{R}^d$ with $\{X_1^{(\ell)}, \dots, X_n^{(\ell)}\}$ being i.i.d. with law μ_ℓ and using the result for the compactly supported case above, we have, for sufficiently large ℓ,

$$\sup_{h=1_A : A \in \mathcal{R}} \bigl|Eh(W_\ell) - Eh(Z)\bigr| \leq \frac{C}{\sigma_*^2}\sqrt{\frac{\log^3 d}{n}}\log n.$$

Moreover, it is also easy to verify that the density of W_ℓ converges almost everywhere to that of W as $\ell \to \infty$. Thus, Scheffé's lemma yields

$$\sup_{h=1_A : A \in \mathcal{R}} \bigl|Eh(W_\ell) - Eh(W)\bigr| \to 0 \quad (\ell \to \infty).$$

This yields the desired result.

Finally, to prove the result in the general case, take $\epsilon > 0$ arbitrarily and let μ_ϵ be the law of the variable $\sqrt{1-\epsilon^2}X_1 + \epsilon\zeta$, where $\zeta \sim N(0, \Sigma)$ and is independent of $\{X_1, \dots, X_n\}$. It is evident that μ_ϵ has covariance matrix Σ and is supported on the whole space $\mathbb{R}^d$. Moreover, μ_ϵ has a log-concave density by Proposition 3.5 in Saumard and Wellner (2014). Hence we have for any $A \in \mathcal{R}$

$$\bigl|E1_A(W) - E1_A(Z)\bigr| \leq \bigl|E1_A(W) - E1_A\bigl(W^\epsilon\bigr)\bigr| + \frac{C}{\sigma_*^2}\sqrt{\frac{\log^3 d}{n}}\log n,$$

where $W^\epsilon := \sqrt{1-\epsilon^2}W + \epsilon\zeta$. Since W has a density and W^ϵ converges in law to W as $\epsilon \to 0$, $|E1_A(W) - E1_A(W^\epsilon)| \to 0$ as $\epsilon \to 0$. Thus, letting $\epsilon \to 0$ and taking the supremum over $A \in \mathcal{R}$ in the above inequality, we complete the proof.

2.5. *Proof of Corollary 1.2.* By Proposition 3.7 in Nourdin, Peccati and Swan (2014), W has a Stein kernel $\tau^W = (\tau_{jk}^W)_{1 \leq j,k \leq d}$ given by

$$\tau_{jk}^W(x) = E\bigl[\langle -DL^{-1}W_j, DW_k\rangle_{\mathfrak{H}} | W = x\bigr], \quad x \in \mathbb{R}^d,$$

where $\langle \cdot, \cdot \rangle_{\mathfrak{H}}$ denotes the inner product of $\mathfrak{H}$, while D and L^{-1} denote the Malliavin derivative and pseudo inverse of the Ornstein–Uhlenbeck operator with respect to X, respectively. By Jensen's inequality and Lemma 2.2 in Koike (2019a), we have

$$\begin{aligned} E\Bigl[\max_{1 \leq j,k \leq d}\bigl|\Sigma_{jk} - \tau_{jk}^W(W)\bigr|\Bigr] &\leq E\Bigl[\max_{1 \leq j,k \leq d}\bigl|\Sigma_{jk} - \langle -DL^{-1}W_j, DW_k\rangle_{\mathfrak{H}}\bigr|\Bigr] \\ &\leq C_q\bigl(\log^{q-1} d\bigr)\overline{\Delta}_W, \end{aligned}$$

where $C_q > 0$ depends only on q. Thus the desired result follows from Theorem 1.1.

2.6. *Proof of Theorem 1.2.* Without loss of generality, we may assume $t < 1$; otherwise, the theorem is trivial because $\sup_{h=1_A:A\in\mathcal{R}} |Eh(W) - Eh(Z)| \le 1$. By exchangeability we have

$$\begin{aligned}
0 &= \frac{1}{2} E\big[\Lambda^{-1} D \cdot \big(\nabla\psi_t(W') + \nabla\psi_t(W)\big)\big] \\
&= E\bigg[\frac{1}{2}\Lambda^{-1} D \cdot \big(\nabla\psi_t(W') - \nabla\psi_t(W)\big) + \Lambda^{-1} D \cdot \nabla\psi_t(W)\bigg] \\
&= E\bigg[\frac{1}{2}\sum_{j,k=1}^{d} \big(\Lambda^{-1} D\big)_j D_k \partial_{jk}\psi_t(W) + \Xi + \Lambda^{-1} D \cdot \nabla\psi_t(W)\bigg],
\end{aligned}$$

where

$$\Xi = \frac{1}{2}\sum_{j,k,l=1}^{d} \big(\Lambda^{-1} D\big)_j D_k D_l U \partial_{jkl}\psi_t\big(W + (1-U)D\big)$$

and U is a uniform random variable on $[0, 1]$ independent of everything else. Combining this with (1.6), (2.7) and (2.9)–(2.10), we obtain

$$ET_t\tilde{h}(W) \le \frac{1}{\sigma_*} E\Big(\max_{1\le j\le d} |R_j|\Big)\sqrt{\log d} + \frac{1}{\sigma_*^2}\Delta_1\big(|\log t| \vee 1\big)\log d + \big|E[\Xi]\big|.$$

To estimate $|E[\Xi]|$, we rewrite it as follows. By exchangeability we have

$$\begin{aligned}
&E\big[\big(\Lambda^{-1} D\big)_j D_k D_l U \partial_{jkl}\psi_t\big(W + (1-U)D\big)\big] \\
&\quad = -E\big[\big(\Lambda^{-1} D\big)_j D_k D_l U \partial_{jkl}\psi_t\big(W' - (1-U)D\big)\big] \\
&\quad = -E\big[\big(\Lambda^{-1} D\big)_j D_k D_l U \partial_{jkl}\psi_t\big(W + UD\big)\big].
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
E[\Xi] &= \frac{1}{4}\sum_{j,k,l=1}^{d} E\big[\big(\Lambda^{-1} D\big)_j D_k D_l U\big\{\partial_{jkl}\psi_t\big(W + (1-U)D\big) \\
&\qquad - \partial_{jkl}\psi_t(W + UD)\big\}\big] \qquad (2.16) \\
&= \frac{1}{4}\sum_{j,k,l,m=1}^{d} E\big[\big(\Lambda^{-1} D\big)_j D_k D_l D_m U(1-2U)\partial_{jklm}\psi_t(W + \tilde{D})\big],
\end{aligned}$$

where $\tilde{D} := UD + V(1-2U)D$ and V is a uniform random variable on $[0, 1]$ independent of everything else. Note that $|U + V(1-2U)| \le U \vee (1-U) \le 1$ and thus $\|\tilde{D}\|_\infty \le \|D\|_\infty$.

Now, note that $\frac{e^{-s}\eta}{\sigma_*\sqrt{1-e^{-2s}}} \le 1/\sqrt{2\log d}$ for every $s \ge t$ by assumption. Hence, (2.8) and Lemma 2.3 imply

$$\sum_{j,k,l,m=1}^{d} \sup_{y\in R(0;\eta)} \big|\partial_{jklm}\psi_t(x+y)\big| \le C\sigma_*^{-4}\frac{1}{t}\log^2 d.$$

Combining this with (2.16) and $\|\tilde{D}\|_\infty \le \|D\|_\infty$, we obtain

$$\begin{aligned}|E\,\Xi| &\le \frac{1}{4}\sum_{j,k,l,m=1}^{d} E\Big[\big|(\Lambda^{-1}D)_j D_k D_l D_m\big| \sup_{y\in R(0;\eta)} \big|\partial_{jklm}\psi_t(W+y)\big|; \|\tilde{D}\|_\infty \le \eta\Big] \\ &\quad + \frac{1}{4}\sum_{j,k,l,m=1}^{d} E\big[\big|(\Lambda^{-1}D)_j D_k D_l D_m \partial_{jklm}\psi_t(W+\tilde{D})\big|; \|D\|_\infty > \eta\big] \\ &\le C\sigma_*^{-4}\frac{\log^2 d}{t}\big(\Delta_2 + \Delta_3(\eta)\big).\end{aligned}$$

Now the desired result follows from (2.6).

2.7. *Proof of Theorem* 1.3. Without loss of generality, we may assume $t < 1$; otherwise, the theorem is trivial because $\sup_{h=1_A:A\in\mathcal{R}} |Eh(W) - Eh(Z)| \le 1$. We follow the proof of Theorem 1.2 and bound $E(\Delta\psi_t(W) - W\cdot\nabla\psi_t(W))$. From the independence of X' and X and the assumption that $E(W) = 0$ and using the telescoping sum, we have

$$\begin{aligned}EW\cdot\nabla\psi_t(W) &= E\big(W - W^{\{1:n\}}\big)\cdot\nabla\psi_t(W) \\ &= \sum_{i=1}^{n} E\big(W^{\{1:(i-1)\}} - W^{\{1:i\}}\big)\cdot\nabla\psi_t(W).\end{aligned}$$

Exchanging X_i with X_i', we have

$$E\big(W^{\{1:(i-1)\}} - W^{\{1:i\}}\big)\cdot\nabla\psi_t(W) = E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)\cdot\nabla\psi_t\big(W^{\{i\}}\big).$$

Therefore,

$$\begin{aligned}&EW\cdot\nabla\psi_t(W) \\ &= \frac{1}{2}\sum_{i=1}^{n} E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)\cdot\big(\nabla\psi_t\big(W^{\{i\}}\big) - \nabla\psi_t(W)\big) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k=1}^{d} E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j\big(W^{\{i\}} - W\big)_k\partial_{jk}\psi_t(W) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k,l=1}^{d} E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j\big(W^{\{i\}} - W\big)_k \\ &\quad \times U\big(W^{\{i\}} - W\big)_l\partial_{jkl}\psi_t\big(W + UV\big(W^{\{i\}} - W\big)\big),\end{aligned}$$

where U, V are independent uniform random variables on $[0, 1]$ and independent of everything else. Exchanging X_i with X_i' gives

$$\begin{aligned}&E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j\big(W^{\{i\}} - W\big)_k U\big(W^{\{i\}} - W\big)_l\partial_{jkl}\psi_t\big(W + UV\big(W^{\{i\}} - W\big)\big) \\ &\quad = -E\big(W^{\{1:i\}} - W^{\{1:(i-1)\}}\big)_j\big(W^{\{i\}} - W\big)_k U\big(W^{\{i\}} - W\big)_l\partial_{jkl}\psi_t\big(W - UV\big(W^{\{i\}} - W\big)\big).\end{aligned}$$

Following similar arguments as in the proof of Theorem 1.2, we obtain the desired result.

2.8. *Proof of Corollary* 1.3. Without loss of generality, we may assume

$$(4\sqrt{5})^6\frac{B_n^2\log^4(dn)}{\sigma_*^4 n} \le 1. \tag{2.17}$$

We prove the assertion in three steps. In Steps 1 and 2, we truncate the random variables and show that the error introduced by the truncation is negligible. In Step 3, we apply Theorem 1.3 to the truncated variable.

Step 1. Set $\kappa_n := B_n\sqrt{5\log(dn)}$. For $i = 1, \dots, n$ and $j = 1, \dots, d$, define

$$\widetilde{X}_{ij} := X_{ij}1_{\{|X_{ij}|\le\kappa_n\}} - EX_{ij}1_{\{|X_{ij}|\le\kappa_n\}},$$

and set $\widetilde{X} := (\widetilde{X}_i)_{i=1}^n$ with $\widetilde{X}_i = (\widetilde{X}_{i1}, \dots, \widetilde{X}_{id})^\top$. Note that $\max_{i,j}|\widetilde{X}_{ij}| \le 2\kappa_n$. Also, since $P(X_{ij}^2 > x) \le 2e^{-x/B_n^2}$ for all $x \ge 0$, Lemma 5.4 in Koike (2019b) yields

$$EX_{ij}^2 1_{\{X_{ij}^2>\kappa_n^2\}} \le Ce^{-\kappa_n^2/B_n^2}\kappa_n^2 \le C\frac{B_n^2\log(dn)}{n^5}. \tag{2.18}$$

Step 2. Let $\widetilde{W} := n^{-1/2}\sum_{i=1}^n \widetilde{X}_i$. On the event $\max_{1\le i\le n}\|X_i\|_\infty \le \kappa_n$, we have

$$|W_j - \widetilde{W}_j| = \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n EX_{ij}1_{\{|X_{ij}|>\kappa_n\}}\right| \le \frac{1}{\sqrt{n}}\sum_{i=1}^n\sqrt{EX_{ij}^2 1_{\{X_{ij}^2>\kappa_n^2\}}} \le C\frac{B_n\sqrt{\log(dn)}}{n^2}.$$

Therefore, Lemma 6.1 in Chernozhukov et al. (2019) yields

$$P\left(\|W - \widetilde{W}\|_\infty > C\frac{B_n\sqrt{\log(dn)}}{n^2}\right) \le P\Big(\max_{1\le i\le n}\|X_i\|_\infty > \kappa_n\Big) \le \frac{1}{2n^4}.$$

From this estimate and the Gaussian anti-concentration inequality, we obtain

$$\sup_{h=1_A:A\in\mathcal{R}}|Eh(W) - Eh(Z)| \le C\left(\frac{1}{2n^4} + \frac{B_n\log d\sqrt{\log(dn)}}{n^2} + \widetilde{\delta}\right),$$

where

$$\widetilde{\delta} := \sup_{h=1_A:A\in\mathcal{R}}|Eh(\widetilde{W}) - Eh(Z)|.$$

Therefore, the proof is completed once we show

$$\widetilde{\delta} \le C\left(\frac{B_n^2\log^4(dn)}{\sigma_*^4 n}\right)^{1/3}.$$

Step 3. We apply Theorem 1.3 to $\widetilde{X}$ and $\widetilde{W}$ with $\eta := 4\kappa_n/\sqrt{n}$. By construction we have $\delta_3(\eta) = 0$. Meanwhile, we have

$$E\left[\max_{1\le j,k\le d}\left|\frac{1}{2n}\sum_{i=1}^n\big((\widetilde{X}'_{ij} - \widetilde{X}_{ij})(\widetilde{X}'_{ik} - \widetilde{X}_{ik}) - E[(\widetilde{X}'_{ij} - \widetilde{X}_{ij})(\widetilde{X}'_{ik} - \widetilde{X}_{ik})]\big)\right|\right]$$
$$\le C\frac{\sqrt{\log d}}{n}E\left[\sqrt{\max_{1\le j,k\le d}\sum_{i=1}^n(X'_{ij} - X_{ij})^2(X'_{ik} - X_{ik})^2}\right] \le C\sqrt{\delta_2\log d},$$

where the first inequality follows from Nemirovski's inequality:

LEMMA 2.5 (Lemma 14.24 in Bühlmann and van de Geer (2011)). *Let Y_i be independent random variables taking values in a measurable space $\mathcal{Y}$ and let $\gamma_1, \dots, \gamma_p$ be real-valued measurable functions on $\mathcal{Y}$ such that $E\gamma_j(Y_i)$ exists. For $m \ge 1$ and $p \ge e^{m-1}$, we have*

$$E\max_{1\le j\le p}\left|\sum_{i=1}^n(\gamma_j(Y_i) - E\gamma_j(Y_i))\right|^m \le [8\log(2p)]^{m/2}E\left[\max_{1\le j\le p}\sum_{i=1}^n\gamma_j^2(Y_i)\right]^{m/2},$$

and the second one follows from the Schwarz inequality. We also have

$$\frac{1}{2n}\sum_{i=1}^{n} E[(\widetilde{X}'_{ij}-\widetilde{X}_{ij})(\widetilde{X}'_{ik}-\widetilde{X}_{ik})] = \frac{1}{n}\sum_{i=1}^{n} E[\widetilde{X}_{ij}\widetilde{X}_{ik}]$$

and $\Sigma_{jk} = \frac{1}{n}\sum_{i=1}^{n} E[X_{ij}X_{ik}]$. Therefore, noting $X_{ij} - \widetilde{X}_{ij} = X_{ij}1_{\{|X_{ij}|>\kappa_n\}} - EX_{ij}1_{\{|X_{ij}|>\kappa_n\}}$, the Schwarz inequality and (2.18) imply that

$$\begin{aligned}
&\max_{1\le j,k\le d}\left|\Sigma_{jk} - \frac{1}{2n}\sum_{i=1}^{n} E[(\widetilde{X}'_{ij}-\widetilde{X}_{ij})(\widetilde{X}'_{ik}-\widetilde{X}_{ik})]\right| \\
&\quad\le \max_{1\le j,k\le d}\max_{1\le i\le n}\bigl(\sqrt{E(X_{ij}-\widetilde{X}_{ij})^2 E\widetilde{X}_{ik}^2} + \sqrt{EX_{ij}^2 E(X_{ik}-\widetilde{X}_{ik})^2}\bigr) \\
&\quad\le C\frac{B_n^2\sqrt{\log(dn)}}{n^{5/2}} \le C\frac{B_n^2\log(dn)}{n}.
\end{aligned}$$

Consequently, we obtain

$$\delta_1 \le C\biggl(\sqrt{\delta_2\log d} + \frac{B_n^2\log(dn)}{n}\biggr).$$

Moreover, Lemma 9 in Chernozhukov, Chetverikov and Kato (2015) yields

$$\begin{aligned}
\delta_2 &\le \frac{C}{n^2}\biggl\{\max_{1\le j\le d} E\biggl[\sum_{i=1}^{n}(\widetilde{X}'_{ij}-\widetilde{X}_{ij})^4\biggr] + E\Bigl[\max_{1\le i\le n}\max_{1\le j\le d}(\widetilde{X}'_{ij}-\widetilde{X}_{ij})^4\Bigr]\log d\biggr\} \\
&\le C\biggl(\frac{B_n^2}{n} + \frac{B_n^4\log^3(dn)}{n^2}\biggr) \le C\frac{B_n^2}{n},
\end{aligned}$$

where the last inequality follows from (2.17). Therefore, for any $t>0$ satisfying $\eta/\sqrt{t} \le \sigma_*/\sqrt{\log d}$, we have

$$\widetilde{\delta} \le C\biggl(\frac{1}{\sigma_*^2}\sqrt{\frac{B_n^2\log^3(dn)}{n}}(|\log t|\vee 1) + \frac{B_n^2\log^2 d}{\sigma_*^4 n}\frac{1}{t} + \sqrt{t}\log d\biggr).$$

Now let $t = (B_n^2\log(dn)/\sigma_*^4 n)^{2/3}$. Then we have

$$\frac{\eta}{\sqrt{t}}\sqrt{\log d} \le 4\sqrt{5}\cdot\frac{B_n\log(dn)}{\sqrt{n}}\cdot\frac{n^{1/3}\sigma_*^{4/3}}{B_n^{2/3}\log^{1/3}(dn)} = 4\sqrt{5}\biggl(\frac{B_n^2\log^4(dn)}{\sigma_*^4 n}\biggr)^{1/6}\sigma_*^2 \le \sigma_*$$

by (2.17). So we can apply the above estimate with this t and obtain

$$\begin{aligned}
\widetilde{\delta} &\le C\biggl\{\sqrt{\frac{B_n^2\log^3(dn)}{\sigma_*^4 n}}\biggl|\log\frac{B_n^2\log(dn)}{\sigma_*^4 n}\biggr| + \biggl(\frac{B_n^2\log^4(dn)}{\sigma_*^4 n}\biggr)^{1/3}\biggr\} \\
&\le C\biggl(\frac{B_n^2\log^4(dn)}{\sigma_*^4 n}\biggr)^{1/3},
\end{aligned}$$

where the last line follows from the inequality $|\log x| \le C/x^{1/6}$ for $0<x\le 1$.

2.9. *Proof of Theorem 1.4.* Without loss of generality, we assume that the right-hand side of (1.8) is finite. Let

$$Y_i = \sum_{i'\in A_i} X_{i'}, \qquad Y_{ii'} = \sum_{i''\in A_{ii'}} X_{i''}.$$

From the independence assumption and $EX_i = 0$, we have, with U being a uniform distribution on $[0, 1]$ and independent of everything else,

$$\begin{aligned} EW \cdot \nabla\psi_t(W) &= \sum_{i=1}^n EX_i \cdot \big(\nabla\psi_t(W) - \nabla\psi_t(W - Y_i)\big) \\ &= \sum_{i=1}^n \sum_{i' \in A_i} \sum_{j,k=1}^d EX_{ij}X_{i'k}\partial_{jk}\psi_t(W - UY_i) \\ &= \sum_{i=1}^n \sum_{i' \in A_i} \sum_{j,k=1}^d EX_{ij}X_{i'k}\big[\partial_{jk}\psi_t(W - UY_i) - \partial_{jk}\psi_t(W - Y_{ii'})\big] \\ &\quad + \sum_{i=1}^n \sum_{i' \in A_i} \sum_{j,k=1}^d EX_{ij}X_{i'k}E\partial_{jk}\psi_t(W - Y_{ii'}). \end{aligned}$$

Because

$$\sum_{i=1}^n \sum_{i' \in A_i} EX_{ij}X_{i'k}E\partial_{jk}\psi_t(W) = \Sigma_{jk}E\partial_{jk}\psi_t(W),$$

we have by (2.11)

$$\begin{aligned} &\big|ET_t\tilde{h}(W)\big| \\ &\quad = \big|E\big[\langle \Sigma, \operatorname{Hess}\psi_t(W)\rangle_{H.S.} - W \cdot \nabla\psi_t(W)\big]\big| \\ &\quad \le \frac{C}{\sigma_*^3\sqrt{t}}(\log d)^{3/2}\sum_{i=1}^n \sum_{i' \in A_i} \sum_{i'' \in A_{ii'}} E\Big[\max_{1 \le j,k,l \le d}\big(|X_{ij}X_{i'k}X_{i''l}| + |X_{ij}X_{i'k}|E|X_{i''l}|\big)\Big]. \end{aligned}$$

Optimizing t gives the desired bound.

APPENDIX

A.1. Proof of Proposition 1.1. It suffices to show that there is a sequence $(x_n)_{n=1}^\infty$ of real numbers such that

$$\rho := \limsup_{n \to \infty} \sqrt{\frac{n}{\log^3 d}}\Big|P\Big(\max_{1 \le j \le d} W_j \le x_n\Big) - P\Big(\max_{1 \le j \le d} Z_j \le x_n\Big)\Big| > 0.$$

We denote by ϕ_1 and Φ_1 the density and distribution function of the standard normal distribution, respectively. For every n, we define $x_n \in \mathbb{R}$ as the solution of the equation $\Phi_1(x)^d = e^{-1}$, that is, $x_n := \Phi_1^{-1}(e^{-1/d})$. Then we have $x_n/\sqrt{2\log d} \to 1$ and $d(1 - \Phi_1(x_n)) \to 1$ as $n \to \infty$ (cf. the proof of Proposition 2.1 in Koike (2019b)). Applying Theorem 1 in Arratia, Goldstein and Gordon (1989) with $I = \{1, \dots, d\}$, $B_\alpha = \{\alpha\}$ and $X_\alpha = 1_{\{W_\alpha > x_n\}}$, we obtain

$$\Big|P\Big(\max_{1 \le j \le d} W_j \le x_n\Big) - e^{-\lambda_n}\Big| \le dP(W_1 > x_n)^2,$$

where $\lambda_n := dP(W_1 > x_n)$. By an analogous argument we also obtain

$$\Big|P\Big(\max_{1 \le j \le d} Z_j \le x_n\Big) - e^{-d(1-\Phi_1(x_n))}\Big| \le d\big(1 - \Phi_1(x_n)\big)^2.$$

Hence we have

$$\begin{aligned} &\Big|P\Big(\max_{1 \le j \le d} W_j \le x_n\Big) - P\Big(\max_{1 \le j \le d} Z_j \le x_n\Big)\Big| \\ &\quad \ge \big|e^{-\lambda_n} - e^{-d(1-\Phi_1(x_n))}\big| - dP(W_1 > x_n)^2 - d\big(1 - \Phi_1(x_n)\big)^2. \end{aligned}$$

Now, since $x_n = O(\sqrt{\log d}) = o(n^{1/6})$ by assumption, Theorem 1 in Petrov ((1975), Chapter VIII) (see also eq. (2.41) in Petrov ((1975), Chapter VIII)) implies

$$\frac{P(W_1 > x_n)}{1 - \Phi_1(x_n)} = \exp\left(\frac{\gamma}{6\sqrt{n}} x_n^3\right) + O\left(\frac{x_n + 1}{\sqrt{n}}\right). \tag{A.1}$$

In particular, since $d(1 - \Phi_1(x_n)) \to 1$, we have $dP(W_1 > x)^2 = O(d^{-1})$ and $d(1 - \Phi_1(x_n))^2 = O(d^{-1})$. Thus we obtain

$$\rho \geq \limsup_{n\to\infty} \sqrt{\frac{n}{\log^3 d}} \left| e^{-\lambda_n} - e^{-d(1-\Phi_1(x_n))} \right| \tag{A.2}$$

because $d^{-1} = o(n^{-1} \log^3 d)$ by assumption. Moreover, using the Taylor expansion of the exponential function around 0, we deduce from (A.1)

$$\lambda_n = d\big(1 - \Phi_1(x_n)\big) + \frac{\gamma}{6\sqrt{n}} x_n^3 + o\left(\frac{x_n^3}{\sqrt{n}}\right)$$

and

$$e^{-\{\lambda_n - d(1-\Phi_1(x_n))\}} = 1 - \frac{\gamma}{6\sqrt{n}} x_n^3 + o\left(\frac{x_n^3}{\sqrt{n}}\right).$$

Therefore, by (A.2) we conclude that

$$\rho \geq e^{-1} \frac{\sqrt{2}|\gamma|}{3}$$

because $x_n / \sqrt{2 \log d} \to 1$. This completes the proof.

A.2. Proof of Lemma 2.2. First we introduce some notation. We denote by ϕ_1 and Φ_1 the density and distribution function of the standard normal distribution, respectively. We set $\bar{\phi}_1(u) := \phi_1(u)/\Phi_1(u)$. Obviously, $\bar{\phi}_1$ is strictly decreasing on $[0, \infty)$.

For a nonnegative integer ν, the νth Hermite polynomial is denoted by H_ν: $H_\nu(u) = (-1)^\nu \phi_1(u)^{-1} \phi_1^{(\nu)}(u)$. When $\nu \geq 1$, we define the functions h_ν and $\bar{h}_\nu$ on $\mathbb{R}$ by

$$h_\nu(u) = H_{\nu-1}(u)\phi_1(u), \qquad \bar{h}_\nu(u) = h_\nu(u)/\Phi_1(u) = H_{\nu-1}(u)\bar{\phi}_1(u) \quad (u \in \mathbb{R}).$$

A simple computation shows

$$h_\nu'(u) = -h_{\nu+1}(u), \qquad \bar{h}_\nu'(u) = -\big\{H_\nu(u) + \bar{h}_\nu(u)\big\}\bar{\phi}_1(u). \tag{A.3}$$

Also, we define the functions λ and Λ on $[0, \infty)$ by

$$\lambda(u) = \frac{\phi_1(u)}{\phi_1(u + 2\eta)} = e^{2u\eta + 2\eta^2}, \qquad \Lambda(u) = \frac{\Phi_1(u)}{\Phi_1(u + 2\eta)} \quad (u \in [0, \infty)).$$

A simple computation shows

$$\Lambda'(u) = \Lambda(u)\big\{\bar{\phi}_1(u) - \bar{\phi}_1(u + 2\eta)\big\}. \tag{A.4}$$

In particular, Λ is nondecreasing on $[0, \infty)$.

To extend the proof for the case with $r = 1$ and $\eta = 0$ to the general case, we need to deduce a bound analogous to (2.1). Roughly speaking, we need to replace ϕ_1 in the middle equation of (2.1) by h_ν to accomplish this. In the derivation of (2.1), it plays a crucial role that ϕ_1 is decreasing on $[0, \infty)$. However, h_ν does not have this property in general, so we will dominate it by an appropriate decreasing function to proceed analogously to the derivation of (2.1). For this purpose, we need to introduce some additional notation. We denote by u_ν the

maximum root of H_ν. For example, $u_1 = 0$, $u_2 = 1$, $u_3 = \sqrt{3}$. It is evident that H_ν is positive and strictly increasing on (u_ν, ∞). We also have $u_1 < u_2 < \cdots$ (see, e.g., Szegö ((1939), Theorem 3.3.2)). Finally, set $M_\nu := \max_{0\le u\le u_\nu} |H_{\nu-1}(u)| < \infty$ and define the function $\tilde{h}_\nu : [0,\infty) \to (0,\infty)$ by

$$\tilde{h}_\nu(u) = M_\nu \phi_1(u) 1_{[0,u_\nu]}(u) + h_\nu(u) 1_{(u_\nu,\infty)}(u) \quad (u \in [0,\infty)).$$

LEMMA A.1. *$\tilde{h}_\nu$ is decreasing on $[0,\infty)$ and $|h_\nu(u)| \le \tilde{h}_\nu(|u|)$ for all $u \in \mathbb{R}$.*

PROOF. Note that $h_\nu'(u) < 0$ when $u > u_\nu$. Then, $\tilde{h}_\nu$ is evidently decreasing on $[0,\infty)$ by construction. The latter claim is also obvious by construction. □

We will also need a counterpart of the latter part of the proof for the case with $r = 1$ and $\eta = 0$. The subsequent two lemmas will be used for this purpose.

LEMMA A.2. *The function $u \mapsto H_\nu(u)\lambda(u) + \bar{h}_\nu(u)$ is strictly increasing on $[u_\nu,\infty)$.*

PROOF. Since $H_\nu(u)\lambda(u) + \bar{h}_\nu(u) = H_\nu(u)\{\lambda(u) - 1\} + \{H_\nu(u) + \bar{h}_\nu(u)\}$ and the function $u \mapsto H_\nu(u)\{\lambda(u) - 1\}$ is nondecreasing on $[u_\nu,\infty)$, it suffices to prove $g := H_\nu + \bar{h}_\nu$ is strictly increasing on $[u_\nu,\infty)$. We have

$$\begin{aligned} g'(u) &= H_\nu'(u) - \{H_\nu(u) + \bar{h}_\nu(u)\}\bar{\phi}_1(u) \\ &= \Phi_1(u)^{-1}\{H_\nu'(u)\Phi_1(u) - h_{\nu+1}(u) - \bar{h}_\nu(u)\phi_1(u)\}. \end{aligned}$$

So we complete the proof once we show $g_1(u) := H_\nu'(u)\Phi_1(u) - h_{\nu+1}(u) - \bar{h}_\nu(u)\phi_1(u) > 0$ for all $u > u_\nu$. We have

$$\begin{aligned} g_1'(u) &= H_\nu''(u)\Phi_1(u) + H_\nu'(u)\phi_1(u) + h_{\nu+2}(u) - \bar{h}_\nu'(u)\phi_1(u) + u\bar{h}_\nu(u)\phi_1(u) \\ &= \nu(\nu-1)H_{\nu-2}(u)\Phi_1(u) + uH_\nu(u)\phi_1(u) - \bar{h}_\nu'(u)\phi_1(u) + u\bar{h}_\nu(u)\phi_1(u), \end{aligned}$$

where the identity $H_{\nu+1}(u) = uH_\nu(u) - H_\nu'(u)$ is used to deduce the last line. Since $H_k(u) > 0$ for $k \le \nu$ and $u > u_\nu$, we have $g_1'(u) > 0$ for $u > u_\nu$. Thus

$$g_1(u) > g_1(u_\nu) = H_{\nu-1}(u_\nu)\{\nu\Phi_1(u_\nu) - \phi_1(u_\nu)\bar{\phi}_1(u_\nu)\} \ge 1/2 - 1/\pi > 0$$

for all $u > u_\nu$. □

Define the functions F_ν and G on $\mathbb{R}^d$ by

$$F_\nu(x) = \sum_{j=1}^d \bar{h}_\nu(x_j)\Lambda(x_j), \qquad G(x) = \prod_{k=1}^d \Phi_1(x_k + 2\eta) \quad (x \in \mathbb{R}^d).$$

LEMMA A.3. *For any $\beta > 0$,*

$$\sup_{x\in[u_\nu,\infty)^d} F_\nu(x)^\beta G(x) = O\big((\log d)^{\beta\nu/2}\big)$$

as $d \to \infty$.

PROOF. Define the function f on $\mathbb{R}^d$ by $f(x) = F_\nu(x)^\beta G(x)$, $x \in \mathbb{R}^d$. First we prove f has a maximizer on $[u_\nu,\infty)^d$. Using (A.3)–(A.4), we obtain

$$\begin{aligned} \partial_l f(x) &= \beta F_\nu(x)^{\beta-1}\{\bar{h}_\nu'(x_l)\Lambda(x_l) + \bar{h}_\nu(x_l)\Lambda'(x_l)\}G(x) + F_\nu(x)^\beta\bar{\phi}_1(x_l + 2\eta)G(x) \\ &= [-\beta\{H_\nu(x_l)\lambda(x_l) + \bar{h}_\nu(x_l)\}\Lambda(x_l) + F_\nu(x)]\bar{\phi}_1(x_l + 2\eta)F_\nu(x)^{\beta-1}G(x). \end{aligned}$$

Now, since $\beta\{H_\nu(x)\lambda(x)+\bar{h}_\nu(x)\}\Lambda(x)\to\infty$ as $x\to\infty$ while $\sup_{x\in[u_\nu,\infty)^d}F_\nu(x)<\infty$, there is a number $\bar{u}\geq u_\nu$ such that for all x with $x_l\geq\bar{u}$, $\partial_l f(x)<0$. This means $\sup_{x\in[u_\nu,\infty)^d}f(x)=\sup_{x\in[u_\nu,\bar{u}]^d}f(x)$ and thus f has a maximizer on $[u_\nu,\infty)^d$.

Let x^* be a maximizer of f on $[u_\nu,\infty)^d$. Then the proof is completed once we show $f(x^*)=O((\log d)^{\beta\nu/2})$ as $d\to\infty$. Let m be the number of components in x^* greater than u_ν. If $m=0$, then

$$f(x^*)\leq\{d\bar{h}_\nu(u_\nu)\Lambda(u_\nu)\}^\beta\Phi_1(u_\nu+2\eta)^d=o(1)$$

as $d\to\infty$, so it suffices to consider the case $m\geq1$. Since f is symmetric, any permutation of x^* is a maximizer of f. Thus, from the definition of m, we may assume $x_1^*,\dots,x_m^*>u_\nu$ and $x_{m+1}^*=\cdots=x_d^*=u_\nu$ without loss of generality. Then, for every $l=1,\dots,m$, we must have $\partial_l f(x^*)=0$. Thus we obtain

$$\{H_\nu(x_1^*)\lambda(x_1^*)+\bar{h}_\nu(x_1^*)\}\Lambda(x_1^*)=\cdots=\{H_\nu(x_m^*)\lambda(x_m^*)+\bar{h}_\nu(x_m^*)\}\Lambda(x_m^*)=\beta^{-1}F_\nu(x^*).$$

Since Λ is nondecreasing, the function $u\mapsto\beta\{H_\nu(u)\lambda(u)+\bar{h}_\nu(u)\}\Lambda(u)$ is strictly increasing on $[u_\nu,\infty)$ by Lemma A.2. Therefore, we have $x_1^*=\cdots=x_m^*=:u^*$ and hence

$$\beta\{H_\nu(u^*)\lambda(u^*)+\bar{h}_\nu(u^*)\}\Lambda(u^*)=F_\nu(x^*)=m\bar{h}_\nu(u^*)\Lambda(u^*)+(d-m)\bar{h}_\nu(u_\nu)\Lambda(u_\nu).$$

Recall that $u^*=u^*(d)$ depends on d. Then, dividing the sequence $\{u^*(d)\}_{d=1}^\infty$ into two parts and separately handling them if necessary, it suffices to consider the following two cases: (i) $u^*\leq\sqrt{8\log d}$ for all d. (ii) $u^*>\sqrt{8\log d}$ for all d. In the first case, we have $\lambda(u^*)=O(1)$ as $d\to\infty$ and thus

$$f(x^*)\leq\beta^\beta\{H_\nu(u^*)\lambda(u^*)+\bar{h}_\nu(u^*)\}^\beta=O((\log d)^{\beta\nu/2})$$

as $d\to\infty$. In the second case, note that $me^{-(u^*)^2/4}\to0$ as $d\to\infty$ because $u^*\geq\sqrt{8\log d}$. This yields $m\bar{h}_\nu(u^*)\Lambda(u^*)\to0$. Hence we obtain

$$f(x^*)=O(\{(d-m)\bar{h}_\nu(u_\nu)\Lambda(u_\nu)\}^\beta\Phi_1(u_\nu+2\eta)^{d-m})=O(1)$$

as $d\to\infty$ because $\Phi_1(u_\nu+2\eta)<1$ and $\bar{h}_\nu(u_\nu)\Lambda(u_\nu)$ is a constant. In the end, we complete the proof. □

Now we are ready to prove Lemma 2.2.

PROOF OF LEMMA 2.2. For every $q\in\{1,\dots,r\}$, set

$$\mathcal{N}_q(r):=\{(\nu_1,\dots,\nu_q)\in\mathbb{Z}^q:\nu_1,\dots,\nu_q\geq1,\nu_1+\cdots+\nu_q=r\},$$
$$\mathcal{J}_q(d):=\{(j_1,\dots,j_q)\in\{1,\dots,d\}^q:j_p\neq j_{p'}\text{ if }p\neq p'\}.$$

Then we have for all $A\in\mathcal{R}$

$$\sum_{j_1,\dots,j_r=1}^d\sup_{y\in R(0;\eta)}\left|\int_A\partial_{j_1,\dots,j_r}\phi(z+y)\,dz\right|$$
$$\leq C_r\sum_{q=1}^r\sum_{(\nu_1,\dots,\nu_q)\in\mathcal{N}_q(r)}\sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)}\sup_{y\in R(0,\eta)}\left|\int_A\partial_{j_1}^{\nu_1}\cdots\partial_{j_q}^{\nu_q}\phi(z+y)\,dz\right|,$$

where $C_r>0$ depends only on r. Therefore, we obtain the desired result once we prove

$$\sup_{A\in\mathcal{R}}\sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)}\sup_{y\in R(0;\eta)}\left|\int_A\partial_{j_1}^{\nu_1}\cdots\partial_{j_q}^{\nu_q}\phi(z+y)\,dz\right|=O((\log d)^{r/2})\quad\text{as }d\to\infty$$

for any (fixed) $(\nu_1,\dots,\nu_q)\in\mathcal{N}_q(r)$ with $q\in\{1,\dots,r\}$.

Take $A = \prod_{j=1}^d (a_j, b_j) \in \mathcal{R}$ arbitrarily and set

$$I_A := \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \sup_{y\in R(0;\eta)} \left| \int_A \partial_{j_1}^{\nu_1} \cdots \partial_{j_q}^{\nu_q} \phi(z+y)\, dz \right|.$$

Then we have

$$\begin{aligned} I_A &= \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \sup_{y\in R(0;\eta)} \left(\prod_{p=1}^q |h_{\nu_p}(b_{j_p}+y_{j_p}) - h_{\nu_p}(a_{j_p}+y_{j_p})| \right) \\ &\quad \times \prod_{k:k\neq j_1,\dots,j_q} \{\Phi_1(b_k+y_k) - \Phi_1(a_k+y_k)\} \\ &\le \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \sup_{y\in R(0;\eta)} \left(\prod_{p=1}^q (|h_{\nu_p}(b_{j_p}+y_{j_p})| + |h_{\nu_p}(a_{j_p}+y_{j_p})|) \right) \\ &\quad \times \prod_{k:k\neq j_1,\dots,j_q} \{\Phi_1(b_k+y_k) + \Phi_1(-a_k-y_k) - 1\}, \end{aligned}$$

where we use the identity $1 - \Phi_1(x) = \Phi_1(-x)$ to deduce the last line. Set $c_j := (|a_j| \wedge |b_j|) \vee \eta$, $j = 1,\dots,d$. Then, we have $\min\{|a_j + y_j|, |b_j + y_j|\} \ge c_j - \eta \ge 0$ for all j. Thus, noting that Φ is increasing and bounded by 1, we obtain by Lemma A.1

$$\begin{aligned} I_A &\le 2^q \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \left(\prod_{p=1}^q \tilde{h}_{\nu_p}(c_{j_p}-\eta) \right) \prod_{k:k\neq j_1,\dots,j_q} \Phi_1(c_k+\eta) \\ &= 2^q \sum_{(j_1,\dots,j_q)\in\mathcal{J}_q(d)} \left(\prod_{p=1}^q \frac{\tilde{h}_{\nu_p}(c_{j_p}-\eta)}{\Phi_1(c_{j_p}-\eta)} \frac{\Phi_1(c_{j_p}-\eta)}{\Phi_1(c_{j_p}+\eta)} \right) \prod_{k=1}^d \Phi_1(c_k+\eta) \\ &\le 2^q \left(\prod_{p=1}^q \sum_{j_p=1}^d \frac{\tilde{h}_{\nu_p}(c_{j_p}-\eta)}{\Phi_1(c_{j_p}-\eta)} \frac{\Phi_1(c_{j_p}-\eta)}{\Phi_1(c_{j_p}+\eta)} \right) \prod_{k=1}^d \Phi_1(c_k+\eta). \end{aligned}$$

Now, since $\sum_p \nu_p = r$, the generalized AM-GM inequality yields

$$\begin{aligned} I_A &\le 2^q \sum_{p=1}^q \frac{\nu_p}{r} \left(\sum_{j_p=1}^d \frac{\tilde{h}_{\nu_p}(c_{j_p}-\eta)}{\Phi_1(c_{j_p}-\eta)} \frac{\Phi_1(c_{j_p}-\eta)}{\Phi_1(c_{j_p}+\eta)} \right)^{r/\nu_p} \prod_{k=1}^d \Phi_1(c_k+\eta) \\ &\le C_r' \sum_{p=1}^q \left\{ \left(\sum_{j=1}^d \bar{\phi}_1(c_j-\eta)\Lambda(c_j-\eta) \right)^{r/\nu_p} \right. \\ &\quad \left. + \left(\sum_{j=1}^d \bar{h}_{\nu_p}(c_j-\eta)\Lambda(c_j-\eta) 1_{(u_{\nu_p},\infty)}(c_j-\eta) \right)^{r/\nu_p} \right\} \prod_{k=1}^d \Phi_1\big((c_k-\eta)+2\eta\big) \\ &\le C_r' \sum_{p=1}^q \Big\{ \sup_{x\in[0,\infty)^d} F_1(x)^{r/\nu_p} G(x) + \sup_{x\in[u_{\nu_p},\infty)^d} F_{\nu_p}(x)^{r/\nu_p} G(x) \Big\}, \end{aligned}$$

where $C_r' > 0$ depends only on r (note that $\bar{h}_\nu$ is positive on $[u_\nu, \infty)$). Consequently, by Lemma A.3 we conclude $\sup_{A\in\mathcal{R}} I_A = O((\log d)^{r/2})$ as $d \to \infty$. □

A.3. Proof of Proposition 1.2. We follow the proof of Proposition 1.4 of Zhai (2018). Let $A \in \mathcal{R}$ be a given hyperrectangle. For a parameter ϵ to be specified later, define

$$A^\epsilon = \Big\{x \in \mathbb{R}^d : \inf_{a\in A} \|x-a\|_\infty \le \epsilon\Big\},$$
$$A_\epsilon = \Big\{x \in \mathbb{R}^d : \inf_{a\in \mathbb{R}^d\setminus A} \|x-a\|_\infty \ge \epsilon\Big\}.$$

Applying the Gaussian anti-concentration inequality in Lemma 2.1 with $Y = (Z^\top, -Z^\top)^\top$ gives

$$P\big(Z \in A^\epsilon \setminus A\big) \le C\epsilon\sqrt{\log d}, \quad \text{and} \quad P(Z \in A \setminus A_\epsilon) \le C\epsilon\sqrt{\log d}.$$

We may regard T as being coupled to Z so that $E[|T-Z|^2] = \mathcal{W}_2(T,Z)^2$. Then

$$\begin{aligned} P(T\in A) &\le P\big(|T-Z|\le\epsilon, T\in A\big) + P\big(|T-Z|>\epsilon\big)\\ &\le P\big(Z\in A^\epsilon\big) + \epsilon^{-2}\mathcal{W}_2(T,Z)^2\\ &\le P(Z\in A) + C\epsilon\sqrt{\log d} + \epsilon^{-2}\mathcal{W}_2(T,Z)^2. \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z\in A) &\le P(Z\in A_\epsilon) + C\epsilon\sqrt{\log d}\\ &\le P\big(|T-Z|\le\epsilon, Z\in A_\epsilon\big) + P\big(|T-Z|>\epsilon\big) + C\epsilon\sqrt{\log d}\\ &\le P(T\in A) + \epsilon^{-2}\mathcal{W}_2(T,Z)^2 + C\epsilon\sqrt{\log d}. \end{aligned}$$

Thus,

$$\big|P(T\in A) - P(Z\in A)\big| \le \epsilon^{-2}\mathcal{W}_2(T,Z)^2 + C\epsilon\sqrt{\log d},$$

and taking $\epsilon = \frac{W_2(T,Z)^{2/3}}{(\log d)^{1/6}}$ gives the result.

Acknowledgments. We thank the two anonymous referees for their careful reading of the manuscript and for their valuable suggestions which led to many improvements. We also thank Victor Chernozhukov for useful comments on the proof of Lemma 2.2.

Funding. Fang X. was partially supported by Hong Kong RGC ECS 24301617 and GRF 14302418 and 14304917, a CUHK direct grant and a CUHK start-up grant. Koike Y. was partially supported by JST CREST Grant Number JPMJCR14D7 and JSPS KAKENHI Grant Numbers JP17H01100, JP18H00836, JP19K13668.

REFERENCES

ANDERSON, N. H., HALL, P. and TITTERINGTON, D. M. (1998). Edgeworth expansions in very-high-dimensional problems. *J. Statist. Plann. Inference* **70** 1–18. MR1649439 https://doi.org/10.1016/S0378-3758(97)00171-7

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D., HANSEN, C. and KATO, K. (2018). High-dimensional econometrics and regularized GMM. *Preprint*. Available at https://arxiv.org/abs/1806.01888.

BENTKUS, V. (2005). A Lyapunov type bound in $\mathbf{R}^d$. *Theory Probab. Appl.* **49** 311–323.

BENTKUS, V. and GÖTZE, F. (1997). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Related Fields* **109** 367–416. MR1481126 https://doi.org/10.1007/s004400050136

BHATTACHARYA, R. N. and HOLMES, S. (2010). An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem. *Preprint.* Available at https://arxiv.org/abs/1003.4254.

BHATTACHARYA, R. N. and RAO, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. MR0436272

BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2013). Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **41** 2479–2512. MR3112923 https://doi.org/10.1214/12-AOP780

BOGATYREV, S. A., GÖTZE, F. and ULYANOV, V. V. (2006). Non-uniform bounds for short asymptotic expansions in the CLT for balls in a Hilbert space. *J. Multivariate Anal.* **97** 2041–2056. MR2301275 https://doi.org/10.1016/j.jmva.2006.01.010

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data. Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

CHATTERJEE, S. (2005). A simple invariance theorem. *Preprint.* Available at https://arxiv.org/abs/math/0508213.

CHATTERJEE, S. (2008a). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859 https://doi.org/10.1214/07-AOP370

CHATTERJEE, S. (2008b). Chaos, concentration, and multiple valleys. Preprint. Available at: https://arxiv.org/abs/0810.4221.

CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

CHATTERJEE, S. and MECKES, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283. MR2453473

CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

CHEN, X. and KATO, K. (2019). Randomized incomplete U-statistics in high dimensions. *Ann. Statist.* **47** 3127–3156. MR4025737 https://doi.org/10.1214/18-AOS1773

CHEN, L. H. Y. and RÖLLIN, A. (2010). Stein couplings for normal approximation. *Preprint.* Available at https://arxiv.org/abs/1003.6039.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017a). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017b). Detailed proof of Nazarov's inequality. *Preprint.* Available at https://arxiv.org/abs/1711.10696.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. MR4009488 https://doi.org/10.1093/restud/rdy065

CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2019). Improved central limit theorem and bootstrap approximation in high dimensions. *Preprint.* Available at https://arxiv.org/abs/1912.10529.

COURTADE, T. A., FATHI, M. and PANANJADY, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898

DUNG, N. T. (2019). Explicit rates of convergence in the multivariate CLT for nonlinear statistics. *Acta Math. Hungar.* **158** 173–201. MR3950207 https://doi.org/10.1007/s10474-019-00917-6

ELDAN, R., MIKULINCER, D. and ZHAI, A. (2020). The CLT in high dimensions: Quantitative bounds via martingale embedding. *Ann. Probab.* **48** 2494–2524. MR4152649 https://doi.org/10.1214/20-AOP1429

ESSEEN, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. MR0014626 https://doi.org/10.1007/BF02392223

FATHI, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. MR3980918 https://doi.org/10.1214/18-AOP1305

FATHI, M. (2021). Higher-order Stein kernels for Gaussian approximation. *Studia Math.* **256** 241–258. MR4172696 https://doi.org/10.4064/sm190415-28-10

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

GÖTZE, F. and ULYANOV, V. V. (2003). Asymptotic distribution of χ^2-type statistics. In *Research Group Spectral Analysis, Asymptotic Distributions and Stochastic Dynamics* Bielefeld, Univ. Preprint 03–033.

GÖTZE, F. and ZAITSEV, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *Ann. Probab.* **42** 354–397. MR3161488 https://doi.org/10.1214/13-AOP839

KIM, Y. T. and PARK, H. S. (2015). Kolmogorov distance for multivariate normal approximation. *Korean J. Math.* **23** 1–10.

KOIKE, Y. (2019a). Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. *Ann. Statist.* **47** 1663–1687. MR3911126 https://doi.org/10.1214/18-AOS1731

KOIKE, Y. (2019b). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* Preprint available at https://arxiv.org/abs/1911.00160.

KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2018). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Preprint.* Available at https://arxiv.org/abs/1804.02605.

LEDOUX, M., NOURDIN, I. and PECCATI, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** 256–306. MR3320893 https://doi.org/10.1007/s00039-015-0312-0

MECKES, E. (2006). *An Infinitesimal Version of Stein's Method of Exchangeable Pairs*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. MR2708805

NAGAEV, S. V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory* (*Tashkent*, 1975) 419–438. Springer. MR0443043

NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

NOURDIN, I., PECCATI, G. and SWAN, Y. (2014). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266** 3170–3207. MR3158721 https://doi.org/10.1016/j.jfa.2013.09.017

NOURDIN, I., PECCATI, G. and YANG, X. (2020). Multivariate normal approximation on the Wiener space: New bounds in the convex distance. *Preprint.* Available at https://arxiv.org/abs/2001.02188.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. MR0388499

PROKHOROV, Y. V. and ULYANOV, V. V. (2013). Some approximation problems in statistics and probability. In *Limit Theorems in Probability, Statistics and Number Theory*. *Springer Proc. Math. Stat.* **42** 235–249. Springer, Heidelberg. MR3079145 https://doi.org/10.1007/978-3-642-36068-8_11

RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554 https://doi.org/10.1214/09-AOP467

SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

SAZONOV, V. V. (1972). On a bound for the rate of convergence in the multidimensional central limit theorem. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 563–581. MR0400351

SONG, Y., CHEN, X. and KATO, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electron. J. Stat.* **13** 4794–4848. MR4038726 https://doi.org/10.1214/19-EJS1643

STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

SZEGÖ, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc., New York. MR0000077

ZHAI, A. (2018). A high-dimensional CLT in $\mathcal{W}_2$ distance with near optimal convergence rate. *Probab. Theory Related Fields* **170** 821–845. MR3773801 https://doi.org/10.1007/s00440-017-0771-3

ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697 https://doi.org/10.3150/17-BEJ939

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512